Turk J Elec Engin, VOL.11, NO.2 2003, © TÜBİTAK

Zero-Crossing Based Demodulation of Minimum Shift Keying

Mine KALKAN
Department of Electronics and Communications, İstanbul Technical University,
Maslak, İstanbul-TURKEY
e-mail:kalkan@itu.edu.tr
Feza KERESTECİOĞLU*
Boğaziçi University, Electrical and Electronics Engineering Department,
Bebek 80815 İstanbul-TURKEY
e-mail:kerestec@boun.edu.tr

Abstract

Minimum shift keying (MSK) modulation has features such as constant envelope, compact spectrum and good error performance, which are all desirable in many digital applications including mobile radio. Numerous receiver structures to demodulate MSK have been suggested, such as correlation receivers, differential detectors and frequency discriminators. MSK is a form of biphase keying and can be detected by a zero-crossing based phase demodulator which gives near optimum performance. In this paper, the bit error performance of a zero-crossing based coherent MSK demodulator is theoretically investigated and a closed-form expression for the bit error rate is derived. The results indicate that the demodulator performs within 0.8-1 dB of the theoratical optimum for MSK. Towards the goal of deriving probability of bit error, it is also shown that under additive white Gaussian noise (AWGN) zero-crossing locations of MSK signals are Gaussian distributed except at very low signal-to-noise ratios.

Key Words: *Minimum shift keying modulation, receiver performance, zero-crossing properties, non-parametric testing.*

1. Introduction

Minimum shift keying (MSK) is a constant envelope, spectrally-efficient modulation scheme which has long been used in digital mobile radio applications including the communications standards such as Digital European Cordless Telecommunications (DECT) and Global System for Mobile Communications (GSM) [1], [2], [3]. In addition to its spectral efficiency, MSK has a good error performance and self-synchronising capability [4].

MSK, conventionally, is a special case of offset quadrature phase-shift keying (OQPSK) with sinusoidal symbol weighting [5]. That is, two sinusoidally-shaped bit streams, one having a bit-period time-offset relative

*Feza Kerestecioğlu is currently on leave for his studies at Kadir Has University, İstanbul

to the other, are employed to modulate orthogonal components of a carrier. The signal in this format is often referred to as precoded MSK [6]. MSK signals can also be obtained by a simpler biphase modulator which accepts the data in serial form [7]. During any keying interval of bit duration, one of the two frequencies, f_1 and f_0, are transmitted, where f_1 and f_0 stand in a special relationship to the baud rate. In view of this, MSK can also be thought of as a form of continuous-phase frequency-shift keying (CPFSK) where the modulation index is 0.5 [4].

Precoded MSK can be coherently detected either by matched filter [5] or by differential decoders [8]. MSK, in biphase keying format, can be detected both coherently and noncoherently. It can be detected as a binary phase-shift keyed (BPSK) signal by a matched filter or noncoherently by a frequency discriminator [7], [9], [10].

MSK, in biphase keying format, can also be detected by a zero-crossing based phase demodulator. In this approach two frequencies, say, $f_1 = 1200$ Hz and $f_0 = 1800$ Hz are transmitted for binary 1 and 0, respectively,with a baud rate of 1200 bps. The resulting signal spectrum is then centred on a carrier located at the mid-frequency between f_1 and f_0 (1500 Hz). This ensures that the signal phase changes $\pm$ 90 relative to the carrier by the end of each period. Coherent detection is achieved by extracting the signal phase from its zero crossings and comparing it to the carrier phase at bit ends.

In this paper the bit error rate (BER) performance of one such receiver is analysed. This requires information on the zero-crossing properties of MSK signals. Therefore, following a brief introduction of MSK signalling and receiver specifications, zero-crossing properties of MSK signals are investigated and asymptotic expressions for the probability density of zero-crossing locations and intervals are derived in Section 2. Using the results of Section 2, an expression for bit error probability is obtained in Section 3 and performance results obtained from the simulations are presented in Section 4. Conclusions are drawn in Section 5.

2. System specifications of the receiver

The block diagram of the zero-crossing based coherent MSK demodulator is shown in Figure 1. The system is basically made up of a phase detector followed by a carrier and clock recovery and decision stages. The received signal $r(t) + n(t)$, after noise limiting, is fed to a zero-crossing detector (ZCD) which generates a sequence of positive impulses at its output. Note that, in order to realise a ZCD, one would hardlimit the input waveform, differentiate to accentuate the zero-crossing points and pass through a full-wave rectifier to eliminate negative pulses due to negative-going zero crossings. The resultant signal is then low-pass filtered and integrated over a bit period to produce an output proportional to the instantaneous signal phase. Here, it would be useful to note that, theoretically, samples of the accumulated signal phase at each zero crossing could have been obtained by directly integrating the ZCD output. However, this would enable one to know the phase values only where the zero crossings occur. As will be explained below, mid-bit and end-bit instantaneous phase values are needed for clock and carrier recovery. A low-pass filter (LPF) cascaded with the ZCD will provide the phase value at any instant over the bit period. The end-bit sample is used for the decision and the mid-bit value, together with the end-bit value, by the clock and carrier recovery stage.

A coherent decision is made by comparing the end-bit sample to a recovered carrier reference. If the signal phase is larger than the carrier phase a binary 0 is selected. Otherwise, the decision is made in favour of a binary 1.

The receiver in Figure 1 can also operate noncoherently. In this case, carrier recovery is not necessary

and the decision will be based on a comparison of two measured phase samples from the received signal; namely, one at the start of the bit period and the other at the end of the bit period. A decision will be made by looking at the direction of the phase change over the bit period.

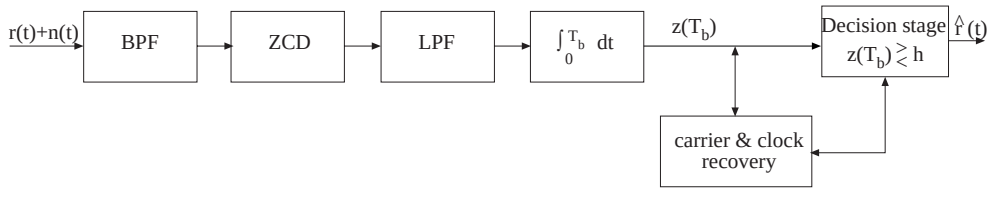

Figure 1. A coherent MSK receiver.

The carrier recovery stage uses the mid-bit phase measurements from the signal to extract the phase reference. If the phase is rising, relative to the carrier phase during the bit period, then the mid-bit phase difference should be 45^o more than the one at the start of the bit period. Similarly, if the phase is falling with respect to the carrier phase then the mid-bit phase difference should be 45^o less than its value at the start of the bit period. By comparing the mid-bit phase with the expected one, the size and direction of the carrier error is obtained and a proportion of this error is used to correct the phase reference [11].

On the other hand, clock recovery is achieved by using a peak search algorithm. It uses three phase samples: two mid-bit samples from two consecutive bits and an end-bit sample between them. Considering the possibility that the phase trajectory relative to the carrier phase has a local maximum (or minimum) at the end of a bit period (this happens whenever a change in the received binary symbol occurs), the algorithm compares the differences in phase between three samples to determine the timing error.

3. Zero-Crossing Properties of MSK Signals

Zero-crossing based demodulation is one of the various schemes used in the detection of frequency modulated signals [12], [13], [14]. Both analogue or digital FM-modulated signals can be recovered from the zero crossings of the received signal; in the case of analogue FM this is done by determining the average number of zero crossings over a certain period and estimating the instantaneous signal frequency from this measurement [15]. In digital FM, normally, more "exact" techniques are used. The currently investigated MSK modem is one where the decision is made on the basis of one zero-crossing difference (i.e., two zero crossings will be received for binary 1 and three for binary 0). The instantaneous signal phase obtained by the integration of interpolated (by the LPF) zero crossings is therefore highly sensitive to the locations of the zero crossings as well as to their number. Gaussian noise added to the MSK signal will result in displacements and density changes in its zero crossings, as shown in Figure 2. This will cause fluctuations in the measured phase and thus erroneous symbol decisions. The error performance analysis of the receiver, therefore, demands information on the statistical properties of zero crossings.

The received MSK signal is simply modelled as a sinusoidal wave plus additive narrow-band Gaussian noise, that is

$$Z(t) = A\cos\omega t + n(t) \tag{1}$$

where ω is the signal frequency and $\mathrm{n}(t)$, narrow-band noise in a quadrature form, is

$$n(t) = n_c(t)\cos\omega t + n_s(t)\sin\omega t. \tag{2}$$

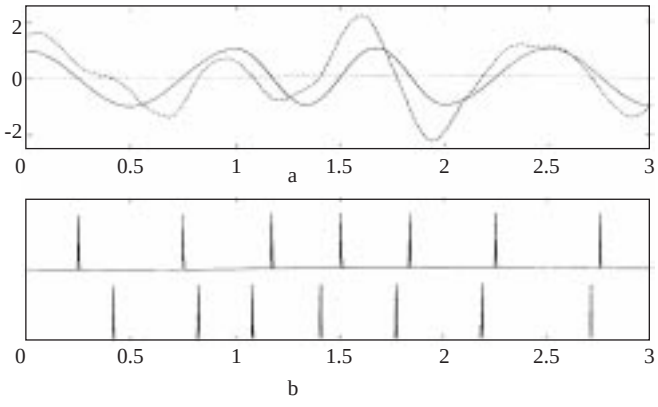

Figure 2. Zero crossings of MSK signals: a) Pure and noisy waveforms (SNR=3 dB) b) Representation of associated zero crossings.

The envelope and phase representation of the signal Z(t) is

$$Z(t) = R(t)\cos(\omega t + \theta(t)) \tag{3}$$

where the envelope R(t) is given by

$$R(t) = \sqrt{(A + n_c(t))^2 + n_s^2(t)}, \tag{4}$$

and the phase $\theta(t)$ is

$$\theta(t) = tan^{-1}\frac{n_s(t)}{n_c(t) + A}. \tag{5}$$

It is well known that the joint pdf of R and θ is given as [16]

$$P_{R,\theta}(r,\theta) = \frac{r}{2\pi N}\exp\left(-\frac{(r^2 + A^2 - 2Ar\cos\theta)}{2N}\right) r > 0, -\pi < \theta < \pi \tag{6}$$

where $N = E\{n^2(t)\}$. From (3), the zero-crossing instants of the received signal will satisfy

$$\omega t_k + \theta = (2k-1)\frac{\pi}{2}. \tag{7}$$

Hence, time of the k$_{th}$ zero crossing is

$$t_k = \frac{(2k-1)\frac{\pi}{2} - \theta}{\omega}. \tag{8}$$

Note that since the process $\theta(t)$ is strict sense stationary, its distribution will be independent of time; therefore, in the analysis the time argument of $\theta(t)$ is dropped. As can be seen from (8), the pdf of the location of the zero crossings can be obtained from the pdf of θ. The pdf of θ, P$_\theta(\theta)$, can be obtained by integrating out r in the joint pdf P$_{R,\theta}(r,\theta)$ as

$$\begin{aligned} P_\theta(\theta) &= \int_0^\infty p_{R,\theta}(r,\theta)dr \\ &= \int_0^\infty \frac{r}{2\pi N}\exp\left(-\frac{r^2+A^2-2Ar\cos\theta+A^2\cos^2\theta-A^2\cos^2\theta}{2N}\right)dr \\ &= \frac{1}{2\pi N}\exp\left(-\frac{A^2\sin^2\theta}{2N}\right)\int_0^\infty r\exp\left(-\frac{(r-A\cos\theta)^2}{2N}\right)dr. \end{aligned} \tag{9}$$

Let $u = (r - A\cos\theta)/\sqrt{N}$. Then the integral on the right-hand side will be

$$\begin{aligned} &\int_{-A\cos\theta/\sqrt{N}}^\infty \left(\sqrt{N}u + A\cos\theta\right)\exp\left(-\frac{u^2}{2}\right)\sqrt{N}du \\ &= N\int_{-A\cos\theta/\sqrt{N}}^\infty u\exp\left(-\frac{u^2}{2}\right)du + A\cos\theta\sqrt{N}\int_{-A\cos\theta/\sqrt{N}}^\infty \exp\left(-\frac{u^2}{2}\right)du \\ &= N\exp\left(-\frac{A^2\cos^2\theta}{2N}\right) + A\cos\theta\sqrt{2\pi N}Q(-\frac{A\cos\theta}{\sqrt{N}}) \end{aligned} \tag{10}$$

where $Q(x)$ is a complementary Gaussian cumulative distribution function defined as [6]

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \exp\left(-\frac{1}{2}u^2\right)du.$$

It follows that

$$p_\theta(\theta) = \frac{1}{2\pi}\exp\left(-\frac{A^2}{2N}\right) + \frac{A\cos\theta}{\sqrt{2\pi N}}Q(-\frac{A\cos\theta}{\sqrt{N}})\exp\left(-\frac{A^2\sin^2\theta}{2N}\right). \tag{11}$$

Let $\Gamma = A/\sqrt{N}$. Then $\mathrm{p}_\theta(\theta)$ can be written as

$$p_\theta(\theta) = \frac{1}{2\pi}\exp\left(-\frac{\Gamma^2}{2}\right) + \frac{\Gamma\cos\theta}{\sqrt{2\pi}}Q(-\Gamma\cos\theta)\exp\left(-\frac{\Gamma^2\sin^2\theta}{2}\right). \tag{12}$$

For large Γ and small θ, (12) reduces to Gaussian probability distribution function, that is

$$p_\theta(\theta) \approx \frac{\Gamma}{\sqrt{2\pi}}\exp\left(-\frac{\Gamma^2\theta^2}{2}\right). \tag{13}$$

Here, notice that $\Gamma^2/2$ corresponds to SNR. In Figure 3, $p_\theta(\theta)$ given by Eq. (12), is plotted for various Γ. Here, θ is in radians. It is seen from the figure that $p_\theta(\theta)$ resembles the Gaussian shape more as Γ gets larger. For small SNR values it tends to approximate the uniform distribution as expected.

From (8) and (13) the zero-crossing locations have a variance of $1/(\Gamma\omega)^2$. Thus, the Gaussian random variables representing the zero-crossing locations of the received 1's and 0's can be given as

$$x_{1k} \sim N\left(\frac{(2k-1)\,\pi/2}{\omega_1}, \frac{1}{\Gamma^2\omega_1^2}\right) \quad k = 1, 2 \tag{14a}$$

and

$$x_{0k} \sim N\left(\frac{(2k-1)\pi/2}{\omega_0}, \frac{1}{\Gamma^2\omega_0^2}\right) k = 1, 2, 3. \tag{14b}$$

where the index k has been used to distinguish between consecutive zero crossings. The confidence level, at which the approximation in (13) holds (thus the zero-crossing locations can be assumed to be Gaussian distributed), can be determined by conducting a nonparametric statistical test. The following subsection deals with this issue.

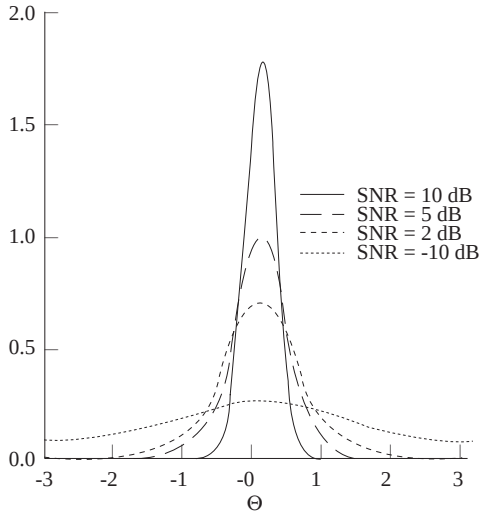

Figure 3. Probability density of θ.

3.1. Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (K-S) one-sample test is suitable for testing the result in (14a,b). The test involves examining random samples with some unknown probability distribution in order to test the null hypothesis that the unknown distribution is in fact a known, specified function. This is done by comparing the experimental cumulative frequency distribution function (CDF) to the theoretical CDF, which should be expected under the null hypothesis. If there is a good agreement between these two, the null hypothesis is accepted, otherwise it is concluded that the unknown probability distribution is not the one in the null hypothesis [17].

Let $S_n(x)$ and $F(x)$ denote the experimental and the specified theoretical distributions, respectively, where n stands for the sample size. In view of the analysis in the preceding section, the null hypothesis is chosen as

$\boldsymbol{H}_o$: The probability density of zero-crossing locations is Gaussian.

The K-S test can then be conducted by taking the following steps:

1. The theoretical CDF, $\mathrm{F}(x)$, expected under the null hypothesis is obtained.

2. The observed values $\{x_{1,\ldots,}x_n\}$ are arranged in an empirical CDF, $S_n(x)$.

3. The test statistic $T = \max_{1\leq i\leq n}|F(x_i) - S_n(x_i)|$ is calculated.

4. T is compared to a predetermined threshold h and $\boldsymbol{H}_o$ is accepted if $T \leq h$.

The critical values of the probability of falsely rejecting $\boldsymbol{H}_o$, when in fact it is true (i.e., type I error), associated with particular values of h are tabulated in the literature [18]. These probabilities are called levels of significance and are denoted by α. Common values of α are $0.20 \sim 0.01$. The significance level of 0.20 is the one at which the strongest confirmation of the null hypothesis is obtained.

For the computation of the test statistics, samples of the zero-crossing locations of the waveform corresponding to binary 1 under additive Gaussian noise are obtained. Figure 4 shows a histogram of zero-crossing locations in an interval of width $\mathrm{T}_b/2$ for a sample size of $n = 3000$. The SNR has been chosen as 5 dB. In producing the histogram the waveform has been sampled at a rate of $100/T_b$. The empirical CDF obtained from these sample values and the theoretical CDF are plotted in Figure 5. The test statistic T is found to be 0.0102. It can be seen from the tabulated statistics [18] that the probability of type I error corresponding to this value of T is larger than the significance level of 0.20. Thus the test passes most confidently. The same procedure has been repeated for various SNR values. The results show that even at a low SNR value of 3 dB, the test passes at a significance level of 0.10 and it is needless to say that the larger the SNR the higher is the significance level. We note that similar results hold in the case of binary 0, and so they are not repeated here.

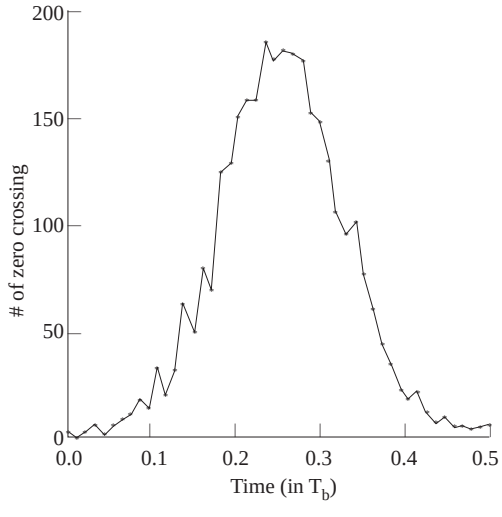

Figure 4. Histogram of zero-crossing locations.

4. Bit Error Probabilities for the Coherent MSK Demodulator

In order to derive the bit error probability one should obtain the decision variable z, which is the accumulated phase at the end of a bit period. One method of doing this is to calculate the spectrum of an aperiodic impulse train and convolve it with the frequency response of the LPF. Integration of the inverse transform of this result over the period T_b would then give the variable z. The MSK signal is made up of randomly interleaved signals of two different frequencies. Thus, when the zero crossings are detected, one would have an aperiodic sequence of impulses even in the absence of noise. The spectrum of a periodic impulse train

with random shifts due to the noise is given in the literature [19]; however, that of the aperiodic pulse train seems to be quite difficult, if not impossible, to obtain.

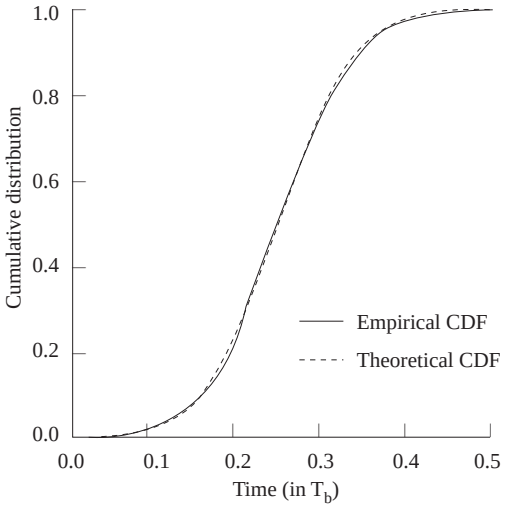

Figure 5. Cumulative distribution of zero-crossing locations.

A relatively practical way of obtaining z is to consider a rectangular time window of length T_b cascaded to the ZCD output (this will, of course, require almost perfect clock recovery). Multiplication of the impulse train with this type of window function results in a certain number of impulses selected over a bit period, as illustrated in Figure 6. The spectrum of a signal like this can be expressed in a simple form and this allows one to obtain z analytically, as shown in the subsection below.

4.1. Derivation of the bit error probability

Assume that the impulse train at the ZCD output is sequentially multiplied by a rectangular window of length T_b s. Therefore, in the noise-free case the resulting signal will be two impulse functions, one shifted by $T_b/2$ s relative to the other, when a one is sent. In the case of a transmitted zero there will be three impulses separated by $T_b/3$ s. Recall from Section 2 that when Gaussian noise is added to the signal zero crossings will be randomly shifted away from where they should be in the noise-free case and these locations themselves will be Gaussian random variables. Thus, in the case of a noisy one the Fourier Transform of the signal is (see Figure 6a)

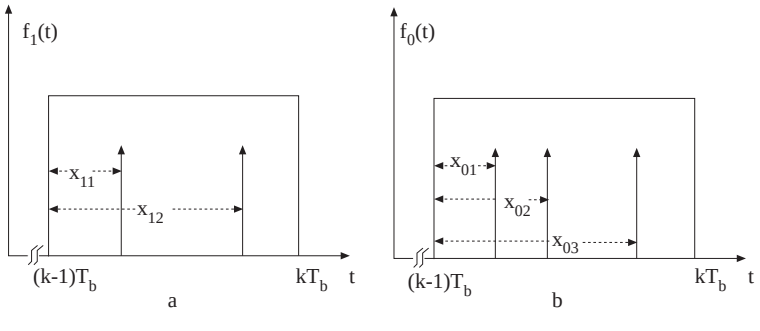

Figure 6. Representation of shifted zero-crossings incorporating a window function. a) Binary 1 b) Binary 0.

$$F\left(\omega, x_{11}, x_{12}\right) = e^{-j\omega x_{11}} + e^{-j\omega x_{12}} \tag{15}$$

where x_{1i} denotes the occurrence time (location) of the i^{th} zero crossing. After the low-pass filtering, the frequency-domain representation of the signal will be

$$X(f) = \begin{cases} e^{-j\omega x_{11}} + e^{-j\omega x_{12}} & \omega \leq \omega_c \\ 0 & elsewhere \end{cases} \tag{16}$$

where ω_c is the filter cut-off frequency (in radians). The inverse transform of (16) yields

$$s_1(t) = \frac{\sin(\omega_c(t - x_{11}))}{\pi(t - x_{11})} + \frac{\sin(\omega_c(t - x_{12}))}{\pi(t - x_{12})}. \tag{17}$$

The decision variable z_1 will then be the integrator output:

$$z_1 = \int_0^{T_b} \frac{\sin(\omega_c(t - x_{11}))}{\pi(t - x_{11})} dt + \int_0^{T_b} \frac{\sin(\omega_c(t - x_{12}))}{\pi(t - x_{12})} dt. \tag{18}$$

Substituting the variable transform $x = \omega_c(t - x_{1i})$ in (18),

$$z_1 = \int_{-\omega_c x_{11}}^{\omega_c(T_b - x_{11})} \frac{1}{\pi}.\frac{\sin x}{x} dx + \int_{-\omega_c x_{12}}^{\omega_c(T_b - x_{12})} \frac{1}{\pi}.\frac{\sin x}{x} dx \quad . \tag{19}$$

Noting that $\int_0^a \sin x/x dx$ is known as the sine integral, $\mathrm{Si}(a)$, and is tabulated in the literature [20], (19) is finally written as

$$z_1 = \frac{1}{\pi}\left(Si(\omega_c(T_b - x_{11})) + Si(\omega_c x_{11}) + Si(\omega_c(T_b - x_{12})) + Si(\omega_c x_{12})\right). \tag{20}$$

Similarly, in the case of a noisy "zero" the transform of the signal (after windowing operation) will be

$$F(\omega, x_{01}, x_{02}, x_{03}) = e^{-j\omega x_{01}} + e^{-j\omega x_{02}} + e^{-j\omega x_{03}} \quad . \tag{21}$$

Taking the same steps as in (15)-(20), the decision variable z_0 is obtained as

$$z_0 = \int_{-\omega_c x_{01}}^{\omega_c(T_b - x_{01})} \frac{1}{\pi}.\frac{\sin x}{x} dx + \int_{-\omega_c x_{02}}^{\omega_c(T_b - x_{02})} \frac{1}{\pi}.\frac{\sin x}{x} dx + \int_{-\omega_c x_{03}}^{\omega_c(T_b - x_{03})} \frac{1}{\pi}.\frac{\sin x}{x} dx. \tag{22}$$

In addition, (22) can be written in the form

$$z_0 = \frac{1}{\pi}\left(Si(\omega_c(T_b - x_{01})) + Si(\omega_c x_{01}) + Si(\omega_c(T_b - x_{02})) + Si(\omega_c x_{02}) + Si(\omega_c(T_b - x_{03})) + Si(\omega_c x_{03})\right) \tag{23}$$

The LPF and integrator outputs of binary 1 and 0 are shown in Figure 7. It is seen from these figures that the LPF, basically, interpolates between the zero crossings and this results in a smooth increase in variables $z_1(t)$ and $z_0(t)$.

The next step in obtaining the error probability is to find the pdf's of the decision variables z_1 and z_0. Both z_1 and z_0 are functions of random variables with known (Gaussian) statistics. However, their densities cannot be determined in a closed form from these statistics, because x_{11}, for example, cannot be expressed as a function of z_1 and x_{12} (see (19)).

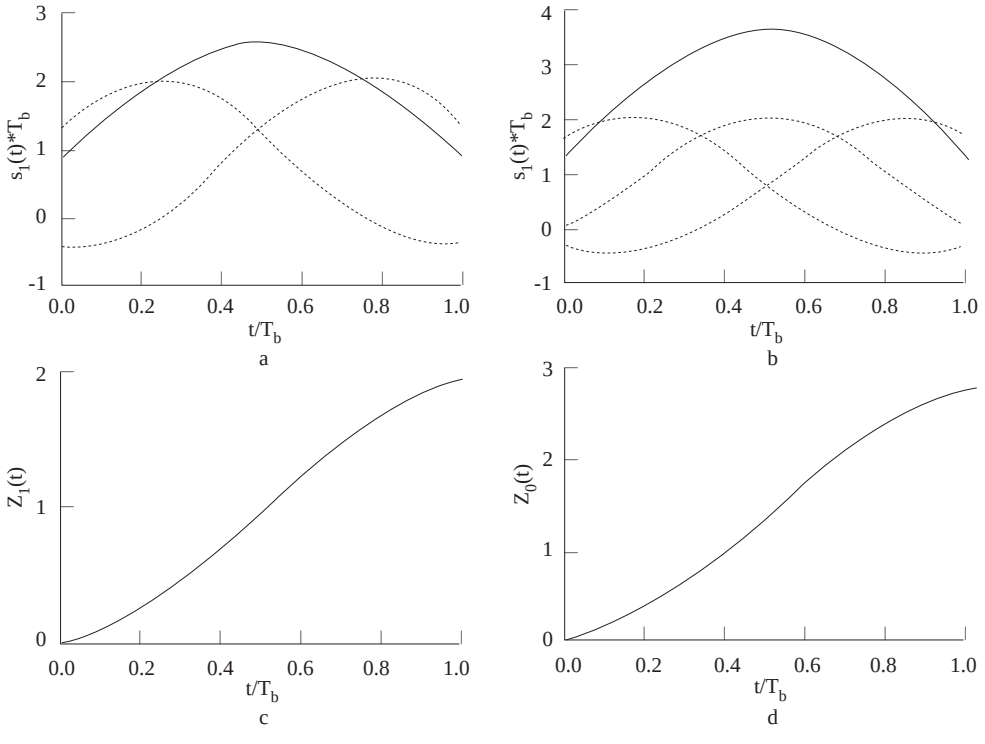

Figure 7. LPF and integrator outputs: a) and b) Binary 1, c) and d) Binary 0.

Nevertheless, one can obtain approximate expressions for z_1 and z_0 and then determine the desired densities. Let us express z_1 in (19) as

$$z_1 = \frac{1}{\pi}(f(x_{11}) + f(x_{12})) \tag{24}$$

where $f(x_{11})$ and $f(x_{12})$ denote respectively the first and the second integral terms on the right-hand side. The variable z_1 can be linearly approximated by expanding $f(x_{11})$ and $f(x_{12})$ into their Taylor series about the points $x_{11} = T_b/4$ and $x_{12} = 3T_b/4$, respectively, and neglecting terms of order higher than two. Thus,

$$f(x_{11}) \approx f(T_b/4) + f'(T_b/4)(x_{11} - T_b/4). \tag{25}$$

Clearly, such an approximation is valid if the deviation of x_{11} from its mean is small, i.e., if the SNR is reasonably high. After straightforward calculations (25) can be written as

$$f(x_{11}) \approx Si(A_1) + Si(A_2) - \sin A_1 + \frac{1}{3}\sin A_2 + \frac{4}{T_b}\left(\sin A_1 - \frac{1}{3}\sin A_2\right) x_{11} \tag{26}$$

where $A_1 = \omega_c T_b/4$ and $A_2 = 3\omega_c T_b/4$. Similarly an expansion for $f(x_{12})$ will yield

$$f(x_{12}) \approx f(3T_b/4) + f'(3T_b/4)(x_{12} - 3T_b/4). \tag{27}$$

After some manipulations,

$$f(x_{12}) \approx Si(A_1) + Si(A_2) + 3\sin A_1 - \sin A_2 - \frac{4}{T_b}\left(\sin A_1 - \frac{1}{3}\sin A_2\right) x_{12}. \tag{28}$$

We have from (26), (28) and (24)

$$z_1 \approx \frac{2}{\pi}\left(Si(A_1) + Si(A_2) + \sin A_1 - \frac{1}{3}\sin A_2\right) + \frac{4}{\pi T_b}\left(\frac{1}{3}\sin A_2 - \sin A_1\right)(x_{12} - x_{11}). \tag{29}$$

Finally, (29) can be expressed as

$$z_1 = C_1 + C_2(x_{12} - x_{11}) \tag{30}$$

where

$$C_1 = \frac{2}{\pi}\left(Si(A_1) + Si(A_2) + \sin A_1 - \frac{1}{3}\sin A_2\right) \tag{31a}$$

and

$$C_2 = \frac{4}{\pi T_b}\left(\frac{1}{3}\sin A_2 - \sin A_1\right). \tag{31b}$$

Similarly the decision variable z_0 can be written from (22) as

$$z_0 = \frac{1}{\pi}\left(f(x_{01}) + f(x_{02}) + f(x_{03})\right). \tag{32}$$

The linear approximation to $f(x_{01})$ about the point $x_{01} = T_b/6$ yields

$$f(x_{01}) \approx Si(A_3) + Si(A_4) - \sin A_3 + \frac{1}{5}\sin A_4 + \frac{6}{T_b}\left(\sin A_3 - \frac{1}{5}\sin A_4\right) x_{01} \tag{33}$$

where $A_3 = \omega_c T_b/6$ and $A_4 = 5\omega_c T_b/6$. Notice that the point $T_b/6$, which is the centre point of this approximation, is the expected value of occurrence time of the first zero crossing for a "zero" signal. The expansion for $f(x_{02})$ will be about the point $T_b/2$ (i.e., $E\{x_{02}\}$):

$$f(x_{02}) \approx 2Si(A_5) \tag{34}$$

where $A_5 = \omega_c T_b/2$. It follows that

$$f(x_{03}) \approx Si(A_3) + Si(A_4) + 5\sin A_3 - \sin A_4 + \frac{6}{T_b}\left(\frac{1}{5}\sin A_4 - \sin A_3\right) x_{03} \tag{35}$$

where A_3 and A_4 are defined as before and the expansion is about the point $E\{x_{03}\} = 5T_b/6$. Substituting (32), (33) and (34) into (31)

$$z_0 \approx \frac{2}{\pi}\left(Si(A_3) + Si(A_4) + 2\sin A_3 - \frac{2}{5}\sin A_4 + Si(A_5)\right) + \frac{6}{\pi T_b}\left(\frac{1}{5}\sin A_4 - \sin A_3\right)(x_{03} - x_{01}), \tag{36}$$

i.e.,

$$z_0 = C_3 + C_4(x_{03} - x_{01}) \tag{37}$$

where the constants C_3 and C_4 are

$$C_3 = \frac{2}{\pi}\left(Si(A_3) + Si(A_4) + 2\sin A_3 - \frac{2}{5}\sin A_4 + Si(A_5)\right) \tag{38a}$$

and

$$C_4 = \frac{6}{\pi T_b}\left(\frac{1}{5}\sin A_4 - \sin A_3\right). \tag{38b}$$

Having put z_1 and z_0 into forms as in (30) and (37), respectively, one can now determine their densities in terms of the known densities of zero-crossing locations, x_{1k} and x_{0k}. It can be seen from (30) that z_1 is a function of $(x_{11} - x_{12})$, which is the distance between two consecutive zero crossings in the case of a transmitted one. Recall from Section 2 that zero-crossing locations are Gaussian distributed. The distance between the zero crossings, which is in fact the difference of two Gaussian random variables, is therefore Gaussian distributed with a mean being the difference of their means and a variance which equals the sum of their variances. Let δ_1 denote the difference random variable $(x_{11} - x_{12})$. Thus the pdf of δ_1 would be given as

$$p_{\delta_1}(\delta) = \frac{1}{\sigma_{\delta_1}\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(\delta - T_b/2)^2}{\sigma_{\delta_1}^2}\right) \tag{39}$$

where $T_b/2$ and σ_δ^2 are the mean and the variance of δ respectively. Assuming x_{11} and x_{12} are statistically independent, we have $\sigma_\delta^2 = 2\sigma_1^2$, where σ_1^2 is the variance of zero-crossing locations in the case of a binary 1. Thus from (14a)

$$\sigma_{\delta_1}^2 = 2\sigma_1^2 = \frac{1}{2\pi^2\Gamma^2 f_1^2}. \tag{40}$$

It can be seen from (30) and (39) that the pdf of z_1 can be written as

$$p_{z_1}(z) = \frac{1}{|C_2|\sigma_{\delta_1}\sqrt{2\pi}} \exp\left(-\frac{(z-(T_b/2+C_1))^2}{2C_2^2\sigma_{\delta_1}^2}\right). \tag{41}$$

Similar to the steps as in (39)-(41), the density of z_0 can be obtained as

$$p_{z_0}(z) = \frac{1}{|C_4|\sigma_{\delta_0}\sqrt{2\pi}} \exp\left(-\frac{(z-(2T_b/3+C_3))^2}{2C_4^2\sigma_{\delta_0}^2}\right) \tag{42}$$

where the variance $\sigma_{\delta_0}^2$ is given as (see (14b))

$$\sigma_{\delta_0}^2 = 2\sigma_0^2 = \frac{1}{2\pi^2\Gamma^2 f_0^2} \ . \tag{43}$$

For the currently considered binary MSK there are two ways in which errors can occur. That is, an error occurs when the signal $s_1(t)$ (i.e., binary 1) is transmitted, but hypothesis H_0 (hypothesis that a zero was sent) is chosen or $s_0(t)$ (i.e., binary 0) is transmitted and hypothesis H_1 is chosen. Therefore, the probability of error is

$$P_B = P(H_0|s_1)P(s_1) + P(H_1|s_0)P(s_0) \tag{44}$$

where $P(s_1)$ and $P(s_0)$ are a priori probabilities of sending binary 1 and binary 0, respectively. As explained in Subsection 6.1.1, the detection is achieved by comparing the phase sample taken at the end of the bit period with the estimate of the carrier phase. This corresponds to comparing the decision variable z (being z_1 or z_0) to a threshold h which is mid-way between the expected values of z_1 and z_0. Since the lower frequency is used for binary 1, z_1 (i.e., the phase sample) is supposed to be less than the threshold h for correct detection. Thus, assuming $P(s_1) = P(s_0) = 1/2$, (44) can be expressed as

$$P_B = \frac{1}{2}P(z_1 > h) + \frac{1}{2}P(z_0 < h). \tag{45}$$

The probability that z_1 exceeds h can be computed by integrating $P_{z_1}(z)$ between the limits h and ∞

$$P(z_1 > h) = \int_h^\infty \frac{1}{|C_2|\sigma_{\delta_1}\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(z-(T_b/2+C_1))^2}{C_2^2\sigma_{\delta_1}^2}\right)dz. \tag{46}$$

Substituting $(z-(T_b/2+C_1))/C_2\sigma_{\delta_1} = u$ and $dz = C_2\sigma_{\delta_1}du$ in (46)

$$P(z_1 > h) = \int_{[h-(T_b/2+C_1)]/C_2\sigma_{\delta_1}}^\infty \exp(-\frac{1}{2}u^2)du. \tag{47}$$

It follows that

$$P(z_1 > h) = Q\left(\frac{h - (T_b/2 + C_1)}{C_2\sigma_{\delta_1}}\right). \tag{48}$$

Similarly, $P(z_0 < h)$ is obtained by integrating $P_{z_0}(z)$ between the limits $-\infty$ and h. This yields, after some manipulations,

$$P(z_0 < h) = Q\left(\frac{(2T_b/3 + C_3) - h}{C_4\sigma_{\delta_0}}\right). \tag{49}$$

Substituting (48) and (49) into (45), P_B is expressed as

$$P_B = \frac{1}{2}\left\{Q\left(\frac{h - (T_b/2 + C_1)}{C_2\sigma_{\delta_1}}\right) + Q\left(\frac{(2T_b/3 + C_3) - h}{C_4\sigma_{\delta_0}}\right)\right\}. \tag{50}$$

The probability of bit error given in (50) is plotted in Figure 8 together with the optimum error probability for binary MSK[1]. The currently investigated scheme performs with a degradation of about 1.5 dB with respect to the optimum MSK detector. Recall that in the model, on which the error probability analysis has been based, the errors are assumed to be made due to shifted zero-crossing locations only. However, especially at low SNRs, errors can also be due to extra or missing zero crossings, which are other manifestations of additive noise. Thus, P_B in (50) sets a lower bound on the system's error rate which, in actual occasions, will be higher. Nevertheless, in the next section where the theoretical results are compared to the simulations, some empirically obtained rules will be presented which can reduce the performance degradation due to such extra or missing zero crossings.

5. Simulation Results

The coherent demodulator described in Section 1 is digitally simulated under perfect carrier and clock recovery to obtain the error performance. A high sampling rate of 100 samples/bit is chosen to ensure accurate measurement of the noisy signal phase. The LPF used is a sixth-order Butterworth with optimum cut-off frequency of 1200 Hz. Figure 9 shows the theoretical and simulated BER curves. In the SNR region of 4-8 dB the simulation results are about $0.8 \sim 1$ dB worse than the theoretical. The difference diminishes for larger SNR. This is partly due to the fact that the approximations involved in the derivation of the BER (i.e., Gaussianity of zero-crossing locations and the linear approximations to z_0 and z_1) hold better for high SNR values. Nevertheless, it can also be an indication of another error-causing event, which has not been considered in the original model: at low-to-medium SNRs, noise also results in extra or missing zero crossings (in addition to those shifted) in the received signal. This causes a sudden increase (if extra) or decrease (if missing) in the signal phase resulting in erroneous detection. However, a zero-crossing correction algorithm can be used to correct such errors as will be shown in the subsection below.

[1]The optimum error rate for coherently detected binary MSK is given as [6]
$P_B = Q\left(\sqrt{2E_b/N_0}\right)$ where E_b is the energy per bit and N_0 is the noise power spectral density.

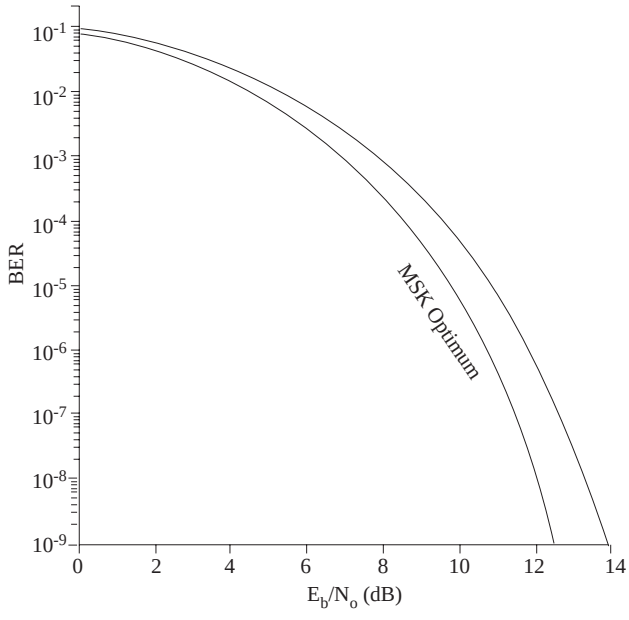

Figure 8. Bit error probability for optimally detected MSK and the receiver of Figure 1.

5.1. A zero-crossing correction rule

An algorithm to avoid the detection errors caused by extra or missing zero crossings would, basically, count the number of zero crossings in a bit period then decide on whether extras or missings occurred and if they did how many.

The simulations are run for 3000 bits to determine the likelihood of extras and missings in the case of two binary waveforms, 1200 Hz and 1800 Hz, respectively. Table 1 shows the results for a noise bandwidth of $W = 800$ Hz. Note that $SNR = 2$ dB here.

Table 1. Distribution of number of zero crossings (800 Hz).

# of zero crossings		0	1	2	3	4
# of bits	$f_1 = 1200Hz$	0	11	2676	151	162
	$f_0 = 1800Hz$	0	27	182	2727	64

Thus, in the case of a binary 1, two extra crossings are most likely to occur. For binary 0, however, a single missing is the most possible. On the other hand, the results with doubled noise bandwidth (i.e. W = 1600 Hz) without changing the SNR are given in Table 2.

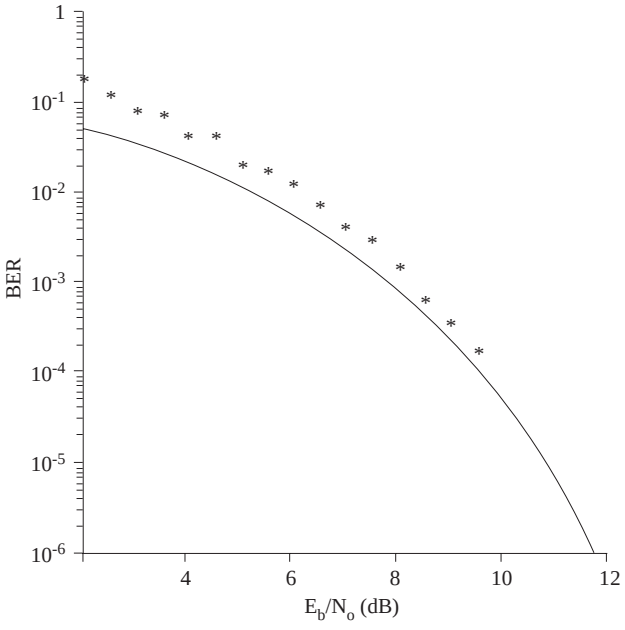

Figure 9. Theoretical BER and simulation results.

It can be concluded from these statistics that the lower frequency is more likely to get two extra crossings while the higher will receive single missings with the highest probability. The logical explanation to this is that noise components significantly higher than the signal frequency are likely to cause extra zero crossings, while the lower ones would result in waveforms with missing zero crossings. Note that when W = 800 Hz (i.e. the noise band is 1100-1900 Hz) the noise would mostly contain frequencies higher than a signal frequency of 1200 Hz. However, if the signal frequency is 1800 Hz the noise will be composed of relatively lower frequencies. When the noise bandwidth is doubled towards the band 700-2300 Hz, one obtains an increased number of extra crossings in the lower frequency signal and missings in the high frequency signal (see Tables 1 and 2)

Table 2. Distribution of number of zero crossings (W = 1600 Hz).

# of zero crossings		0	1	2	3	4	5	6
# of bits	$f_1 = 1200Hz$	0	20	2432	229	310	8	1
	$f_0 = 1800Hz$	0	35	120	2724	91	30	0

In view of these findings some simple rules can be applied to avoid such errors:

1. If the number of zero crossings counted in a bit period is more than three (and, hence, there are definitely extras) some of the impulses at the output of the zero-crossing detector must be removed so as to make the total number two. That is, the correction will be made as if a binary 1 is transmitted during that bit period. This is because, the occurrence of extra crossings under a binary 1 is much more likely than that under binary 0.

2. In the case of having a single zero crossing within a bit period two more will be inserted as if binary 0 is sent, since the probability of the occurrence of a single missing is much higher in the case of a binary 0, compared to that of a binary 1. Evidently, each added zero crossing is placed in the middle of the largest interval between other zero crossing(s) and/or the bit ends so as to yield the most even possible placement.

3. In the cases of having two or three zero crossings no correction should be made.

These rules make use of only the number of zero crossings to rectify obvious errors. It can also be suggested that further information about the zero crossing locations (e.g., their relative distances) can be used to correct possible missing or extra zero crossings even when there are two or three crossings during a bit period. When two crossings are counted, the signal might be a true 1, or a 0 with one missed crossing. Similarly, in the case of three crossings it might either be a 0, or a 1 with an extra crossing. We can see from the Tables 1 and 2 that such events have nonzero probabilities. It has been shown in Section 3 that the distance between the zero crossings is of a Gaussian random variable with a mean $T_b/2$ or $T_b/3$ and a variance $\sigma^2_{\delta_1}$ or $\sigma^2_{\delta_0}$, respectively for binary 1 or 0. If the distance values which are likely under binary 1 and 0 do not overlap one can use this information to find out whether the number of zero crossings is erroneous. For example, two crossings in a bit period whose distance is about $T_b/3$would mean that one zero crossing out of three is missing. Assuming that, in practice, the distances would spread around their means not more than $3\sigma_\delta$, in order to apply such rules reliably it is required that $3\sigma_{\delta_1} + 3\sigma_{\delta_0} \leq T_b/6$. In view of (40) and (43), this condition is satisfied for SNR $\geq$ 6 dB. Thus, such corrections by looking at distances among zero crossings would not be possible for low-to-moderate SNR values.

The MSK receiver of Figure 1 is digitally simulated with the zero-crossing correction rule. It is seen from Figure 10 that, with the correction applied, an improvement of 0.5-0.7 dB is achieved in the BER performance over the SNR range of 3-8 dB. Notice that with this improvement the theoretical values are well confirmed. For larger SNRs, this improvement is lost since extra or missing zero crossings are unlikely to occur. Although the correction of the zero crossings is based on an intuitive rule of thumb, it can result in moderately improved BER performance. However, the simulations show that it is not that useful for SNR values lower than 3 dB.

In addition, it would be useful to indicate that performance gain by the zero-crossing corrections will be more apparent under operations with imperfect carrier recovery. In the case of having extra or missing zero crossings, the signal phase will experience abrupt changes which will worsen carrier recovery. Resynchronisation of the carrier phase will usually take somewhat longer than a bit period (faster recovery may not be suitable, because this will increase the sensitivity of the recovered phase to the noise) [11]. Therefore, a number of bits can be decoded erroneously during the acquisition time. Thus, reversing the effect of extra or missing crossings by the correction rule will also prevent the system performance from being degraded further due to a carrier unlock.

5.2. Noncoherent detector performance

As stated previously, the receiver structure in Figure 1 can also be used for the noncoherent detection of MSK. In this type of operation, the decision will be made by comparing the signal phase by comparing the signal phase sampled at the end of the bit period with the one at the start of the bit period. This will reduce receiver complexity since no carrier recovery is needed. However, the performance will be much poorer compared to the coherent receiver. This is because a detection based on a comparison of two noisy samples is more sensitive to noise than the one between a noisy phase and a phase reference. Figure 11 shows the BER of a noncoherent receiver obtained by simulation. It is seen that the degradation is about 4.5 dB compared to the performance of the coherent receiver (see Figure 9).

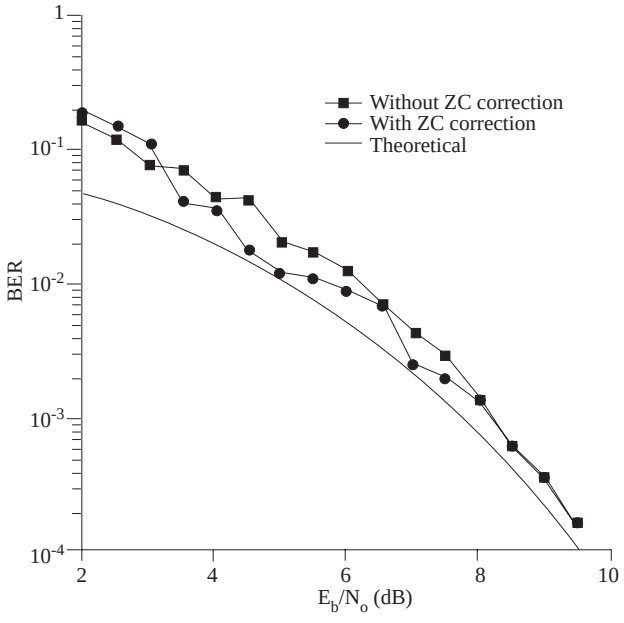

Figure 10. Improvement in BER performance by zero-crossing correction.

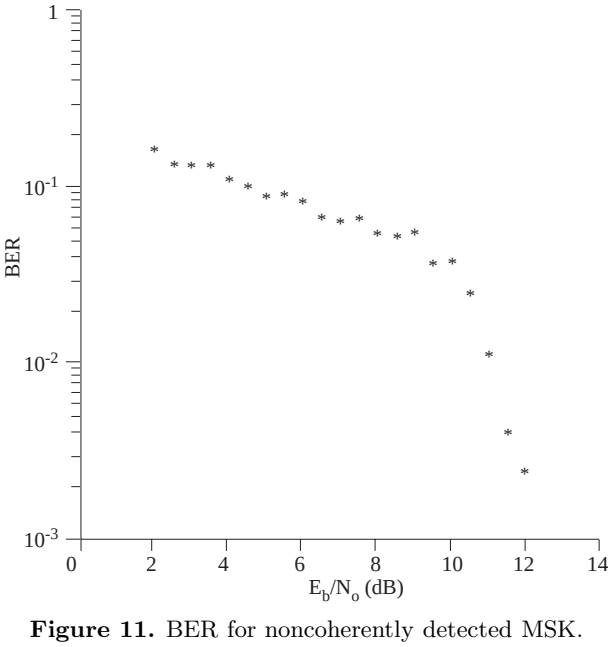

Figure 11. BER for noncoherently detected MSK.

6. Conclusions

In this paper, the bit error performance of a zero-crossing-based MSK demodulator has been theoretically investigated. A closed form expression for BER was derived. The results have indicated that the system performs within 0.8-1 dB of the theoretical optimum for MSK. This is a much better performance than those of many types of MSK receivers; namely, differentially coherent ones or frequency discriminators [6]. Thus, with its good error performance and simplicity, the receiver is a good candidate for many digital applications.

It has been shown that the location of a zero crossing is Gaussian distributed except at low SNR values. In this sense, bit error probability is an asymptotic expression and is more reliable at medium to high SNR

values. This is not only because of the approximations involved in the distribution of zero crossings, but also due to extra and missing zero crossings occuring at low SNRs.

The probability of the zero-crossing intervals obtained from this result has been shown to be useful in the performance analysis of the coherent receiver. In fact, analysis of any zero-crossing-based digital FM demodulator would require such information; therefore this result may also prove to be useful in the analysis of other receivers.

References

[1] K. Feher, "Modems for emerging digital cellular mobile radio systems", *IEEE Transactions on Vehicular Technology*, Vol. VT-40, pp. 355–365, 1991.

[2] Y.L. Huang, K.D. Fan and C.C. Huang, "A fully digital noncoherent and coherent GMSK receiver architecture with joint symbol timing error and frequency offset estimation", *IEEE Transactions on Vehicular Technology*, Vol. VT-49, pp. 863–873, 2000.

[3] R. Prasad, W. Mohr and W. Konhauser, *Third Generation Mobile Communication Systems*, Boston, Artech House, 2000.

[4] T. Rappaport, *Wireless Communications: Principles and Practice,* New Jersey Prentice Hall, 2002.

[5] S.A. Gronemeyer and A.L. McBride, "MSK and offset QPSK modulation", *IEEE Transactions on Communications*, Vol. COM-24, pp. 809–820, 1976.

[6] B. Sklar, *Digital Communications: Fundamentals and Applications*, Englewood Cliffs Prentice-Hall, 1988.

[7] F. Amoroso and J.A. Kivett, "Simplified MSK signaling technique", *IEEE Transactions on Communications,* Vol. COM-25, pp. 433–440, 1977.

[8] G.H. Kaleh, "A differentially coherent receiver for minimum shift keying signal", *Journal on Selected Areas in Communications*, Vol. 7, pp. 99–106, 1989.

[9] S. Pasupathy, "Minimum Shift Keying: A spectrally efficient modulation", *IEEE Communications Magazine*, Vol. 17, pp. 14–22, 1979.

[10] W.P. Osborn and M.B. Luntz, "Coherent and noncoherent detection of CPFSK", *IEEE Transactions on Communications*, Vol. COM-22, pp. 1023–1036, 1974.

[11] R.J. Murray and R.W. Gibson, "A coherent digital demodulator for minimum shift key and related modulation schemes, *Philips Journal of Research*, Vol. 39, pp. 1–10, 1984.

[12] H.B. Voelcker, "Zero-crossing properties of angle-modulated signals", *IEEE Transactions on Communications*, Vol. COM-20, pp. 307–315, 1972.

[13] M. Schwartz, *Information Transmission, Modulation, and Noise*, Auckland McGraw-Hill, 1980.

[14] K. Piwnicki, "Modulation methods related to sine-wave crossings", *IEEE Transactions on Communications*, Vol. COM-31, pp. 503–508, 1983.

[15] S.K. Ray, "Zero-crossing-based approximate demodulation of wide-deviation FM", *IEE Proceedings*, Vol. 131, pp. 47–51, 1984.

[16] K.S. Shanmugam, *Digital and Analog Communication Systems*, New York John Wiley & Sons, 1985.

[17] S. Siegel, *Nonparametric Statistics for the Behavioral Sciences*, New York McGraw-Hill, 1956.

[18] W.J. Conover, *Practical Nonparametric Statistics*, 2nd ed., New York John Wiley & Sons, 1980.

[19] H.E. Rowe, *Signals and Noise in Communication Systems*, New York Van Nostrand, 1965.

[20] M.A. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, New York Dover, 1964.