

Contributions to the Understanding of the MSK Modulation

Dayan Adionel Guimarães

Abstract—This tutorial deals with key aspects of the MSK (Minimum Shift Keying) modulation, aiming at unveiling some of its hidden concepts. Signal generation and demodulation are analyzed in detail. Common questions concerning the study of the MSK modulation are addressed and answered, e.g. the similarities and differences among MSK, Sunde's FSK (Frequency Shift Keying) and SQPSK (Staggered Quaternary Phase-Shift Keying) or OQPSK (Offset QPSK); the relation among the modulating data stream, its differentially-decoded version, the frequency shifts and the phase shifts of the modulated signal, and the MSK signal-space representation.

Index Terms—MSK, FSK, SQPSK and OQPSK modulations.

Resumo—Este tutorial trata de aspectos chave sobre a modulação MSK, objetivando revelar alguns dos seus conceitos muitas vezes não revelados explicitamente. A geração e a demodulação do sinal MSK são analisadas em detalhe. Ao longo do trabalho procura-se responder a algumas questões intrigantes relacionadas com, por exemplo, as similaridades e diferenças entre as modulações MSK, FSK (de Sunde) e SQPSK ou OQPSK e a relação entre a seqüência moduladora, sua versão decodificada diferencialmente, os desvios de frequência e de fase do sinal modulado e a representação do sinal MSK no espaço euclidiano.

Palavras chave— Modulações MSK, FSK, SQPSK e OQPSK.

I. INTRODUCTION

The Minimum Shift Keying (MSK) modulation, also known as “fast FSK” [1], was first considered during the early 60s and 70s [2]-[4], and its characteristics have gained the attention of the scientific community during the subsequent decades.

MSK modulation has features such as constant envelope, compact spectrum and good error performance, which are all desirable in many digital communication systems. Its utilization goes from the Global System for Mobile Communication (GSM), in which a Gaussian-filtered MSK (GMSK) modulation is employed, to micro-satellite communications, positioning and navigation systems, hybrid optical/wireless communication systems, deep space communications and, more recently, to the Blue Ray disc technology [5], only to mention a few examples.

Like many recently rediscovered technologies developed several years, or even decades ago, the MSK modulation seems to be one more idea whose time has come.

Although covered in many papers and good books on Digital Communications, some of the concepts of this modulation are hidden or difficult to understand, representing opportunities for alternative approaches, like the one adopted in this tutorial. This approach is intended to help everyone who wants to have an understanding about the MSK modulation, especially the practicing engineers and the first-level graduate students in Telecommunications. It addresses some key questions about the MSK modulation, such as:

- 1 – To which extent the MSK modulation can be regarded as a special case of the conventional Sunde's [6] [7, p. 381] FSK (Frequency Shift Keying) modulation?
- 2 – To which extent the MSK modulation can be detected in the same way as the Sunde's FSK modulation?
- 3 – To which extent the MSK modulation can be regarded as a special case of the SQPSK or OQPSK (Staggered or Offset QPSK) modulation?
- 4 – To which extent the frequency and phase shifts of an MSK signal are related to the modulating data sequence?
- 5 – To which extent the phase shifts of an MSK signal can be related to the phase transition diagram on its signal-space representation?

The remaining of this work is organized as follows: Section II addresses some fundamental concepts about the signal-space representation, the complex representation of signals and systems, and the minimum separation between tones in an orthogonal FSK signaling. Section III is devoted to the analysis of the signal construction from the signal-space expansion and the complex representation approaches. The MSK spectral content, receiver structure and system performance are also analyzed in Section III. Further attributes and uses of the MSK modulation are summarized in Section IV, and Section V addresses the answers to the questions highlighted above, concluding the work.

II. BASIC CONCEPTS

In this section the reader are invited to revisit some fundamental concepts about signal-space representation and complex representation of signals and systems. Although applicable to the study of digital communications in general, these two concepts are essential for the study at hand, and will give us insight on different forms of MSK signal generation and detection. Additionally, the minimum tone separation for coherent detection of orthogonal FSK is analyzed, aiming at

Manuscript received on December 9, 2006.

D. A. Guimarães (dayan@inatel.br) is with INATEL - Instituto Nacional de Telecomunicações. Av. João de Camargo, 510 - Santa Rita do Sapucaí - MG - Brazil - 37540-000.

justifying the term *minimum* in the name of the MSK modulation.

A. Signal-space representation

The signal-space representation is constructed on the basis of linear combination theory, and it is very analogous to the vector algebra theory. Let us define an N -dimensional Euclidian space spanned by N orthogonal axes. Let us also define a set of orthogonal vectors $\{\phi_j\}$, $j = 1, 2, \dots, N$, normalized in the sense that they have unit length. These vectors are said to be *orthonormal* and to form an *orthonormal basis*.

Any vector \mathbf{v}_i , $i = 1, 2, \dots, M$ in the Euclidian space can be generated through the linear combination

$$\mathbf{v}_i = \sum_{j=1}^N v_{ij} \phi_j \quad (1)$$

where the coefficients v_{ij} correspond to the projection of the i -th vector on the j -th base vector. Their values can be determined by the dot product (or inner product) between \mathbf{v}_i and ϕ_j , that is

$$v_{ij} = \mathbf{v}_i^T \phi_j \quad (2)$$

where the superscript T denotes matrix transposition, $\mathbf{v}_i = [v_{i1} \ v_{i2} \ \dots \ v_{iN}]^T$ and ϕ_j is also an N -dimensional vector with a 1 in the j -th position and zeros otherwise, that is $\phi_j = [0 \ 1 \ 0 \ \dots \ 0]^T$ for $j = 2$ as an example.

Figure 1 illustrates these concepts for a two-dimensional ($N = 2$) Euclidian space and for two vectors ($M = 2$). The axes were labeled in a way to resemble the orthonormal base-vectors.

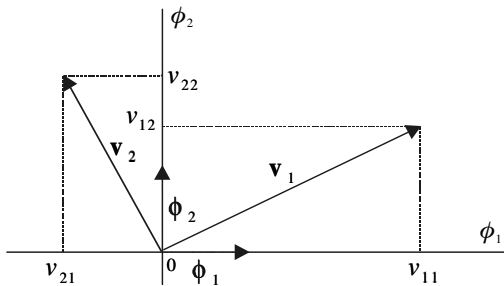


Figure 1. Vector-space representation for $M = 2$ and $N = 2$.

In a similar way, one can use the Euclidian space to represent coefficients that, in a linear combination, will give rise to signals instead of vectors. Then we have the signals

$$s_i(t) = \sum_{j=1}^N s_{ij} \phi_j(t), \quad i = 1, 2, \dots, M \quad (3)$$

where, now, the set $\{\phi_j(t)\}$ comprises N orthonormal *base-functions*, one function being orthogonal to each other and having unit energy, that is:

$$\int_0^T \phi_i(t) \phi_j(t) dt = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (4)$$

The set of functions $\{\phi_j(t)\}$ are also said to be *orthonormal* and to form an *orthonormal basis*.

Through Figure 1 it can be seen that the value of a coefficient is proportional to a measure of the orthogonality between the analyzed vector and the corresponding base-vector: the greater the orthogonality, the lesser the value of the coefficient. By analogy to the vector algebra, we can determine the values of the coefficients in (3) through a measure of orthogonality between the analyzed waveform and the corresponding base-function, which leads intuitively to

$$s_{ij} = \int_0^T s_i(t) \phi_j(t) dt, \quad \begin{cases} i = 1, 2, \dots, M \\ j = 1, 2, \dots, N \end{cases} \quad (5)$$

In fact (5) has a formal mathematical justification, which can be obtained by operating generically with (3) and (4):

$$\begin{aligned} \int_0^T x(t) y(t) dt &= \int_0^T \sum_{j=1}^N x_j \phi_j(t) \sum_{k=1}^N y_k \phi_k(t) dt \\ &= \sum_{j=1}^N \sum_{k=1}^N x_j y_k \int_0^T \phi_j(t) \phi_k(t) dt \\ &= \sum_{j=1}^N x_j y_j = \mathbf{x}^T \mathbf{y} \end{aligned} \quad (6)$$

Expression (6) states that the correlation in time domain has the inner product as its equivalent in the vector domain.

We are now ready to define the signal-space representation: since knowing the set of coefficients and base-functions is as good as to know the waveform signals themselves, we can also represent signals in a Euclidian space. In this representation we use points instead of vectors, to avoid polluting unnecessarily the graph. This kind of plot is also called *signal constellation*. Figure 2 shows a two-dimensional signal-space used to represent the signals $s_1(t)$ and $s_2(t)$ through the corresponding *signal vectors* \mathbf{s}_1 and \mathbf{s}_2 .

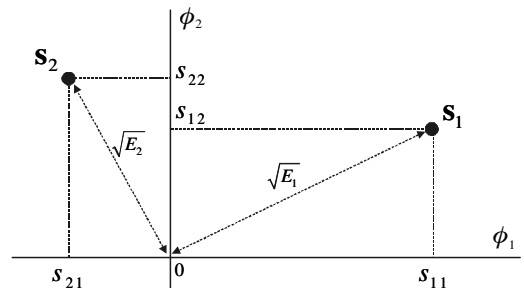


Figure 2. Signal-space representation for $M = 2$ and $N = 2$.

As can be noticed from Figure 2, the norm of a signal vector, that is, the length of this vector can be determined in the light of equation (6) by:

$$\sqrt{s_{i1}^2 + s_{i2}^2} = \sqrt{\mathbf{s}_i^T \mathbf{s}_i} = \sqrt{\int_0^T s_i^2(t) dt} = \sqrt{E_i} \quad (7)$$

Generally speaking, the distance from any signal vector to the origin of the coordinates is equal to the square root of the corresponding signal energy:

$$E_i = \int_0^T s_i^2(t) dt = \mathbf{s}_i^T \mathbf{s}_i = \sum_{j=1}^N s_{ij}^2 = \|\mathbf{s}_i\|^2 \quad (8)$$

As a complementary result, the squared Euclidian distance between two signal vectors is obtained through

$$d_{ik}^2 = \|\mathbf{s}_i - \mathbf{s}_k\|^2 = \sum_{j=1}^N (s_{ij} - s_{kj})^2 = \int_0^T [s_i(t) - s_k(t)]^2 dt \quad (9)$$

The concepts just described will be used later for the understanding of a particular form for the MSK signal generation and detection.

B. Complex representation of signals and systems

We start by reviewing the concept of Hilbert transform. Following [7] and [8], let $g(t)$ be a signal with Fourier transform $G(f)$. The Hilbert transform of $g(t)$ and the corresponding inverse transform are defined respectively by

$$\hat{g}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{g(\tau)}{t - \tau} d\tau = \int_{-\infty}^{\infty} g(\tau) \frac{1}{\pi(t - \tau)} d\tau \quad (10)$$

and

$$g(t) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\hat{g}(\tau)}{t - \tau} d\tau \quad (11)$$

In (10) we can identify that the Hilbert transform of $g(t)$ is the convolution between $g(t)$ and the function $1/\pi$.

By recalling that a convolution in the time domain corresponds to a multiplication in the frequency domain, and by using the Fourier transform pair

$$\frac{1}{\pi t} \Leftrightarrow -j \operatorname{sgn}(f), \quad (12)$$

where $\operatorname{sgn}(f)$ is the *sign* function or *signum* function defined by

$$\operatorname{sgn}(f) = \begin{cases} 1, & f > 0 \\ 0, & f = 0 \\ -1, & f < 0 \end{cases}, \quad (13)$$

then we can write:

$$\hat{G}(f) = -j \operatorname{sgn}(f) G(f) \quad (14)$$

Analyzing (14) we can see that the Hilbert transform of $g(t)$ corresponds to a phase shift of -90° for the positive frequencies of $G(f)$ and $+90^\circ$ for the its negative frequencies.

Let us now make use of another definition: the *analytic signal* or *pre-envelope* of $g(t)$:

$$g_+(t) = g(t) + j\hat{g}(t) \quad (15)$$

from where, using (14) and the definition of the *signum* function given in (13), we can obtain

$$G_+(f) = G(f) + \operatorname{sgn}(f)G(f) = \begin{cases} 2G(f), & f > 0 \\ G(0), & f = 0 \\ 0, & f < 0 \end{cases} \quad (16)$$

Now, consider a band-pass signal $g(t)$ whose bandwidth is essentially confined in $2W$ Hz and is small compared to the its carrier frequency f_c . According to (16), the analytic spectrum $G_+(f)$ is centered about f_c and contains only positive frequency components. Then, using the frequency-shifting property of the Fourier transform we can write:

$$g_+(t) = \tilde{g}(t) \exp(j2\pi f_c t) \quad (17)$$

where $\tilde{g}(t)$ is called the *complex envelope* of the signal $g(t)$ and it is clearly a low-pass signal.

Since $g_+(t)$ is a band-pass signal, we can determine the low-pass signal $\tilde{g}(t)$ through a frequency translation of $g_+(t)$ back to about $f = 0$. Using again the frequency-shifting property of the Fourier transform we can write

$$\begin{aligned} \tilde{g}(t) &= g_+(t) \exp(-j2\pi f_c t) \\ &= [g(t) + j\hat{g}(t)] \exp(-j2\pi f_c t) \end{aligned} \quad (18)$$

or, equivalently,

$$g(t) + j\hat{g}(t) = \tilde{g}(t) \exp(j2\pi f_c t) \quad (19)$$

Since the signal $g(t)$ is the real part of the left side of the expression above, we can obtain a very useful representation:

$$g(t) = \operatorname{Re}[\tilde{g}(t) \exp(j2\pi f_c t)] \quad (20)$$

Generally speaking, $\tilde{g}(t)$ can be a complex quantity, which can be expressed in the Cartesian form by:

$$\tilde{g}(t) = g_I(t) + jg_Q(t) \quad (21)$$

where the subscripts I and Q stand for *in-phase* and *quadrature*. Then, by substituting (21) in (20) we have, after some simplifications:

$$g(t) = g_I(t) \cos(2\pi f_c t) - g_Q(t) \sin(2\pi f_c t) \quad (22)$$

Both $g_I(t)$ and $g_Q(t)$ are low-pass signals and are called the in-phase component and the quadrature component of the signal $g(t)$, respectively. This is why we call $\tilde{g}(t)$ the *equivalent low-pass* version of the band-pass signal $g(t)$. This result will be used later on in this tutorial to describe a particular form for the MSK signal generation and detection.

Rewriting expression (21) in its polar form we have:

$$\tilde{g}(t) = a(t) \exp[j\theta(t)], \quad (23)$$

from where, using (20), we can obtain

$$\begin{aligned}
g(t) &= \text{Re}[\tilde{g}(t)\exp(j2\pi f_c t)] \\
&= \text{Re}\{a(t)\exp[j\theta(t)]\exp(j2\pi f_c t)\} \\
&= a(t)\cos[2\pi f_c t + \theta(t)]
\end{aligned} \tag{24}$$

In (24), $a(t) = |\tilde{g}(t)|$ is the envelope of the band-pass signal $g(t)$, or the amplitude modulated component of $g(t)$, and $\theta(t)$ is its phase, or the phase-modulated component of $g(t)$. This result will also be used later as a means for understanding the MSK signal generation.

Taking the Fourier transform of $g(t)$ we know to obtain its frequency content. If $g(t)$ is a voltage signal, then the magnitude of its Fourier transform will result in a, say, “voltage spectral density”. Then, using (20) we get:

$$G(f) = \mathfrak{F}\{g(t)\} = \int_{-\infty}^{\infty} \left\{ \text{Re}[\tilde{g}(t)e^{j2\pi f_c t}] \right\} e^{-j2\pi f t} dt \tag{25}$$

Using the identity $\text{Re}[C] = \frac{1}{2}[C + C^*]$ in (25), and applying the Fourier transform properties $x^*(t) \Leftrightarrow X^*(-f)$ and $x(t)\exp(j2\pi f_c t) \Leftrightarrow X(f - f_c)$, we obtain:

$$\begin{aligned}
G(f) &= \frac{1}{2} \int_{-\infty}^{\infty} \left[\tilde{g}(t)e^{j2\pi f_c t} + \tilde{g}^*(t)e^{-j2\pi f_c t} \right] e^{-j2\pi f t} dt \\
&= \frac{1}{2} \left[\tilde{G}(f - f_c) + \tilde{G}^*(-f - f_c) \right]
\end{aligned} \tag{26}$$

If $g(t)$ is a sample function of an stationary random process $G(t)$, it has infinity energy and, hence, its Fourier transform does not exist. In this case the spectral content of $G(t)$ is given by its *power spectral density* (PSD), which is obtained from the Fourier transform of the auto-correlation function $R_G(\tau)$ of the random process, as follows [8, p. 67]:

$$S(f) = \int_{-\infty}^{\infty} R_G(\tau)\exp(-j2\pi f \tau)d\tau \tag{27}$$

The PSD for a stationary random process can also be estimated through [7, p. 51]:

$$S(f) = \lim_{\Pi \rightarrow \infty} \frac{1}{\Pi} E \left[|G_{\Pi}(f)|^2 \right] \tag{28}$$

where $G_{\Pi}(f)$ is the Fourier transform obtained from the sample process $g_{\Pi}(t)$, which is $g(t)$ truncated from $-\Pi/2$ to $\Pi/2$. The function $|G_{\Pi}(f)|^2$ is called the *energy spectral density* of the energy signal $g_{\Pi}(t)$. If the signal is deterministic, (28) can also be used, without the expectation operation [11, p. 31].

However, if the Fourier transform $G(f)$ exists and is exact, according to which was stated before equation (25) $S(f)$ can be simply determined by the squared-modulus of $G(f)$, that is,

$$S(f) = |G(f)|^2 = \frac{1}{4} \left[\tilde{G}(f - f_c) + \tilde{G}^*(-f - f_c) \right]^2 \tag{29}$$

Using a simplified notation, and the fact that $|C|^2 = CC^*$, we can rewrite (31) as follows:

$$\begin{aligned}
S(f) &= \frac{1}{4} \left[|X(f) + X^*(-f)|^2 \right] \\
&= \frac{1}{4} \left\{ [X(f) + X^*(-f)][X^*(f) + X(-f)] \right\} \\
&= \frac{1}{4} \left[X(f)X^*(f) + X^*(-f)X(-f) \right. \\
&\quad \left. + X(f)X(-f) + X^*(-f)X^*(f) \right] \\
&= \frac{1}{4} \left[|X(f)|^2 + |X(-f)|^2 \right. \\
&\quad \left. + X(f)X(-f) + X^*(-f)X^*(f) \right]
\end{aligned} \tag{30}$$

By recognizing that $X(f)$ and $X(-f)$ are band-limited, band-pass signals, the products $X(f)X(-f)$ and $X^*(f)X^*(-f)$ in (30) vanish to zero. Going back to the normal notation, we get:

$$\begin{aligned}
S(f) &= \frac{1}{4} \left[|\tilde{G}(f - f_c)|^2 + |\tilde{G}(-f - f_c)|^2 \right] \\
&= \frac{1}{4} \left[S_B(f - f_c) + S_B(-f - f_c) \right]
\end{aligned} \tag{31}$$

Equation (31) states that we can easily obtain the power spectral density $S(f)$ of a band-pass signal by translating the power spectral density $S_B(f)$ of the low-pass equivalent, and its mirror image, to the frequencies f_c and $-f_c$, respectively, and multiplying the result by $1/4$.

C. Minimum frequency separation for coherent detection

It may be somewhat obvious for some readers that MSK is a form of orthogonal frequency shift keying modulation, but our aim in this subsection is to give reasons for the term *minimum* in the Minimum Shift Keying nomenclature.

To be coherently orthogonal in the signaling interval T , two cosine functions with different frequencies must satisfy

$$\int_0^T \cos(2\pi f_1 t)\cos(2\pi f_2 t)dt = 0 \tag{32}$$

Using the identity $\cos\alpha\cos\beta = \frac{1}{2}[\cos(\alpha - \beta) + \cos(\alpha + \beta)]$ in the expression above we obtain:

$$\frac{1}{2} \int_0^T \cos[2\pi(f_1 - f_2)t]dt + \frac{1}{2} \int_0^T \cos[2\pi(f_1 + f_2)t]dt = 0 \tag{33}$$

from where, after some manipulations, we get:

$$\frac{\sin[2\pi(f_1 - f_2)T]}{4\pi(f_1 - f_2)} + \frac{\sin[2\pi(f_1 + f_2)T]}{4\pi(f_1 + f_2)} = 0 \tag{34}$$

Since for practical purposes the sum $f_1 + f_2 \gg 1$, the second term in the left-hand side of (34) is approximately zero, which results in

$$\begin{aligned}
\sin[2\pi(f_1 - f_2)T] &= 0 \\
\Rightarrow (f_1 - f_2) &= \frac{k}{2T}, k \text{ inteiro}
\end{aligned} \tag{35}$$

Then, the minimum frequency separation between tones for an orthogonal FSK with coherent detection is

$$(f_1 - f_2) = \frac{1}{2T}, \quad (36)$$

which justifies the name Minimum Shift Keying for the MSK modulation.

III. MSK SIGNAL GENERATION AND DETECTION

In this section, the MSK signal generation and detection are analyzed in detail, based first on a complex representation approach, and then, based on a signal-space representation approach. The MSK power spectral density is also considered. However, first we introduce the basics about the MSK and the conventional binary FSK modulation. At the end of the section these modulations are revisited, aiming at establishing their similarities and differences from the design of the transmitter and the receiver perspective.

The receiver structures and performances considered in this section assume that the system operates on an Additive White Gaussian Noise (AWGN) channel.

A. MSK and conventional binary FSK

A continuous-phase, frequency-shift keying (CPFSK) signal can be described as a phase-modulated signal using (24), as shown by:

$$s(t) = \sqrt{\frac{2E_b}{T_b}} \cos[2\pi f_c t + \theta(t)] \quad (37)$$

where E_b is the average energy per bit and T_b is the bit duration.

The time derivative of the phase evolution $\theta(t)$ in (37) gives rise to the CPFSK instantaneous angular frequency shift. Then, in a given bit interval $\theta(t)$ increases or decreases linearly, depending on the desired transmitted tone, as described by:

$$\theta(t) = \theta(0) \pm \frac{\pi h}{T_b} t, \quad 0 \leq t \leq T_b \quad (38)$$

where $\theta(0)$ accounts for the accumulated phase history until instant $t = 0$ and h is a measure of the frequency deviation. If $h = 1$ we have the conventional form of binary FSK modulation, also known as Sunde's FSK [6] [7, p. 381], in which the tone separation is obtained from (38) as $1/T_b$ Hz.

Generalizing (38), at any time instant the phase evolution can be determined by

$$\theta(t) = \theta(0) + \frac{\pi h}{T_b} \int_0^t b(t) dt \quad (39)$$

where $b(t) \in \{\pm 1\}$ is the waveform related to the information sequence, such that a -1 represents a bit 0 and a $+1$ represents a bit 1.

The modulated signal described by (37) and (39) can be generated by means of a continuous-phase VCO (voltage controlled oscillator) having $b(t)$ as its input, and configured with center frequency f_c Hz and gain $h/(2T_b)$ Hz/volt.

Example 1 - Suppose we want to transmit the information sequence [1 0 1 0 0 1 1]. Following (39), with $h = 1$, we shall have the phase evolution illustrated in Figure 3. Also in Figure 3 are plotted the waveform $b(t)$ and the resultant FSK modulated signal $s(t)$ for $f_c = 2/T_b$ Hz. The resultant tones are then at frequencies $f_1 = 5/(2T_b)$ Hz and $f_2 = 3/(2T_b)$ Hz.

A careful look at Figure 3 shows that phase transitions from one bit to the next lead to the same value, using modulo 2π algebra (a phase transition of $+\pi$ is equal to a phase transition of $-\pi$, modulo 2π). Then, the receiver is not able to explore any phase information in the conventional Sunde's FSK modulation.

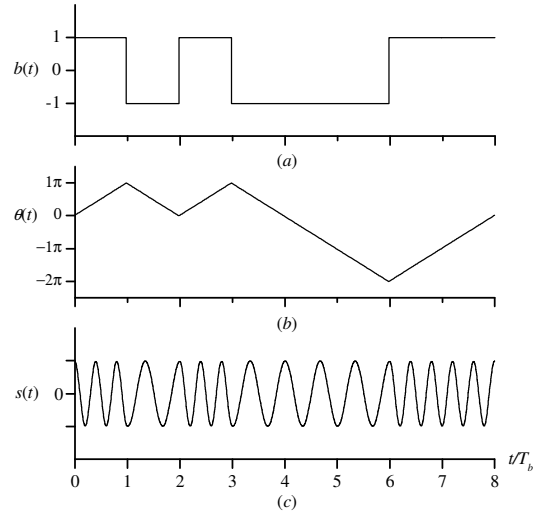


Figure 3. Information sequence (a), phase evolution (b) and modulated signal (c) for the Sunde's FSK modulation.

Now, let us make $h = 1/2$ in (39). In this case we have the minimum tone separation of $1/(2T_b)$ Hz, and, through (37), we shall generate an MSK signal.

Example 2 - Suppose again that we want to transmit the information sequence [1 0 1 0 0 1 1]. According to (39), with $h = 1/2$, we shall have the phase evolution shown in Figure 4. The waveform $b(t)$ and the resultant MSK modulated signal $s(t)$ for $f_c = 1/T_b$ Hz are also plotted. The resultant tones are at frequencies $f_1 = 5/(4T_b)$ Hz and $f_2 = 3/(4T_b)$ Hz.

As can be noticed from Figure 4, phase transitions from a bit to the next one lead to different values, modulo 2π . Then, it is possible to explore some phase information with the MSK modulation. This is indeed the motivation for the use of MSK: the receiver can explore phase transitions in order to benefit from this additional information to improve performance.

B. MSK signal generation and detection from the complex representation approach

The generation of $s(t)$ through (37) and (39), though straightforward from the implementation point of view, brings no or little insight on how the receiver can be constructed in order to explore the phase information in the modulated signal. Then we are forced to obtain alternative mathematical models for representing $s(t)$.

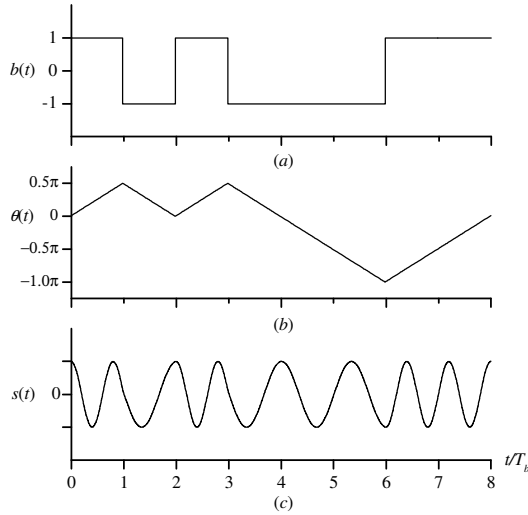


Figure 4. Information sequence (a), phase evolution (b) and modulated signal (c) for the MSK modulation.

To simplify matters, consider initially only the first bit interval. Using $h = 1/2$ in (38) and the identity $\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta$, we can rewrite (37) as follows:

$$\begin{aligned} s(t) &= \sqrt{\frac{2E_b}{T_b}} \cos\left[\theta(0) \pm \frac{\pi}{2T_b}t\right] \cos(2\pi f_c t) \\ &\quad - \sqrt{\frac{2E_b}{T_b}} \sin\left[\theta(0) \pm \frac{\pi}{2T_b}t\right] \sin(2\pi f_c t), \quad 0 \leq t \leq T_b \quad (40) \\ &= s_I(t) \cos(2\pi f_c t) - s_Q(t) \sin(2\pi f_c t) \end{aligned}$$

Making use of (22) and applying again the identity $\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta$ to the in-phase component of $s(t)$, and, without loss of generality, assuming $\theta(0) = 0$, we get

$$\begin{aligned} s_I(t) &= \sqrt{\frac{2E_b}{T_b}} \cos\left[\pm \frac{\pi}{2T_b}t\right], \quad -T_b \leq t \leq T_b \quad (41) \\ &= \pm \sqrt{\frac{2E_b}{T_b}} \cos\left(\frac{\pi}{2T_b}t\right) \end{aligned}$$

Since $\theta(0) = 0$, before $t = 0$ the phase evolution was a positive or negative slope going towards zero, depending on the previous bit. Then, the result in (41) is an increasing cosine function from $-T_b$ to 0. Thus, $s_I(t)$ can be interpreted as a half-cycle cosine function from the whole interval $(-T_b, T_b]$.

Similarly, the quadrature component of $s(t)$ can be written as follows:

$$s_Q(t) = \pm \sqrt{\frac{2E_b}{T_b}} \sin\left(\frac{\pi}{2T_b}t\right), \quad 0 \leq t \leq 2T_b \quad (42)$$

where we have made use of $\theta(0) = 0$ and of the identity $\sin(\alpha \pm \beta) = \sin\alpha\cos\beta \pm \cos\alpha\sin\beta$. We have also made use of the relation $\cos[\theta(0)] = \cos[\theta(T_b) \mp \pi/2] = \pm \sin[\theta(T_b)] = \pm 1$.

Since $\theta(T_b) = \pm\pi/2$, depending on the information bit during the interval $(0, T_b]$, we shall have $\sin[\theta(t)]$ going towards zero during the interval T_b to $2T_b$, regardless the information bit during this interval. Thus, $s_Q(t)$ can be viewed as a half-cycle sine function from the whole interval $(0, 2T_b]$, the polarity of which depending on the information bit during the interval $[0, T_b]$.

Using the results (41) and (42) in (40), we obtain:

$$\begin{aligned} s(t) &= \pm \sqrt{\frac{2E_b}{T_b}} \cos\left(\frac{\pi}{2T_b}t\right) \cos(2\pi f_c t) \\ &\quad \mp \sqrt{\frac{2E_b}{T_b}} \sin\left(\frac{\pi}{2T_b}t\right) \sin(2\pi f_c t) \quad (43) \end{aligned}$$

where the polarity of both terms in a given bit interval are not necessarily the same.

Following [4, p. 18], we can rewrite (43) as:

$$\begin{aligned} s(t) &= \sqrt{\frac{2}{T_b}} a_I(t) \cos\left(\frac{\pi}{2T_b}t\right) \cos(2\pi f_c t) \\ &\quad - \sqrt{\frac{2}{T_b}} a_Q(t) \sin\left(\frac{\pi}{2T_b}t\right) \sin(2\pi f_c t) \quad (44) \end{aligned}$$

where we have defined $a_I(t)$ and $a_Q(t)$ as random sequences of rectangular pulses with amplitudes $\pm\sqrt{E_b}$ and duration $2T_b$ seconds. These sequences are associated to the polarities of the half-cycle cosine and sine functions as follows: if $a_I(t)$ is positive, $s_I(t)$ follows the function $\cos\{\lceil\pi/(2T_b)\rceil t\}$; if $a_I(t)$ is negative, $s_I(t)$ corresponds to $-\cos\{\lceil\pi/(2T_b)\rceil t\}$. The same happens with $s_Q(t)$: if $a_Q(t)$ is positive, $s_Q(t)$ follows the function $\sin\{\lceil\pi/(2T_b)\rceil t\}$; if $a_Q(t)$ is negative, $s_Q(t)$ corresponds to $-\sin\{\lceil\pi/(2T_b)\rceil t\}$.

From the above discussion we can conclude that, depending on the information bit to be transmitted, the in-phase and quadrature components of $s(t)$ can change their polarities each $2T_b$ seconds, and that the half-cycle cosine and sine functions are offset from each other by T_b seconds. However, we are not still able to easily obtain the information sequence responsible for generating a given sequence of polarities. This would demand us to come back to the general analysis presented in Section III-A, specifically to equation (37), thus making difficult the visualization of the implementation issues for the MSK modem.

Then, for the time being we assume a given sequence of pulses for $s_I(t)$ and $s_Q(t)$, and later we determine the information sequence based on the analysis of this assumption. A general rule will arise from this analysis.

Example 3 – In a 8-bit interval, let $s_I(t)$ and $s_Q(t)$ assume the sequence of half-cycle cosine and sine pulses shown in Figure 5. For reference, in this figure the functions $\cos\{\lceil\pi/(2T_b)\rceil t\}$ and $-\sin\{\lceil\pi/(2T_b)\rceil t\}$ are also plotted, in dashed lines, and are given the polarities of the waveforms $a_I(t)$ and $a_Q(t)$. Combining the waveforms in Figure 5 according to (40), we get the results in Figure 6. In this figure the waveforms $s_I(t)$ and $s_Q(t)$ are also plotted, in dashed lines. The carrier

frequency in this example is $f_c = 1/T_b$ Hz. The resultant tones are then at frequencies $f_1 = 5/(4T_b)$ Hz and $f_2 = 3/(4T_b)$ Hz.

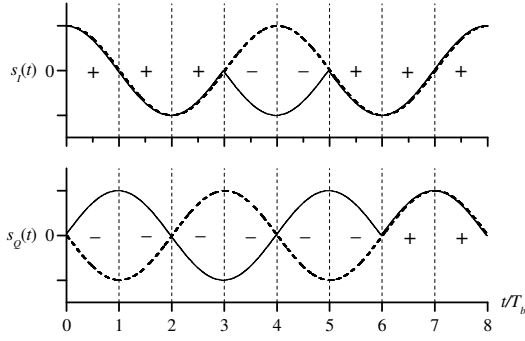


Figure 5. Generation of the MSK signal: base-band in-phase and quadrature components.

Observing the modulated signal $s(t)$ in Figure 6 we can notice that, if a bit 1 is associated to the tone of greater frequency, the corresponding modulating sequence should be $\mathbf{d} = [1\ 1\ 1\ 0\ 0\ 1\ 0\ 0]$. Let us now define a new sequence \mathbf{i} in which the exclusive-or (XOR) operation between a given bit and the previous one results in a bit of the sequence \mathbf{d} . This new sequence is $\mathbf{i} = [1\ 0\ 1\ 0\ 0\ 0\ 1\ 1\ 1]$. Sequence \mathbf{d} can be seen as a *differentially decoded* version of \mathbf{i} . Additionally, suppose that the sequence \mathbf{i} is parallelized to form the sequences of odd and even symbols of duration $2T_b$, $\mathbf{i}_o = [1\ 1\ 0\ 1\ 1]$ and $\mathbf{i}_e = [0\ 0\ 0\ 1]$, respectively. Now, suppose that each symbol of these sequences is converted to $\pm\sqrt{E_b}$. The great achievement here is that these new parallel sequences, if they are off-set to each other T_b seconds, are exactly the waveforms $a_I(t)$ and $a_Q(t)$.

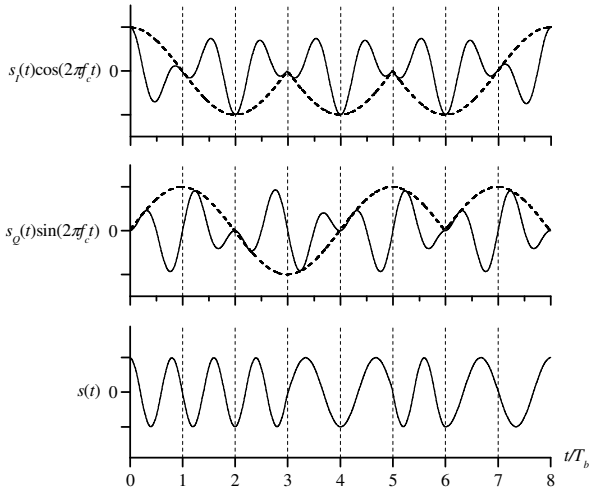


Figure 6. Generation of the MSK signal: modulated in-phase and quadrature components and the resultant MSK signal.

We then conclude that the MSK signal can be generated through (44), where the waveforms $a_I(t)$ and $a_Q(t)$ are the serial-to-parallel (S/P) converted version of the information sequence, with bit 1 converted to $+\sqrt{E_b}$ and bit 0 converted to $-\sqrt{E_b}$. Additionally, the sequence $a_Q(t)$ has to be offset T_b seconds from $a_I(t)$, before they multiply the corresponding

remaining terms in (44). Figure 7 illustrates the structure of the MSK modulator constructed according to complex representation approach just described.

The MSK signal just analyzed can also be generated by means of a VCO configured with center frequency f_c Hz and gain $1/(4T_b)$ Hz/volt. However, since the frequency shifts in the modulated signal do not directly correspond to the information sequence, the input of the VCO must be the differentially decoded version of this information sequence, converted to $\{\pm 1\}$.

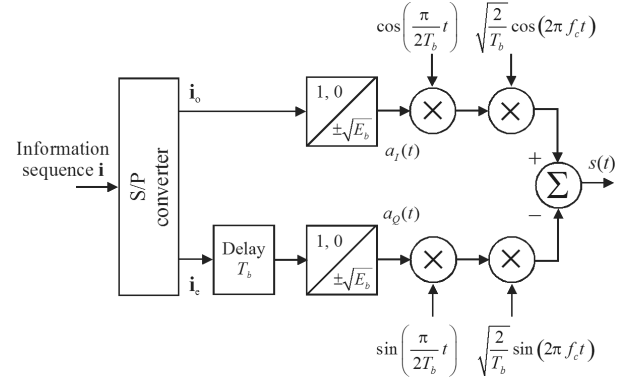


Figure 7. MSK modulator constructed according to the complex representation approach.

Some authors claim that the MSK modulation is a special form of OQPSK (or SQPSK) modulation where the pulse shaping are half-cycle cosine and sine functions, instead of the rectangular shaping functions used in OQPSK. However, in despite of being true, this statement must be carefully interpreted. From (43) we can see that, in fact, the shapes of the pulses that modulate the quadrature carriers are half-cycle cosine and sine functions. Nevertheless, they are not a simple reshaping of the waveforms $a_I(t)$ and $a_Q(t)$. Before modulating the quadrature carriers, $a_I(t)$ and $a_Q(t)$ are modified by the polarities of the waveforms $\cos\{\lceil\lceil\pi/(2T_b)\rceil t\rceil\}$ and $\sin\{\lceil\lceil\pi/(2T_b)\rceil t\rceil\}$ in each $2T_b$ interval. But we can make a small modification in the above structure to implement the MSK modulation in the same way we implement an OQPSK modulator, the unique difference being the shape of the pulses that modulate the quadrature carriers. We just have to use the modulus $|\cos\{\lceil\lceil\pi/(2T_b)\rceil t\rceil\}|$ and $|\sin\{\lceil\lceil\pi/(2T_b)\rceil t\rceil\}|$ in (44). The resultant structure is shown in Figure 8. In this figure we have used additional simplifications to make the MSK modulator structure closer to a more practical one: the quadrature carriers were generated from a single oscillator and the pulse-shaping functions were implemented via low-pass filters with identical impulse responses given by

$$h(t) = \begin{cases} \sin\left(\frac{\pi}{2T_b}t\right), & 0 \leq t \leq 2T_b \\ 0, & \text{otherwise} \end{cases} \quad (45)$$

The inputs to these filters are PAM (Pulse Amplitude Modulation) sequences having very short durations (approximating unit impulses) and amplitudes of $+\sqrt{E_b}$.

We can recall that at the beginning of Section III-B we have made the assumption that $\theta(0) = 0$. This assumption was adopted only to facilitate the mathematical description of the MSK modulation. In fact, from an implementation perspective, any initial phase is allowed for the quadrature carriers. However, regardless of this initial phase, the designer must only guarantee the correct phase alignment among the quadrature carriers, the pulse shaping functions and the sequences $a_I(t)$ and $a_Q(t)$.

In the light of the similarities between the MSK and OQPSK modulations, we are now able to understand possible structures for the MSK demodulator. We know that a conventional QPSK modulator can be interpreted as two BPSK (Binary Phase-Shift Keying) modulators, each of them making use of one of the two quadrature carriers. Then, the QPSK demodulator can be implemented as two independent BPSK demodulators. The decisions made by each of these demodulators are parallel-to-serial (P/S) converted to form the estimate of the transmitted bit sequence. The OQPSK demodulator follows the same rule, with the difference that one of the estimated parallel sequences is offset T_b seconds from the other. Then, before P/S conversion these sequences must be aligned in time.

If we use $|\cos\{\lceil\pi/(2T_b)\rceil t\}|$ and $|\sin\{\lceil\pi/(2T_b)\rceil t\}|$ in the modulator of Figure 7 or adopt a more practical solution, as shown in Figure 8, we must use $\cos\{\lceil\pi/(2T_b)\rceil t\}$ and $\sin\{\lceil\pi/(2T_b)\rceil t\}$ in the demodulator in Figure 9. In this case the correspondence between the MSK and the OQPSK demodulators exists, the unique difference being the shape of the pulses that multiply the quadrature carriers. If we do not apply the modulus operation at the modulator we have its complex representation approach realization and, in this case, we just have to use the original $\cos\{\lceil\pi/(2T_b)\rceil t\}$ and $\sin\{\lceil\pi/(2T_b)\rceil t\}$ in the demodulator of Figure 9.

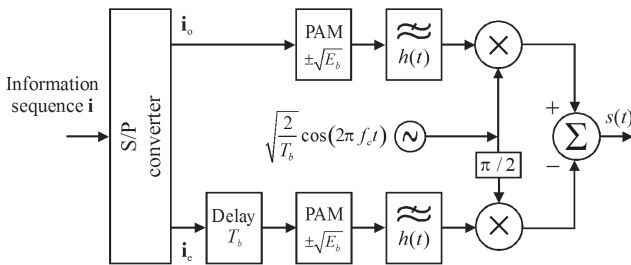


Figure 8. A more practical MSK modulator constructed according to the equivalence with the OQPSK modulation.

The received signal in Figure 9 is coherently correlated, in one arm of the receiver, with the result of the multiplication between the in-phase carrier and the shaping function $\cos\{\lceil\pi/(2T_b)\rceil t\}$. In the other arm, the received signal is correlated with the result of the multiplication between the quadrature carrier and the shaping function $-\sin\{\lceil\pi/(2T_b)\rceil t\}$. These correlations are made in a $2T_b$ seconds interval, reflecting the duration of the half-cycle cosine and sine

functions, and are time-aligned with these functions. The estimated sequences \hat{i}_e and \hat{i}_o are then time-aligned and P/S converted to form the estimate of the transmitted sequence, \hat{i} .

If, for some reason, it is necessary to represent a bit 1 in the sequence \mathbf{d} by the tone of lower frequency, the only thing we have to do is to invert the minus signal in the summation block in Figure 7 or Figure 8, and invert the minus signal in the bottom multiplier block in Figure 9.

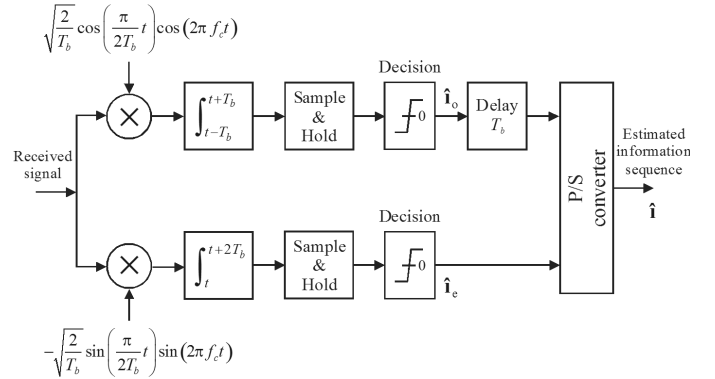


Figure 9. MSK demodulator constructed according to the complex representation approach.

C. MSK signal generation and detection from the signal-space representation approach

We are now able to determine the orthonormal base-functions responsible for generating the MSK signal. Recalling that we are talking about a binary orthogonal signaling, the base-functions can be directly obtained from (44) as follows:

$$\phi_1(t) = \sqrt{\frac{2}{T_b}} \cos\left(\frac{\pi}{2T_b}t\right) \cos(2\pi f_c t) \quad (46)$$

$$\phi_2(t) = \sqrt{\frac{2}{T_b}} \sin\left(\frac{\pi}{2T_b}t\right) \sin(2\pi f_c t) \quad (47)$$

These base-functions, differently from what is stated in [7, p. 390], are defined for any interval, not only from 0 to T_b .

Comparing (44) with (46) and (47) we readily see that the MSK signal vectors are determined by the amplitudes of the waveforms $a_I(t)$ and $a_Q(t)$ defined in (44), in each bit interval:

$$\mathbf{s}_i = \begin{bmatrix} s_{i1} \\ s_{i2} \end{bmatrix} = \begin{bmatrix} \pm\sqrt{E_b} \\ \pm\sqrt{E_b} \end{bmatrix}, \quad i = 1, 2, 3, 4 \quad (48)$$

Then, as shown in Figure 10, the signal-space diagram for the MSK modulation comprises four signal vectors, despite of MSK be a binary modulation. The mapping between these vectors and the information bits is determined via the differentially decoded version of the information bits. The following example is meant to clarify these statements.

Example 4 – Let the sequence of signal vector polarities be $[+ -]$, $[+ -]$, $[+ -]$, $[- -]$, $[- -]$, $[+ -]$, $[+ +]$ and $[+ +]$, generated on a bit-by-bit basis. In this sequence, the polarities on the left refer to s_{i1} , and those on the right refer to s_{i2} . These

polarities are the same as those considered in Figure 5 and, as we already know from Example 3, they are associated to the information sequence $\mathbf{i} = [1\ 0\ 1\ 0\ 0\ 0\ 1\ 1\ 1]$ and to its differentially decoded version $\mathbf{d} = [1\ 1\ 1\ 0\ 0\ 1\ 0\ 0]$. From this example it is possible to draw the mapping between the signal vectors and the differentially decoded version of the information bits, as shown in Table I.

TABLE I – MAPPING BETWEEN THE MSK SIGNAL VECTORS AND THE DIFFERENTIALLY DECODED VERSION OF THE INFORMATION BITS

i	Bits	Signal vector coordinates	
		s_{i1}	s_{i2}
1	0	$+\sqrt{E_b}$	$+\sqrt{E_b}$
2	1	$+\sqrt{E_b}$	$-\sqrt{E_b}$
3	0	$-\sqrt{E_b}$	$-\sqrt{E_b}$
4	1	$-\sqrt{E_b}$	$+\sqrt{E_b}$

Since MSK is a continuous phase modulation, no abrupt phase transition occurs when a symbol changes. The circumference in Figure 10 illustrates this smooth phase transitions between any pair of symbols. They can be observed in a x - y plot, with $s_I(t)$ applied to the x -axis and $s_Q(t)$ applied to the y -axis (see Figure 5).

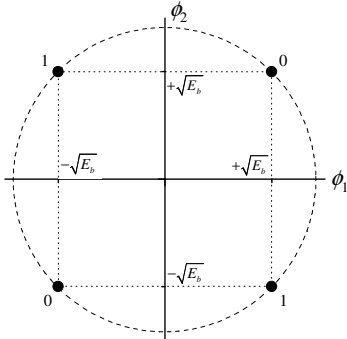


Figure 10. MSK constellation.

Observing (46) and (47) we see that the base-functions $\phi_1(t)$ and $\phi_2(t)$ correspond to the modulation of the quadrature carriers by the waveforms $\cos\{\lceil\pi(2T_b)\rceil t\}$ and $\sin\{\lceil\pi(2T_b)\rceil t\}$, respectively. Comparing (46) and (47) with (44), we see that the base-function $\phi_1(t)$ are multiplied by $a_I(t)$, the base-function $\phi_2(t)$ is multiplied by $a_Q(t)$, and the results are added to form the MSK signal $s(t)$. Figure 11 illustrates the generation of the MSK signal from this signal-space representation approach. The signal vector polarities associated to the waveforms $a_I(t)$ and $a_Q(t)$ are the same as those used in Example 4.

As we did with the complex representation approach, now we shall construct the modulator structure based on the signal-space representation. As a matter of fact, if we group together the two upper mixers and group together the two lower mixers in Figure 7 this job is already done. But we shall manipulate the base-function expressions to get an alternative structure. First, let us expand $\phi_1(t)$ using the identity $\cos\alpha\cos\beta = \frac{1}{2}[\cos(\alpha - \beta) + \cos(\alpha + \beta)]$:

$$\begin{aligned}\phi_1(t) &= \sqrt{\frac{2}{T_b}} \cos\left(\frac{\pi}{2T_b}t\right) \cos(2\pi f_c t) \\ &= \sqrt{\frac{1}{2T_b}} \cos(2\pi f_2 t) + \sqrt{\frac{1}{2T_b}} \cos(2\pi f_1 t)\end{aligned}\quad (49)$$

Now, let us expand $\phi_2(t)$ using the identity $\sin\alpha\sin\beta = \frac{1}{2}[\cos(\alpha - \beta) - \cos(\alpha + \beta)]$:

$$\begin{aligned}\phi_2(t) &= \sqrt{\frac{2}{T_b}} \sin\left(\frac{\pi}{2T_b}t\right) \sin(2\pi f_c t) \\ &= \sqrt{\frac{1}{2T_b}} \cos(2\pi f_2 t) - \sqrt{\frac{1}{2T_b}} \cos(2\pi f_1 t)\end{aligned}\quad (50)$$

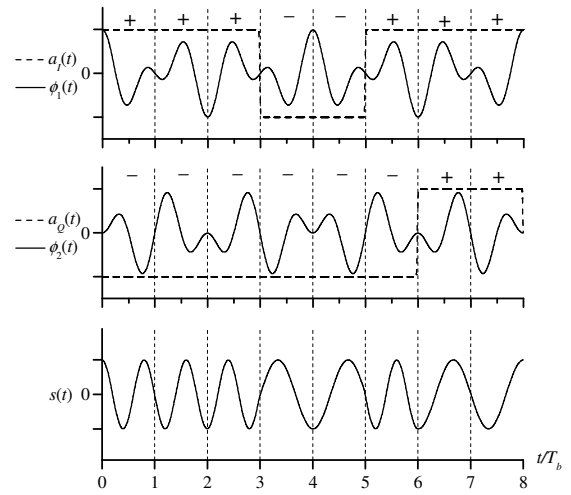


Figure 11. Generation of the MSK signal: base-functions, coefficients and the resultant MSK signal.

Figure 12 shows the MSK modulator constructed according to the interpretation of expressions (49) and (50). The two cosine functions are multiplied to generate the tones with frequencies f_1 and f_2 , according to (49). Each of these tones is selected through the band-pass filters shown in this figure, and the results are combined according to (49) and (50) to generate the base-functions. Finally, these base-functions are multiplied by the corresponding waveforms $a_I(t)$ and $a_Q(t)$ and the results are added-up to form the MSK signal. The approach at hand can also consider the demodulator shown in Figure 9, where we readily identify the use of the base-functions $\phi_1(t)$ and $\phi_2(t)$ feeding the correlators.

We can see that, operating in different ways with the mathematical model of the MSK signal, it is possible to construct different, but equivalent structures. More structures would be possible if an alternative mathematical model were adopted. These comments are also valid to the construction of the MSK demodulator. In [11, pp. 299-307] the reader can find several forms for the implementation of an MSK modem, along with different approaches on its construction.

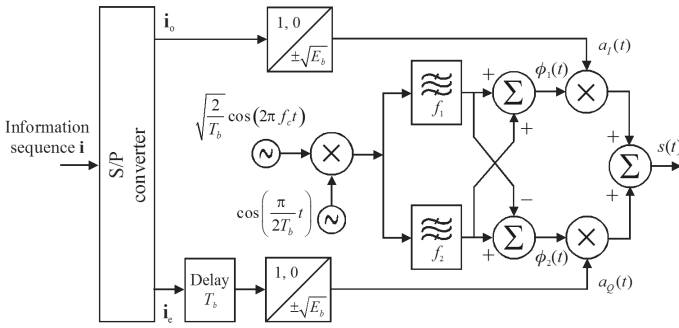


Figure 12. MSK modulator constructed according to the signal-space representation approach.

D. Bit error probability for the MSK modulation

We can see through Figure 8 and equation (44) that the modulator transmits two independent sequences using two quadrature carriers, and through Figure 9 we can see that the demodulator detects these sequences independently. Consequently, we can state that the modulator can be interpreted as formed by two independent BPSK-like modulators and that the demodulator can be interpreted as formed by two independent BPSK-like demodulators. The difference to the conventional BPSK modulator and demodulator is the presence of half-cycle sine and cosine pulse-shaping functions. The energy per symbol for each of these two component BPSK modulators is easily found to be

$$\begin{aligned} \xi &= \frac{2E_b}{T_b} \int_0^{2T_b} \cos^2\left(\frac{\pi}{2T_b}t\right) \cos^2(2\pi f_c t) dt \\ &= \frac{2E_b}{T_b} \int_0^{2T_b} \sin^2\left(\frac{\pi}{2T_b}t\right) \sin^2(2\pi f_c t) dt = E_b \end{aligned} \quad (51)$$

where, for simplification purposes, we have adopted the carrier frequency f_c as an integer multiple of $1/(2T_b)$. The energy per MSK symbol is the sum of the symbol energies in the quadrature modulated carriers, that is $E = 2E_b$, a value that can also be obtained from the constellation in Figure 10.

Confusions may arise here: the duration of one bit is of course T_b seconds, and we must make the bit decisions in a bit-by-bit basis. But the phase information at the MSK receiver is explored in $2T_b$ seconds intervals, so that the effective energy collected by this receiver corresponds to observations made during intervals of $2T_b$ seconds.

From the above discussion we can conclude that the bit error probability for the MSK modulation on the AWGN channel, considering equally-likely bits, can be determined by the average of the bit error probabilities for the two component BPSK detectors [8, p. 271], which results in:

$$P_b = \frac{1}{2} \operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right), \quad (52)$$

where N_0 is the AWGN power spectral density and $\operatorname{erfc}(u)$ is the complimentary error function of the argument. This result shows that the performance of the MSK modulation is the same as the performance of the BPSK and QPSK modulations,

and is 3 dB more energy-efficient than the conventional BFSK with coherent detection [7, p. 418].

E. MSK signal generation and detection from a conventional Sunde's FSK approach

Suppose now that we aim at generating an MSK signal using the conventional FSK approach, but with the minimum tone separation $(f_1 - f_2) = 1/(2T_b)$ Hz. The modulator would appear like in Figure 13. This form of FSK signal generation guarantees phase continuity only if the tone separation is a multiple of $1/T_b$ and the carrier frequency is a multiple of $1/(2T_b)$. Then, the modulated signal in Figure 13 will show phase discontinuities, which does not correspond to an MSK signal. MSK and binary FSK signals are the same if they are generated according to (37) and (39), using $h = 1/2$.

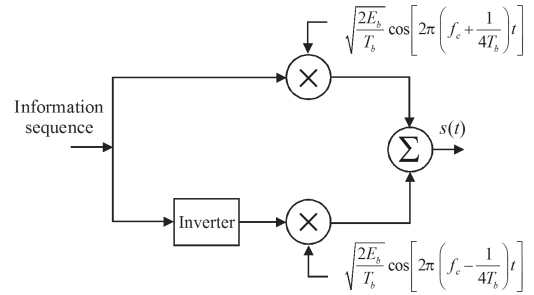


Figure 13. A try for generating an MSK signal from the conventional binary FSK implementation approach.

Now, following [9], suppose that we want to detect an MSK signal using a conventional coherent FSK demodulator. We would be tempted to think that it is just necessary to correlate the received signal with base-functions formed by the cosine tones with frequencies f_1 and f_2 , during T_b seconds intervals, and that the decision would be made in favor of the greatest correlator output. However, the phase continuity and phase dependency imposed by the MSK signal construction do not permit the use of the above approach. This is illustrated in Figure 14, where we have plotted an MSK signal and the cosine base-functions with frequencies f_1 and f_2 separated by $1/(2T_b)$ Hz. Observe that, in several intervals, there are no phase coherence between the modulated signal and the base-functions with the same frequency, a behavior that would lead to detection errors.

Let us elaborate a little bit more on this issue. From Figure 14 we can see that when no phase coherence occurs, the MSK signal is at 180° out of phase from the corresponding base-function. Then, by comparing the magnitudes of the correlators outputs we are still able to make correct decisions. But we cannot forget that, unless the MSK signal is generated directly from the realization of (37) and (39) with $h = 1/2$, the estimated bits would correspond to a differentially decoded version of the information bits. To get the estimates of the information bits we have to apply the inverse operation on the estimated bits through the exclusive OR (XOR) between a given bit and the previous XOR result (see Example 3 and the corresponding comments). However, this operation can lead to the opposite decisions, since a differentially decoded 1 can result from the information sequence 01 or 10, and a differentially decoded 0 can result from the information

sequence 00 or 11. Inserting a differential coder at the transmitter input and a differential decoder at the receiver output easily solves this ambiguity problem.

Finally, we shall have the transmitter and receiver structures shown in Figure 15.

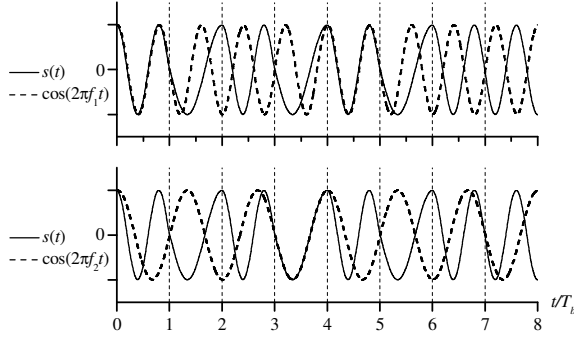


Figure 14. MSK signal $s(t)$, the $\cos(2\pi f_1 t)$ and $\cos(2\pi f_2 t)$.

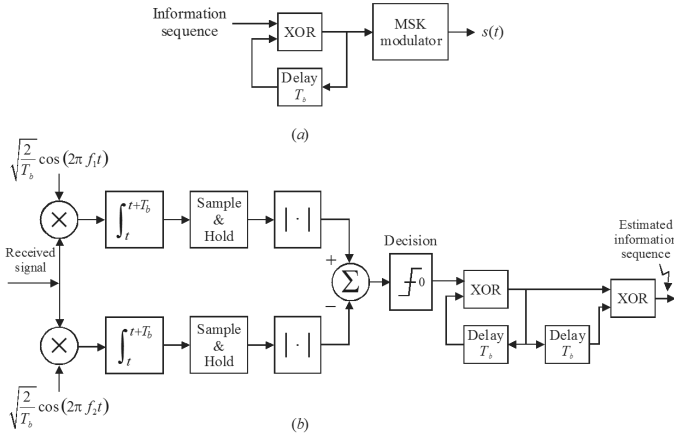


Figure 15. MSK modulator with conventional FSK detection: modified MSK transmitter (a) and detection via a modified coherent binary FSK receiver (b).

Since the receiver in Figure 15 is not exploring any phase information, we expect a worse performance as compared to the one provided by the appropriate MSK receiver. Furthermore, although the channel noise is Gaussian, the noise in the decision variable is not. Then, the analytical process for obtaining an expression for the bit error probability P_e for the receiver under investigation is quite involved and is beyond the scope of this work. Nevertheless, a numerical calculation of P_e was made and a simulation of the system in Figure 15 was carried out. Both results agreed and showed that the performance lies in between a coherently detected and a non-coherently detected binary FSK, as shown in Figure 16, and is approximately 3.05 dB worse than the P_e obtained with the MSK receiver. This is an attractive result, since the P_e curves for the coherent and the non-coherent FSK differs asymptotically in about 1 dB [7 p. 418], and we are using a transmitted signal that has the most compact spectrum among the coherent and orthogonal CPFSK modulations [9].

Using a more practical and simplified approach, the MSK modulator in Figure 15-a can be replaced by a VCO, eliminating the need for the three differential circuits used by the complete system. This alternative was also simulated and

the BER was the same as the one obtained with the simulation of the complete system depicted by Figure 15.

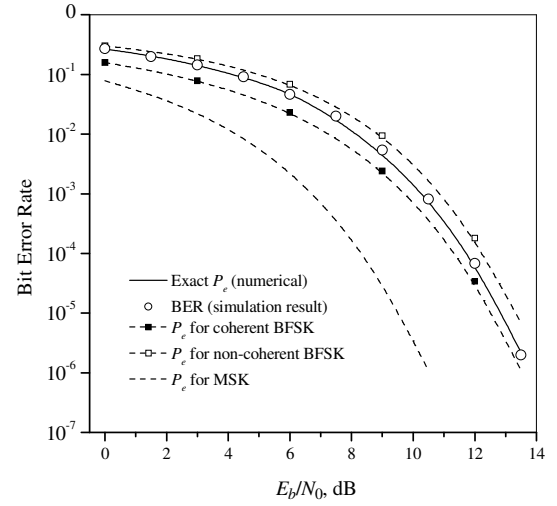


Figure 16. Performance results for MSK, coherent and non-coherent BFSK and for the system depicted in Figure 15. The channel is AWGN [9].

F. Power spectral density of the MSK signal

We saw in Section II that in order to obtain the PSD of a modulated signal, we can determine the PSD of its complex envelope representation and, using (31), convert the result to the desired PSD. According to (22), the MSK signal can be written as:

$$s(t) = s_I(t) \cos(2\pi f_c t) - s_Q(t) \sin(2\pi f_c t) \quad (53)$$

from where the complex envelope given by (21) is

$$\tilde{s}(t) = s_I(t) + js_Q(t) \quad (54)$$

For the MSK modulation, the low-pass in-phase and quadrature components in (54) are random waveforms in which the pulses with duration $2T_b$ can assume positive or negative values according to:

$$s_I(t) = \sum_k I_k p(t - 2kT_b), \quad -\infty \leq k \leq \infty \quad (55)$$

$$s_Q(t) = \sum_k Q_k p(t - 2kT_b), \quad -\infty \leq k \leq \infty$$

where $p(t)$ is the shaping pulse with half-cycle sine format:

$$p(t) = \sqrt{\frac{2E_b}{T_b}} \sin\left(\frac{\pi}{2T_b} t\right), \quad 0 \leq t \leq 2T_b, \quad (56)$$

and $\{I_k\}$ and $\{Q_k\}$ are random antipodal sequences $\in \{\pm 1\}$ associated to the odd and even information bits, respectively (see Example 3) or, equivalently, associated to the waveforms $a_I(t)$ and $a_Q(t)$ in (44).

It is a well-known result that the power spectral density of a random antipodal sequence can be determined by dividing the energy spectral density (ESD) of the shaping pulse by the pulse duration [7, p. 48] [8, p. 207]. By recalling that the ESD of a pulse is the squared-modulus of its Fourier transform, then

the PSD of $s_I(t)$, which is equal to the PSD of $s_Q(t)$, can be easily determined. Furthermore, we know that the in-phase and quadrature components of the MSK signal are independent to each other. Then, the PSD of (54) can be obtained through

$$S_B(f) = 2 \frac{|P(f)|^2}{2T_b}, \quad (57)$$

and the PSD of the MSK signal can be finally obtained using the above result in (31).

Following the procedure just described, the PSD of the base-band MSK signal in (54) can be obtained from [8, p. 214] and is given by

$$S_B(f) = \frac{32E_b}{\pi^2} \left(\frac{\cos 2\pi f T_b}{1 - 16f^2 T_b^2} \right)^2 \quad (58)$$

Equation (58) is plotted in Figure 17, along with the base-band PSD of the QPSK modulation, for comparison purposes. To draw this figure, both MSK and QPSK signals were set to the same average power.

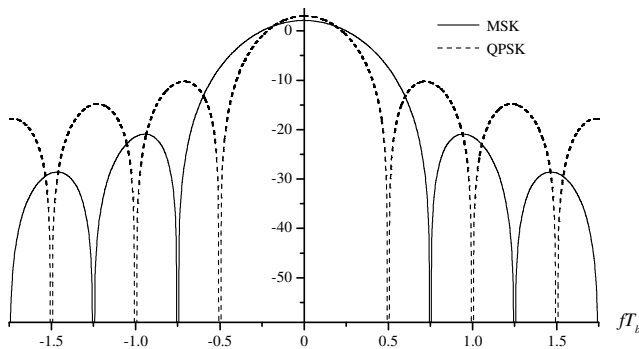


Figure 17. Normalized base-band PSD, in dBm/Hz, for the MSK and the QPSK modulations with the same average power.

It can be seen from Figure 17 that, although the main lobe of the MSK spectrum is wider than that of the QPSK one, the PSD of the MSK decreases faster with frequency. For QPSK, approximately 90% of the modulated signal power is concentrated in the main lobe. For MSK, this quantity increases to approximately 99%. This is a desired attribute of the MSK modulation, which makes it attractive due to easy filtering and, consequently, low adjacent channel interference.

Detailed and more complete considerations about the power spectral characteristics of continuous-phase modulated signals can be found in [8, pp. 209-219].

IV. FURTHER ATTRIBUTES AND USES OF THE MSK

In this section we summarize some MSK-related topics concerning additional attributes and applications of this modulation. We start by revisiting the application of the MSK in the recently-developed Blue-Ray technology [5], and as the base for implementing the GMSK modulation used, for instance, in the GSM standard [7, pp. 396-400]. In the case of the GSM standard, a Gaussian-filtered version of the information sequence is applied to an MSK modulator, resulting in the GMSK signal. This is done to increase the

spectral efficiency of the MSK modulation, with the penalty of a possibly small reduction in performance due to inter-symbol interference introduced by the Gaussian filtering process.

As mentioned at the beginning of this paper, the MSK modulation is also attractive because of its constant envelope, a characteristic that can be observed in all FSK-type modulations. Although M -PSK modulations also have constant envelopes, this is valid only if no filtering is applied to the signal. When the modulated signal is filtered before going through some non-linear distortion, such as non-linear amplification, out-of-band and in-band spurious can be generated due to envelope fluctuations that occur during abrupt phase transitions. Non-constant envelopes can also show high peak-to-average power ratios (PAPR), making it difficult the project of high dynamic range and power-efficient non-linear amplifiers. The MSK modulation, even after filtering, has low PAPR, becoming attractive in these cases.

The MSK modulation can also be viewed as a special form of coded-modulation scheme in which the phase continuity restrictions introduce some sort of redundancy and, consequently, error correction capabilities. This attribute is explored in detail in [10].

In [12], J. K. Omura, et. al apply the MSK modulation to achieve code-division multiple access (CDMA) capability in a spread spectrum system.

Finally, although MSK is usually associated to the binary case, that is, $M = 2$, its concepts are generalized to the M -ary case in [13] and [14]. A multi-amplitude, continuous-phase modulation approach is considered in [8, pp. 200-203], where the signal amplitude is allowed to vary, while the phase trajectory is constrained to be continuous. Generalized MSK is also considered in [15].

V. CONCLUSIONS

We are now armed with enough concepts to give possible answers (A) to the questions (Q) listed at the end of Section I:

Q: To which extent the MSK modulation can be regarded as a special case of the Sunde's FSK modulation? *A:* We saw that MSK is in fact a special form of FSK with the minimum tone separation for orthogonality and coherent detection. However, the MSK signal construction gives to the receiver the ability to explore phase information for performance improvement, which does not happen with the conventional FSK modulation. As we saw in Section III-E, the conventional binary FSK signal with minimum tone separation does not correspond to an MSK signal and does not exhibit phase continuity for all bit transitions.

Q: To which extent the MSK modulation can be detected as the conventional Sunde's FSK modulation? *A:* From the analysis in Section III-E we conclude that an MSK signal can be detected as a conventional binary FSK, but it is necessary to make modifications at the transmitter and at the receiver, according to the block diagram shown in Figure 15. Since this modified receiver explores no phase information, the performance will not be the same as that provided by the appropriate MSK receiver.

Q: To which extent the MSK modulation can be regarded as a special case of the SQPSK or OQPSK (Staggered or

Offset QPSK) modulation? *A:* The MSK modulation is indeed a special form of OQPSK (or SQPSK) modulation, where the pulse shaping are half-cycle cosine and sine functions instead of the rectangular shaping functions used in OQPSK. But this is not a direct interpretation of the MSK signal construction. To shown perfect equivalence with the OQPSK modulation, the MSK transmitter must be implemented according to Figure 8. The receiver structure is kept unchanged, according to the block diagram shown in Figure 9.

Q: To which extent the frequency and phase shifts of an MSK signal are related to the modulating data sequence? *A:* If the modulated signal is generated through the realization of (37) and (39), using $h = 1/2$, then there will be a direct correspondence, that is, bit 0 will be represented by the tone with frequency, say, f_2 (or vice-versa), and bit 1 will be represented by the tone with frequency f_1 (or vice-versa). However, by generating the MSK signal through the other ways shown is this tutorial, the frequency shifts will correspond to a differentially decoded version of the modulating data sequence.

Q: To which extent the phase shifts of an MSK signal can be related to the phase transition diagram on its signal-space representation? *A:* The MSK signal is constructed in a way that, besides phase continuity, it exhibits phase transitions that helps the receiver improve the detection performance. This is done because phase transitions from one bit to the next lead to different values, modulo 2π (see Figure 4). A bit one increases the phase in $\pi/2$ radians and a bit 0 decreases the phase in $\pi/2$ radians. If these bits are or are not the information bits, it depends on how the MSK signal is generated: directly via (37) or indirectly (see former question and answer). Concerning the phase shifts of an MSK signal, they cannot be directly mapped on the signal-space symbol transitions. Two reasons support this conclusion: firstly, since a given signal-space diagram can represent a base-band or a band-pass signaling, it is not always able to represent phase transitions of a modulated signal, though it can happen with some modulations, such as M -PSK and M -QAM. Secondly, discrete points in a signal space cannot represent continuous-phase signals, because the phase of the carrier is time-variant [8, pp. 199-200]. As an example, two consecutive ones correspond to the same coordinates in Figure 10, but we know that the carrier phase changes $+\pi/2$ radians from its preceding value, in a continuous way. A solution to this is to have a three-dimensional diagram with axes $s_i(t)$, $s_q(t)$ and t , in which the phase trajectory can be recorded [8, pp. 194-195]. Figure 18 illustrates this representation.

REFERENCES

- [1] R. de Buda, "Coherent demodulation of frequency-shift keying with low deviation ratio", *IEEE Trans. on Communications*, vol. COM-20, no. 3, pp. 429-436, June 1972.
- [2] M. L. Doelz and E. H. Heald, "Minimum shift data communication system", *United States Patent 2,917,417*, March 28, 1961.
- [3] S. A. Groameyer and A.L. McBride, "MSK and offset QPSK modulation", *IEEE Trans. on Communications*, August 1976.
- [4] S. Pasupathy, "Minimum Shift Keying: A Spectrally Efficient Modulation", *IEEE Communications Magazine*, vol. 17, no. 4, pp. 14-22, July 1979.

- [5] *Blu-Ray Disc Recordable Format – Part 1: Physical Specifications*. Available at http://www.blu-raydisc.com/assets/downloadablefile/BD-R_Physical_3rd_edition_0602f1-13322.pdf (last access: August, 06, 2007).
- [6] E. D. Sunde, "Ideal binary pulse transmission by AM and FM", *Bell Systems Technical Journal*, vol. 38, pp. 1357-1426, Nov. 1959.
- [7] S. Haykin, *Communication Systems*, 4th Edition - John Wiley and Sons, Inc.: New York, USA, 2001.
- [8] J. G. Proakis, *Digital Communications – 3rd Edition*, McGraw Hill, Inc.: USA, 1995.
- [9] D. A. Guimarães, A Simple FFSK Modulator and its Coherent Demodulator, *IEICE Trans. Fundamentals*. Vol. E91-A, No. 3, pp. 909-910, March 2008.
- [10] H. Leib, S. Pasupathy, "Error Control Properties of Minimum Shift Keying", *IEEE Communications Magazine*, vol.31 No.1, pp. 52-61, January 1993.
- [11] S. Benedetto, and E. Biglieri, *Principles of Digital Transmission With Wireless Applications*. Kluwer Academic and Plenum Publishers: New York, 1999.
- [12] J. K. Omura et. al., "MSK spread-spectrum receiver which allows CDMA operations", *United States Patent 5,963,585*, October 5, 1999.
- [13] M. K. Simon, "A generalization of minimum shift keying (MSK) type signaling based upon input data symbol pulse shaping", *IEEE Trans. on Communications*, vol. COM-24, pp. 845-856, August 1976.
- [14] I. Korn, "Generalized MSK", *IEEE Trans. on Information Theory*, vol. IT-26, no. 2, pp. 234-238, March 1980.
- [15] R. Sadr and J. K. Omura, "Generalized minimum shift-keying modulation techniques", *IEEE Trans. on Communications*, Volume 36, Issue 1, pp. 32-40, Jan 1988.

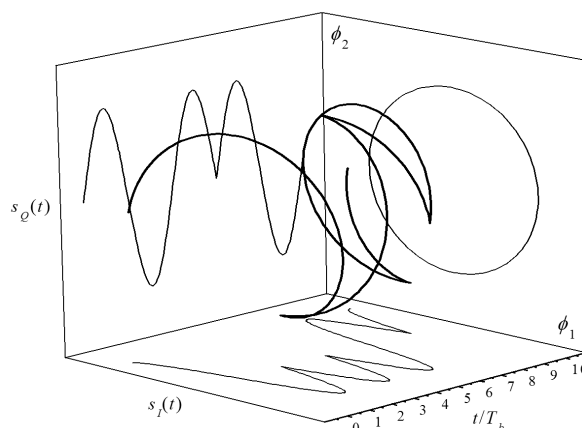


Figure 18. Phase trajectory of an MSK signal. The projections of this trajectory on all planes are also shown.



Dayan Adionel Guimarães was Born in Carrancas, MG, Brazil, on March 01, 1969. He holds the titles: Electronics Technician (ETE "FMC", 1987), Electrical Engineer (Inatel, 1994), Specialist in Data Communication Engineering (Inatel, 2003) and in Human Resources Management (FAI, 1996), Master in Electrical Engineering (Unicamp, 1998) and Doctor in Electrical Engineering (Unicamp, 2003).

From 1988 to 1993 he developed equipment for Industrial Instrumentation and Control, and was also the Manufacturing and Product Engineering Supervisor at *SENSE Sensores e Instrumentos*. Since January 1995 he is Professor at Inatel where, for eight years, he was responsible for the structure that supports practical teaching activities for the Electrical Engineering undergraduate course. His research includes the general aspects on Digital and Mobile Communications, specifically Multi-Carrier CDMA systems, and coding for fading channels, specifically Block Turbo Codes.

Dr. Dayan is member of the *Telecomunicações* magazine's Editorial Board, member of the Inatel's Master Degree Counseling Board and of the IEICE (*Institute of Electronics, Information and Communication Engineers*), Japan.