Noncoherent Block Demodulation of MSK with Inherent and Enhanced Encoding

Harry Leib, Member, IEEE, and Subbarayan Pasupathy, Fellow, IEEE

Abstract—The study considers noncoherent demodulation and coding of minimum shift keying (MSK). Using the framework of noncoherent maximum-likelihood block estimation (N-MLBE), it is shown that there is a close relation between known noncoherent MSK demodulation structures, such as envelope and differential receivers, and schemes derived from the N-MLBE principle. When the observation interval is increased, the performance of MSK with N-MLBE tends to that of binary coherently detected orthogonal BFSK. Furthermore, a new demodulation strategy, reduced block noncoherent estimation (RBNE) is introduced and shown to improve the performance of noncoherent MSK, beyond that of coherently detected orthogonal FSK. When RBNE is used with MSK and the observation interval is increased, the performance approaches that of antipodal signaling. The key feature that makes all these structures give an improved performance is multi-symbol noncoherent processing with exploitation of the MSK inherent coding properties induced by its phase continuity. Further performance improvement by the use of binary block codes is considered for MSK with N-MLBE. It is shown that simple binary block codes with low bandwidth expansion which exploit the inherent MSK memory can give significant gains.

I. INTRODUCTION

M

inimum shift keying (MSK) is a form of continuous phase binary frequency shift keying (BFSK) with modulation index 0.5 [1], [2]. Because of the phase continuity, MSK has an inherent memory or redundancy and hence can be viewed as a rate 1/2 coding scheme [3]. The role that the phase continuity plays in coherent communication schemes and the coding features that it induces have been extensively investigated in the coded-modulation literature ([3], [4] and references), however, its role in a noncoherent system does not seem to have attracted much attention. This paper focuses on noncoherent MSK, as a first step in understanding the coding aspects of the phase continuity in noncoherent coded-modulation.

In many of the applications where MSK is considered [5]–[8], problems due to fading, Doppler shifts, etc., suggest that noncoherent demodulation (i.e., demodulation without carrier-phase synchronization) may be more desirable. As a result, various structures such as envelope receivers, differential receivers and FM discriminator receivers have been proposed for noncoherent demodulation of MSK [8]–[10]. Since these structures were developed independently, the interrelations between them and optimal structures are not clear; this serves as another motivation for the present study. Recently, [11]–[13], it has been shown that noncoherent demodulation for M-ary phase shift keying (MPSK) can perform as well as coherent detection. Using a detection theory framework, optimal noncoherent demodulation structures for MPSK have been derived that perform as well as coherent detection. The main vehicle through which this improved performance has been obtained is multi-symbol noncoherent crosscorrelation which exploits the fact that the unknown carrier phase is time invariant over the observation interval. Could similar methods work for MSK as well?

In this work we explore the possibility of improving the performance of noncoherently detected MSK by using a detection theory framework. It is shown that multi-symbol noncoherent crosscorrelation is not enough to bring the performance close to that of ideal coherent demodulation of MSK (antipodal signaling). The phase continuity of MSK is a key factor which must be exploited even in a noncoherent detection setting, and this work presents one possible method to do so. Phase continuity induces coding features into this modulation format which when properly utilized can bring the performance close to antipodal signaling even with noncoherent demodulation. These inherent coding features can be enhanced by additional coding; and hence, this work considers the use of binary block coding with noncoherent MSK and shows how to construct simple high rate codes which give significant gain.

The results of this paper can be extended to other modulation schemes that have inherent coding properties due to phase continuity, such as Continuous Phase Modulation, or due to
II. NONCOHERENT BLOCK ESTIMATION OF MSK

When viewed as digital FM modulation, the complex envelope of an MSK signal \( s(t) \) can be expressed as:

\[
\tilde{s}(t) = e^{j\phi(t)}, \quad -\infty < t < \infty
\]

\( \phi(t) \) is the information conveying phase, namely,

\[
\phi(t) = \pi \sum_{n=-\infty}^{\infty} a(n) q[t - (n-1)T], \quad -\infty < t < \infty \tag{1}
\]

with phase function \( q(t) \) given by \( q(t) = \begin{cases} 0, & t \leq 0 \\ t/T, & 0 < t \leq T \\ 1/2, & t > T \end{cases} \).

The fact that \( q(t) = 1/2 \) for \( t \geq T \) constrains the initial phase in each symbol interval to be identical to the final phase in the preceding interval, which induces memory or coding features into this modulation format. The binary equiprobable information symbols \( a(n) = \pm 1 \) are transmitted at a rate of 1/T. It is assumed that the channel is modeled as additive white Gaussian. The complex envelope of the received signal is \( \tilde{r}(t) = e^{j\phi(t)} + \tilde{w}(t) \) where \( \tilde{w}(t) \) is complex white Gaussian noise with two-sided power spectral density \( N_0 [\text{W/Hz}] \).

When MSK is used with noncoherent block demodulation, the decision variables of the optimum noncoherent receiver (optimum in the minimum probability of block error sense) with observation interval \((0,NT)\), \( N = 1,2,\cdots \) are given by [14, p. 291],

\[
|X[a_i(N)]| = \left| \int_{0}^{NT} \tilde{r}(t)e^{-j\phi(t)}a_i(N) dt \right|, \quad i = 1,2,\cdots,2^N \tag{2.a}
\]

where

\[
\phi[t; a_i(N)] = \phi(0) + \pi \sum_{n=1}^{N} a_i(n) q[t - (n-1)T], \quad 0 \leq t \leq N \tag{2.b}
\]

is the information-conveying phase generated by the \( N \)-tuple

\[ a_i(N) = [a_i(1), a_i(2), \cdots, a_i(N)], \quad a_i = \pm 1. \]

The receiver chooses the sequence \( a_i(N) \) that maximizes \( |X[a_i(N)]| \) and thus uses a noncoherent maximum likelihood block estimation (N-MLBE) strategy.

It is seen from (2) that the initial phase \( \phi(0) \) of the MSK waveforms does not affect the decision variables \( |X[a_i(N)]| \). This initial phase reflects the phase continuity of the modulation format outside the observation interval. When properly exploited with coherent demodulation, the phase continuity gives a significant increase in performance equivalent to that of BPSK [1]. With N-MLBE only the phase continuity within the observation interval matters. Therefore we assume, \( \phi_i(0) = 0 \) without loss of generality.

In order to gain some insight into the structure of N-MLBE we write (2) as

\[
|X[a_i(N)]| = \left| \sum_{n=1}^{N} x[n, a_i(n)] \right|
\]

\( x[n, a_i(n)] = \int_{(n-1)T}^{NT} \tilde{r}(t)e^{-j\phi(t)}a_i(n) dt, \quad n = 1,2,\cdots,N \) and \( a_i(n) = \{a_i(1), \cdots, a_i(n)\} \). Over the interval \((n-1)T < t \leq nT\) we have \( \phi[t; a_i(N)] = (\pi/2) \sum_{k=1}^{n} a_i(k) + \pi a_i(n)(t - (n-1)T)/(2T) \) where \( \sum_{k=1}^{n} a_i(k) = 0 \). Therefore,

\[
x[n, a_i(n)] = |x[n, a_i(n)]|e^{j\theta[n, a_i(n)]}
\]

where \( \theta[n, a_i(n)] = \arg\{x[n, a_i(n)]\} = \int_{(n-1)T}^{nT} \tilde{r}(t)e^{-j\phi(t)}a_i(n) dt \) and \( y[n, a_i(n)] = \int_{(n-1)T}^{nT} \tilde{r}(t)e^{-j\phi(t)}a_i(n) dt \). Substituting (4.a) in (3) the decision variables \( |X[a_i(N)]| \) can be expressed in terms of the single symbol correlations \( y[n, a_i(n)] \) as

\[
|X[a_i(N)]| = \sum_{n=1}^{N} \exp \left\{ -j(\pi/2) \sum_{k=1}^{n-1} a_i(k) \right\} y[n, a_i(n)] \tag{4.b}
\]

Substitution of (5)–(6) in (3) yields

\[
|X[a_i(N)]|^2 = \sum_{n=1}^{N} |y[n, a_i(n)]|^2 + 2 \sum_{n=1}^{N} \sum_{m=n+1}^{N} |y[n, a_i(n)]| |y[m, a_i(m)]| \cos \left\{ \alpha[m, a_i(m)] - \alpha[n, a_i(n)] - (\pi/2) \sum_{k=1}^{n-1} a_i(k) \right\}. \tag{7}
\]

Let us consider (7) for some particular cases. For \( N = 1 \) we have \( |X[a_i(1)]|^2 = |y[1, a_i(1)]|^2, i = 1,2 \). The decision rule \( |y[1, 1]|^2 < |y[1,-1]|^2 \) can be easily recognized as
Fig. 1. Structure of the $N$-MLBE of MSK receiver.

the classical noncoherent BFSK receiver, which uses only the envelopes of the single symbol correlations (i.e., matched filtering and envelope detection). For $N = 2$, (7) becomes

$$
|X[a_i(2)]|^2 = |y[1, a_i(1)]|^2 + |y[2, a_i(2)]|^2 \\
+ 2|y[1, a_i(1)]||y[2, a_i(2)]||
$$

$$
\cdot \cos \{a[2, a_i(2)] - a[1, a_i(1)] - (\pi/2)a_i(1)\}
$$

where $i = 1, \ldots, 4$. This receiver for $N = 2$ uses the phase differences in a form of differential demodulation. It also uses the magnitudes of the correlators output, $|y[1, a_i(1)]|$, $|y[2, a_i(2)]|$, since in MSK, the information is also carried by these magnitudes. For a modulation format like PSK where all the information is contained only in the phases of the decision variables, this structure reduces to the classical differential coherent receiver [15, p. 230]. For $N = 3$ the decision variables are

$$
|X[a_i(3)]|^2 = \sum_{n=1}^{3} |y[n, a_i(n)]|^2 + 2|y[2, a_i(2)]||y[1, a_i(1)]||
$$

$$
\cdot \cos \{a[2, a_i(2)] - a[1, a_i(1)] - (\pi/2)a_i(1)\} \\
+ 2|y[3, a_i(3)]||y[2, a_i(2)]||
$$

$$
\cdot \cos \{a[3, a_i(3)] - a[2, a_i(2)] - (\pi/2)a_i(2)\} \\
+ 2|y[3, a_i(3)]||y[1, a_i(1)]||
$$

$$
\cdot \cos \{a[3, a_i(3)] - a[1, a_i(1)] - (\pi/2)a_i(1) - (\pi/2)a_i(2)\}
$$

which are generated by using the first-order phase differences $a[2, a_i(2)] - a[1, a_i(1)], a[3, a_i(3)] - a[2, a_i(2)]$, as well as the second-order phase difference, $a[3, a_i(3)] - a[1, a_i(1)]$.

Therefore, for an observation interval of three symbols we have multi-differential as well as envelope detectors; and, the relation between the $N$-MLBE principle and envelope and differential structure becomes clear. The performance gain that can be obtained by increasing the observation interval is considered next.

III. PERFORMANCE ANALYSIS OF N-MLBE DEMODULATION OF MSK

First, let us consider a simplified case where only two information blocks could be transmitted over the observation interval $(0, NT]$, $a_i(N) = [a_i(1), \ldots, a_i(N)]$ and $a_m(N) = [a_m(1), \ldots, a_m(N)]$. The pairwise probability of error $P_e(i/m)$ is the probability that when $a_m(N)$ was transmitted, the demodulator decides in favor of $a_i(N)$, and can be obtained by applying the theory of noncoherent demodulation (for example, [14, p. 293]) to the MSK format as has been done in [16]–[18]. It is shown in Appendix A that for large SNR ($\gamma$) and $|\rho_{im}| > 0$, $P_e(i/m)$ has the following asymptotic form:

$$
P_e(i/m) \approx \sqrt{[N - d^2_{NC}(i, m)/2]/[N - d^2_{NC}(i, m)]}
$$

$$
\cdot Q[\sqrt{\gamma} d_{NC}(i, m)]
$$

(8)

where $\gamma = T/(2N\sigma)$ is the SNR,

$$
d_{NC}(i, m) = N(1 - |\rho_{im}|)
$$

(9)

and $\rho_{im} = 1/(NT) \int_0^T e^{j[\omega_m(t) - \omega_i(t)]}dt$ is the crosscorrelation coefficient between two MSK signals. From (8)
it is seen that $P_e(i/m)$ is characterized by the parameter $d_{NC}(i,m)$ which depends on the two information blocks, $a_i(N), a_m(N)$. Therefore $d_{NC}(i,m)$ can be interpreted as a distance-type function which plays the same fundamental role as the Euclidean distance does for coherent detection [17]. When $|\rho_{im}| = 0$, we have noncoherently detected orthogonal signals and the exact result $P_e(i/m) = (1/2)\exp(-\gamma N/2)$.

The magnitude of the crosscorrelation is

$$|\rho_{im}| = (1/N) \left| \sum_{k=1}^{N} \rho_{im}(k) \right| \tag{10a}$$

where

$$\rho_{im}(k) = \text{sinc}([a_i(k) - a_m(k)]/4) \times \exp\left\{-(\pi/2) \sum_{n=1}^{k-1} [a_i(n) - a_m(n)] \right\} + (\pi/4)[a_i(k) - a_m(k)] \tag{10b}$$

and \(\text{sinc}(x) = \sin(\pi x)/\pi x\). It is seen from (10b) that \(\rho_{im}(k), \rho_{im}(k)\) can be purely real or purely imaginary. When \(a_i(k) = a_m(k), \rho_{im}(k)\) is purely real with magnitude 1. When \(a_i(k) \neq a_m(k), \rho_{im}(k)\) is purely imaginary with a magnitude of \(2/\pi\). Therefore $|\rho_{im}|$ has a form,

$$|\rho_{im}| = (1/N)\sqrt{(N - q_R)^2 + q_I^2}/\pi^2, \quad q_I \leq w_{im} \leq q_R \leq N \tag{11}$$

where $q_R, q_I$ are integers, and $w_{im}$ is the Hamming distance between $a_i(N)$ and $a_m(N)$. Since $N(\text{Re}\{\rho_{im}\}) = N - q_R$ and $N(1 - \text{Re}\{\rho_{im}\}) = d_{EC}^2(i,m)$ where $d_{EC}(i,m)$ is the Euclidean distance between the MSK waveforms that are generated by $a_i(N)$ and $a_m(N)$ we have

$$q_R = d_{EC}^2(i,m), \quad q_I = \left\{ \begin{array}{ll}
2N - d_{EC}^2(i,m), & d_{EC}^2(i,m) \leq N \quad (\text{Re}\{\rho_{im}\} \geq 0), \\
2N - d_{EC}^2(i,m), & d_{EC}^2(i,m) > N \quad (\text{Re}\{\rho_{im}\} < 0) \end{array} \right. \tag{12}$$

We call the quantity $d_{EC}^2(i,m)$ the circular square Euclidean distance for obvious reasons. For MSK $d_{EC}^2(i,m)$ is an integer and $w_{im} \leq d_{EC}^2(i,m) \leq N$. Table 1 presents a list of error patterns (difference codeword pairs where the 2 and -2 components are replaced by 1) with corresponding distances $d_{EC}^2$.

From (11) and (12) we have the upper bound

$$|\rho_{im}| \leq (1/N)\sqrt{(N - d_{EC}^2(i,m))^2 + d_{EC}^2(i,m)4}/\pi^2 \tag{13a}$$

and hence, from (9) and (13a), we have the lower bound for $d_{NC}^2(i,m)$,

$$d_{NC}^2(i,m) \geq N - \sqrt{(N - d_{EC}^2(i,m))^2 + d_{EC}^2(i,m)4}/\pi^2 = D_0^2 \left[ d_{EC}^2(i,m), N \right]. \tag{13b}$$

This lower bound is always achieved by some error patterns and therefore it is closely related to a bound in performance. The function $D_0^2(q, N)$ is illustrated in Fig. 2(a), and its properties are investigated in Appendix B. It is seen that by increasing the block-length $N$ it is impossible to increase the minimum of $D_0^2$ beyond 0 dB (= 1). However, if we could find a way to eliminate the error patterns which correspond to $D_0^2[1, N]$, then the minimum of $D_0^2$ could be increased to 3 dB (= 2) by increasing the block-length $N$. This is the basis of a technique for improving the SNR performance of noncoherent MSK which is explored in Section IV. Furthermore if it would be possible to eliminate error patterns which correspond to $D_0^2[q, N]$ also for larger values of $q$ then the minimum value of $D_0^2$ can be further increased with $N$. A technique that achieves this task by using coding is considered in Section V.

Since $|\rho_{im}| \geq (N - d_{EC}(i,m))/N$ we have $d_{NC}(i,m) < d_{EC}^2(i,m)$ which with (13b) gives

$$D_0^2 \left[ d_{EC}^2(i,m), N \right] \leq d_{NC}^2(i,m) < d_{EC}^2(i,m). \tag{14}$$

Appendix B shows that $D_0^2 \left( d_{EC}^2, N \right)$ increases with $N$, and

$$\lim_{N \to \infty} D_0^2 \left( d_{EC}^2, N \right) = d_{EC}^2. \tag{14}$$

Therefore, when $N \to \infty$ we have from (14) $D_0 \to d_{NC}$ which shows that for any error pattern (13b) is tight as $N$ increases.

Now we investigate the rate of convergence of $D_0^2 \left( d_{EC}^2, N \right)$

<table>
<thead>
<tr>
<th>Error patterns</th>
<th>No. of errors</th>
<th>$C(d_{EC}^2, N)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$100000000000$</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>$000000000000$</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>$110000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$101000000000$</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$000000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$000000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$110000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$101000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$010000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$000000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$110000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$100000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$000000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$110000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$101000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$010000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$000000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$110000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$101000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$010000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$000000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$110000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$101000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$010000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$000000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$110000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$101000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$010000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$000000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$110000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$101000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$010000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$000000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$110000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$101000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$010000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$000000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$110000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$101000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$010000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$000000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$110000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$101000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$010000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$000000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$110000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$101000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$010000000000$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$000000000000$</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 1: Noncoherent MSK Error Patterns
izes quite well the performance of noncoherent demodulation of waveforms with small $d_{EC}$.

An asymptotic (high SNR) upper bound to $P_e(d_{EC}, N)$, the pairwise probability of error between sequences with circular Euclidean distance $d_{EC}$ which is tight with increasing $N$ can be derived by using (8), (13.b) and a result from Appendix B which states that $D_2^2(d_{EC}, N) < N/2$, yielding

$$P_e(d_{EC}, N) \lesssim \sqrt{\frac{[N - D_2^2(d_{EC}, N)]/2}{[N - D_2^2(d_{EC}, N)]}} 
Q[\sqrt{D_2(d_{EC}, N)}] \lesssim \frac{3/2}{Q[\sqrt{D_2(d_{EC}, N)}]}.$$ (17)

The probability of block error when all the sequences $a_i(N)$, $i = 1, 2, \cdots, 2^N$ could be transmitted, can be upper bounded by the union bound employed with (17) which yields

$$P_e(N) \lesssim \frac{3/2}{Q[\sqrt{D_2(d_{EC}, N)}]} \sum_{q=1}^{2^N} C(q, N)Q[\sqrt{D_2(d_{EC}, N)}].$$ (18)

where $C(q, N)$ is the number of distinct error patterns that can be generated from information block pairs $a_i(N)$, $a_m(N)$ with $d_{EC}(i, m) = q$.

In order to find $C(q, N)$ (18) one has to consider the effect of the error location on the crosscorrelation given in (10.a, b) and (11). The following can be verified: i) After an even number of successive errors, the real components $p_{im}(k)$ in (10.a, b) do not change sign. ii) After an odd number of successive errors, these components change sign making $y_s$ larger than $w$ in (11). Therefore, a high probability error pattern can not include clusters of odd successive errors in the middle of the block; they can only occur at the edges. Based on these observations $C(q, N)$ has been calculated for $q = 1, 2, 3$ and it is presented in Table I.

The bound of (18) can be further simplified by taking only the largest term in the sum, the one which corresponds to

$$q_0 = \min_{1 \leq q \leq N} \{D_2^2(q, N)\} = D_2^2(q_0, N)$$

where $q_0$ is the value of $d_{EC}$ that corresponds to the error pattern which limits the performance. Thus,

$$P_e(N) \lesssim \frac{3/2}{Q[\sqrt{D_2(d_{EC}, N)}]} \sum_{q=1}^{2^N} C(q_0, N)Q[\sqrt{D_2(d_{EC}, N)}].$$

Appendix B shows that $D_2^2(q, N)$ has a single maximum in the interval $1 \leq q \leq N$; therefore, its minimum is at $q = 1$ or at $q = N$ and

$$D_2^2(q_0, N) = \min \{D_2^2(1, N), D_2^2(N, N)\} = \begin{cases} D_2^2(N, N) = ND_2^2(1, 1), & N = 1, 2, 3 \geq 3 \end{cases}$$ (19.a)

which shows that

$$q_0 = \begin{cases} N, & N = 1, 2, 3 \geq 3 \end{cases}.$$ (19.b)

From (19.b) we see that error patterns with $d_{EC} = 1$ limit the performance for $N \geq 3$. With the values of $C(q, N)$ from Table I we have the tight asymptotic bound,

$$P_e(N) \lesssim \begin{cases} \sqrt{\frac{3/2}{Q[\sqrt{D_2(d_{EC}, 1)}]}}, & N = 1, 2, \sqrt{\frac{6Q[\sqrt{D_2(d_{EC}, 1)}]}{Q[\sqrt{D_2(d_{EC}, 1)}]}}, & N \geq 3 \end{cases}.$$ (20)
be approximated by
which is the performance of MSK with one symbol coherent
error. Since this binary system is identical to the bit error probability can
quite well the probability of symbol error as well.

Table III gives the asymptotic degradation
with respect to coherent orthogonal BFSK.

<table>
<thead>
<tr>
<th>N</th>
<th>degradation [dB]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.4</td>
</tr>
<tr>
<td>2</td>
<td>1.4</td>
</tr>
<tr>
<td>3</td>
<td>0.5</td>
</tr>
<tr>
<td>4</td>
<td>0.3</td>
</tr>
<tr>
<td>5</td>
<td>0.3</td>
</tr>
<tr>
<td>6</td>
<td>0.2</td>
</tr>
</tbody>
</table>

So far we have considered only the probability of block error \( P_e(n) \). The symbol error probability \( P_{se}(N) \), in this binary system is identical to the bit error probability can be approximated by \( P_{se}(N) \approx (w_a/N)P_e(N) \) where \( w_a \) is the average Hamming weight of the error patterns with \( d_{EC} = q_a \).

Since \( w_a \approx q_a \), and with (19.2) we see that (20) characterizes quite well the probability of symbol error as well.

Fig. 3 illustrates the probability of error (20) as a function of SNR. It is seen that when \( N \) is increased the probability of error approaches that of coherently detected orthogonal BFSK which is the performance of MSK with one symbol coherent demodulation [1]. Table II gives the asymptotic degradation with respect to coherent BFSK and shows the performance improvement of 3 dB when the observation interval is increased from one symbol interval to two. With an observation interval of three symbols N-MLBE of MSK is approximately within 0.5 dB of coherent orthogonal BFSK. From Fig. 3 and Table II it is seen that at an error probability of \( 10^{-6} \) the SNR performance is within 0.2 dB from the asymptotic (very low error probability) performance.

The main result of this section is that MSK when noncoherently demodulated over several symbol intervals can be made to perform as well as coherent orthogonal FSK, which is the performance of MSK when coherently demodulated over one symbol interval. It is well known [1] that coherent demodulation of MSK with exploitation of phase continuity yields the performance of antipodal signaling (up to a multiplier constant of 2). This gives a 3 dB improvement in SNR with respect to single symbol coherent demodulation where the phase continuity is not exploited. The techniques that exploit phase continuity in a coherent demodulation context are based on multiple symbol processing (for example, by using Viterbi algorithm based methods). Can phase continuity, or the coding features induced by phase continuity be exploited with noncoherent demodulation? This subject is addressed in Section IV.

IV. REDUCED BLOCK NONCOHERENT ESTIMATION (RBNE) OF MSK

One of the most powerful analytical tools in communication theory is that of signal spaces. In coherent communications, the pairwise probability of error depends on the Euclidean distance (which is a true mathematical distance induced by the inner product or crosscorrelation) and the signal space with this distance forms a Hilbert space. Analysis and synthesis of coherent communication systems can be reduced in many important cases to consideration of geometrical features of this Hilbert space. The use of similar methods for noncoherent communication systems is hampered by the fact that the pairwise probability of error depends on the noncoherent distance \( d_{NC} \), (9), which is not a proper mathematical distance. Therefore, the signal space with the noncoherent distance \( d_{NC} \) is not a Hilbert space and it does not present a very powerful analytical tool such as for coherent communication. Nevertheless, by using (12), (13), and (14) one can use the noncoherent distance \( d_{NC} \) to formulate appropriate constraints in terms of the Euclidean distance \( d_E \), thus making it possible to attach "noncoherent geometrical properties" to the signal space.

The main use of the noncoherent distance \( d_{NC} \) is an indication about the pairwise probability of error. The "closer" the two signals are with respect to \( d_{NC} \) the larger is their pairwise probability of error with noncoherent detection. Two signals are close with respect to the noncoherence distance \( d_{NC} \) if their Euclidean distance satisfies: \( d_{EC}^2 < N \) or \( d_{EC}^2 > N \) which can be written in terms of their circular Euclidean (12) distance as: \( d_{EC}^2 << N \). One of the conclusions of Section III is that for MSK with N-MLBE with \( N \geq 3 \), the performance is limited by error patterns with \( d_{EC}^2 = 1 \), which correspond to single error patterns with an error in the first or last symbol (i.e., those with \( d_{EC}^2 = 1 \) which are highly correlated, or \( d_{EC}^2 = 2N - 1 \) which are highly anticorrelated). Highly correlated as well as highly anticorrelated waveforms are close in terms of the noncoherent distance [also seen from (9)]. This shows that the MSK waveforms are located in clusters of four with respect to the distance \( d_{NC} \) (Fig. 3). The clustering property is also reflected in (18) by \( C(1, N) = 2 \) (i.e., the number of nearest neighbors under the noncoherent distance \( d_{NC} \) is independent of the block length \( N \)). The four signals in the cluster share the same information bits except for the first and last ones and this provides the basis of an improved method for noncoherent MSK demodulations. The clustering property is another manifestation of the inherent coding of MSK, which can be exploited to improve performance.

During an observation interval of \( 0 < t \leq NT(N \geq 3) \), the sequence of transmitted information symbols is \( a(N) = \ldots \)
Fig. 4. Noncoherent MSK signal space structure. Each cluster of four signals corresponds to one of the $2^{N-2}$ binary $N-2$ tuples $[a(2), \ldots, (N-1)]$.

$[a(1), a(2), \ldots, a(N)]$. Assume that we are interested in estimating $a'(N) = [a(2), \ldots, a(N-1)]$ only, i.e., we are estimating only the middle $(N-2)$ symbols out of the $N$ observed symbols. This is equivalent to estimating the cluster and not the particular signal in the cluster. The probability of cluster error is lower than the probability of signal error because the distance between clusters is larger than the distance between signals within a cluster. One can use a composite hypothesis approach for deriving the optimal receiver as in [18], however, this yields a very complicated structure using exponential nonlinearities. A somewhat simpler receiver structure results by using a suboptimal strategy. After obtaining the decision variables, $[X_i(N)]$ as in (2), the decision is made in favor of $a' = [a(N), a(N-1), \ldots, a(2)]$ if one of the four sequences $[\pm 1, a(2), \ldots, a(N-1), \pm 1]$ maximizes $|X_i(N)|$. It can be shown, using the same arguments as in [18], that the performance of this reduced block noncoherent estimation (RBNE) of MSK for large SNR is the same as the optimal demodulation based on composite hypothesis testing.

The analysis of RBNE demodulation follows the same lines as Section III. We start with (18) and note that the error patterns with $d_{EC}^2 = 1$ are not considered errors here; thus the bound of (18) applies with $C(1, N) = 0$. Along similar lines as in Section III we can simplify this bound by using only the largest term in the sum, the term that corresponds to

$$
\min_{2 \leq s \leq N} \left\{ D_s^2(q, N) \right\} = \begin{cases} 
N D_s^2(1, 1), & N = 3, 4 \\
2 D_s^2(1, N/2), & N \geq 5
\end{cases}
$$

(21)

and finally we have,

$$
P_e(N) \lesssim \begin{cases} 
\sqrt{3/2} Q \left[ \sqrt{\gamma N D_s(1, 1)} \right], & N = 3, 4 \\
\sqrt{3/2} N Q \left[ \sqrt{2\gamma N D_s(1, N/2)} \right], & N \geq 5
\end{cases}
$$

(22)

Comparison of (22) with (20) shows that the memory of MSK when correctly utilized (even in a noncoherent setting) can give an improvement of 3 dB in performance, which is the same as that obtained for coherent MSK when the phase continuity is exploited [1].

Fig. 5 illustrates the probability of error (22) as a function of SNR. It is seen that the performance of MSK with RBNE can be made close to antipodal signaling. With ideal coherent demodulation which exploits memory, the performance of MSK is equivalent to antipodal signaling up to a multiplier factor of 2 (due to the inherent differential encoding of MSK) [1]. This multiplier factor is equivalent to a degradation of about 0.5 dB at a probability of error of $10^{-6}$. Therefore, from Fig. 5 we see that MSK with RBNE and $N = 8$ is within 0.5 dB from ideal coherent MSK at practical values of probabilities of error. Table III presents the asymptotic degradation of RBNE of MSK with respect to antipodal signaling. It is seen that with only a block length of three symbols RBNE is within 2.6 dB from antipodal signaling, and when the block length is increased to six symbols the asymptotic degradation is only 0.5 dB. Thus, we see that the block length needed for RBNE in order to have negligible degradation with respect to coherent demodulation is relatively short. Furthermore, for $N \geq 8$ the performance of MSK with RBNE at a probability of error of $10^{-6}$ is within 0.2 dB from its asymptotic (very low probability of error) performance.

Basically, the RBNE strategy ignores the first and last symbols of the observation interval. One could inquire if any additional improvement can be obtained by ignoring more than one symbol at the edges of the observation interval. From Section III we know that the error patterns with $d_{EC}^2 = 2$ can be located anywhere in the block. Therefore, by ignoring more than one symbol at each edge we can not eliminate the effect of these error patterns. By ignoring more than one symbol at each edge, $C(2, N)$ of (18) cannot be made

![Figure 5. Probability of error for MSK with RBNE.](image_url)
zero; it can only be reduced. This is reflected also in the signal space structure of noncoherent MSK. From Fig. 4 we see that if we define a cluster of more than four signals then the distance between these multi-symbol clusters is not larger than the distance between the basic clusters of four symbols. The minimum value of $C(2, N)$ is 2 and this is obtained when we estimate only the middle symbol from the entire block (i.e., one symbol estimation based on $N$ symbol observation, as considered in [16], [17], [18]). The performance of one symbol estimation can be obtained from (22) by letting the multiplicative coefficient $N$ be 2. This shows that the performance of one symbol estimation is asymptotically equivalent to our RBNE strategy, and only for large block lengths $N$ it gives a noticeable SNR improvement over RBNE. However this improvement is counter balanced by an increase in complexity, since only one symbol is estimated at a time, while with RBNE we estimate $N - 2$ symbols.

We conclude this section by comparing several structures for noncoherent MSK demodulation, that have been reported earlier, with the results of this work. These structures are presented in Table IV. From lines 1–3 and 7 of Table IV we see that with an unmatched predetection filter alone and without any postdetection processing, the degradation can not be made less than 3 dB with respect to ideal coherent MSK (i.e., the performance only approaches that of orthogonal signaling). In general, a predetection filter reduces the noise effects on the decision variable at the expense of intersymbol interference. The best trade-off is obtained when a matched filter is used. Because of the phase continuity of MSK, even with a matched predetection filter there is an inherent interference between quadrature components. An attempt to remove this interference by a zero forcing equalizer (inverse filter) has been presented in [9], line 4 in Table IV. With an infinite length equalizer and a single differential detector the degradation is only 1 dB; however, the performance will degrade with a decrease in the length (and hence the observation interval) of the equalizer. Increasing the number of differential detectors as shown in Section II could further decrease the degradation of this method for noncoherent demodulation of MSK. The approach of [22], line 5 in Table IV, uses a single error correcting scheme as postdetection processing. Under the assumption of ideal predetection filtering (i.e., matched filtering without any interference) the performance has been shown to be within 0.3 dB from ideal coherent MSK. However this kind of filtering is impossible with MSK, see [9]. Nevertheless, [22] is an important contribution since it shows explicitly the advantages of exploiting the coding features induced by phase continuity. A scheme similar to [22], but with a predetection filter of the equalized fourth order Butterworth type, line 6 in Table IV, gives a degradation of 2.2 dB. The same scheme without error correction, line 7 in Table IV, is within 3.5 dB from ideal coherent MSK.

Attempts to use the memory of MSK by a postdetection Viterbi algorithm, lines 8 and 9 in Table IV, have resulted in relatively poor performance. This is due to insufficient statistics provided by a single limiter-discriminator or differential detector. The optimal receiver of Section II consists of combinations of envelope and multiple symbol differential detectors. Only such combinations can provide decision variables which are sufficient statistics with the coding features induced by phase continuity. Another approach, line 10 of Table IV, is based on predetection filtering and limiter-discriminator detection with decision feedback that reduces the effects of the interference at the output of the predetection filter. The performance of this structure is 2 dB from ideal coherent MSK, when error propagation effects are neglected.

From this short comparison we see that our RBNE demodulation strategy is asymptotically optimal (for high SNR) in the sense that its performance approaches that of ideal coherent MSK when $N \rightarrow \infty$. However other noncoherent MSK demodulation methods can be found in the literature with performance that is close to ideal coherent MSK. For a finite observation interval we couldn’t find any MSK demodulation structure to compare with except the one proposed in [10]. This MSK demodulation structure is based on matched filtering and differentially coherent detection which uses (in a nonoptimal form) the MSK memory; therefore it is a nonoptimal noncoherent demodulation structure which uses an observation

<table>
<thead>
<tr>
<th>No.</th>
<th>Ref.</th>
<th>Predetection filter</th>
<th>Detection: limiter-discriminator(D)</th>
<th>Postdetection processing</th>
<th>Degradation in dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[19]</td>
<td>Gaussian, $BT = 1$</td>
<td>LD and $D$</td>
<td>None</td>
<td>4.3</td>
</tr>
<tr>
<td>2</td>
<td>[20]</td>
<td>Gaussian, $BT = 1.2$</td>
<td>$D$</td>
<td>None</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>[21]</td>
<td>Spectral raised cosine</td>
<td>$D$</td>
<td>None</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>[9]</td>
<td>Matched filter and zero forcing equalizer</td>
<td>$D$ three symbol delay</td>
<td>None</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>[22]</td>
<td>Matched filter without any interference</td>
<td>$D$ single and double symbol delays</td>
<td>Single error correction</td>
<td>0.3</td>
</tr>
<tr>
<td>6</td>
<td>[6]</td>
<td>Butterworth $BT = 1.1$</td>
<td>$D$ single and double symbol delays</td>
<td>Single error correction</td>
<td>2.2</td>
</tr>
<tr>
<td>7</td>
<td>[6]</td>
<td>Butterworth $BT = 1.1$</td>
<td>$D$ single and double symbol delays</td>
<td>None</td>
<td>3.5</td>
</tr>
<tr>
<td>8</td>
<td>[21]</td>
<td>Raised cosine</td>
<td>$LD$</td>
<td>Viterbi algorithm</td>
<td>5.5</td>
</tr>
<tr>
<td>9</td>
<td>[22]</td>
<td>None</td>
<td>$D$ optimized delay</td>
<td>Viterbi algorithm</td>
<td>3.5</td>
</tr>
<tr>
<td>10</td>
<td>[24]</td>
<td>Raised cosine</td>
<td>$LD$</td>
<td>Decision feedback</td>
<td>2</td>
</tr>
</tbody>
</table>

TABLE IV

Comparison of Noncoherent Demodulation Structures of MSK. The Degradation is with Respect to Antipodal Signaling.
interval of two symbols. The asymptotic degradation with respect to ideal coherent MSK is 5.3 dB. From Table II we have that the asymptotic degradation with respect to ideal coherent MSK of the N-MLBE receiver with $N = 2$ is 4.4 dB. This shows an advantage of 0.9 dB of the N-MLBE approach. From the references of Table IV and from our work it is seen that the ingredients of a good noncoherent MSK demodulation structure are: i) matched predetection filtering, ii) detection by combinations of multiple differential and envelope detectors, iii) postdetection processing that exploits the coding features of MSK. Our conjecture is that these three ingredients are necessary for good noncoherent demodulation of other continuous phase modulation schemes as well.

V. CODING WITH NONCOHERENT BLOCK DEMODULATION OF MSK

The improved performance of the RBNE strategy is due to the fact that the most probable error-patterns of N-MLBE (i.e., single-error patterns with $d_{EC}^2 = 1$) are eliminated and thus the coefficient $C(1, N)$ of (18) is zero. In this sense RBNE can be viewed as a structure that exploits the (most probable) single-error correction capability of MSK. This demodulation method makes use of the inherent coding features of MSK, however, the performance cannot be made better than ideal coherent MSK. In order to further improve performance one has to enhance these inherent coding features. This can be done by additional (external) coding.

The encoded noncoherent MSK system that is considered in this work is presented in Fig. 6. Prior to MSK modulation we use an $(n_c, k_c, w_c)$ binary block encoder, which encodes blocks of $k_c$ information bits into codewords of $n_c$ bits with minimum Hamming distance $w_c$. These encoded bits are modulated by an MSK modulator and transmitted over the channel. The noncoherent MSK demodulator uses an observation interval identical to the codeword length, $n_c = N$. It correlates the complex envelope of the received signal over a demodulation interval of $N$ symbols with each of the $2^{k_c}$ possible transmitted MSK waveforms induced by the binary blocks. The decision is in favor of the information block that maximizes this crosscorrelation. Since $k_c < n_c$, this type of coding involves a bandwidth expansion of $1/R$ where $R = k_c/n_c$ is the code rate. For this type of coding to be practical we request a relatively small $k_c$, which guarantees a reasonable receiver complexity, and large $R$ which guarantees low bandwidth expansion.

An asymptotic upper bound to the block error probability with coding can be found from (18) and it is given by

$$P_{bc}(N) \leq 3/2 \sum_{q=q_c}^{N} C_c(q, N) \left[ \sqrt{\gamma_b} R D_{q}(N, N) \right]$$

(23)

where $\gamma_b = \gamma / R$ is the normalized SNR with respect to the code rate $R$.

The performance measure used in this paper is $G_{\infty}$, the asymptotic coding gain with respect to uncoded MSK with the same demodulation strategy and the same complexity. For MSK with N-MLBE demodulation and with fixed complexity $k_c$ we have from (20) and (23) for $N \geq 3$ $G_{\infty} = \min_{q, q_c \leq N} R D_{q}(k_c, q)/D_{q}(k_c, k_c)$

and using Appendix B it can be easily shown that (see (24) at bottom of page)

where $Z_o = (1 - 4/\pi^2) / (1 + 4/\pi^2) \approx 0.423$.

From (24) we see that $q_c$ and not $w_c$ determines the coding gain. What is the relation of $q_c$ to the minimum Hamming

$$G_{\infty} = \begin{cases} R \frac{D_{q}(k_c, N)}{D_{q}(1,k_c)} = k_c \frac{N}{n_c} \frac{D_{q}(1,N/k_c)}{D_{q}(1,k_c)} & \frac{N}{n_c} \leq Z_o \\ R N \frac{D_{q}(1,1)}{D_{q}(1,k_c)} = k_c \frac{N}{n_c} \frac{D_{q}(1,N/k_c)}{D_{q}(1,1)} & \frac{N}{n_c} > Z_o \end{cases}$$

(24)
distance \( w_e \). For MSK, because of its memory, only a part of the \( w \)-error patterns have \( d_{EC}^2 > w \); the other \( w \)-error patterns have \( d_{EC}^2 = w \). Therefore, there might be situations where all the minimum Hamming weight error patterns which can be generated by the codewords have \( d_{EC}^2 > w \); in these cases \( q_e > w_e \).

When considering code design for noncoherent MSK schemes, a sensible design criterion is maximization of \( q_e \) rather than \( w_e \). Increasing \( w_e \) with a constraint on the rate \( R \) implies that the block-length \( n_e \) and hence the complexity \( k_e \) has to be increased. If \( w_e \) is increased with a constraint on the complexity \( k_e \), then again the block-length \( n_e \) has to be increased which shows that the rate \( R \) decreases. In general one cannot find a high rate short block-length binary code with \( w = 3 \). It is easy to verify that \( w_e = 4 \) implies that the block-length \( n_e \) and hence the complexity \( k_e \) has to be increased.

A linear binary code of block-length \( n_e \) which consists of \( 2^{k_e} \) codewords is a linear subspace of \( GF(q)^n \), the linear space of \( n_e \)-dimensional vectors over \( GF(q) \), the binary Galois field. This linear subspace is the null space of an \( (n_e-k_e) \times n_e \) binary matrix over \( GF(2) \) which is called the parity check matrix of the code. When this code is used with MSK and N-MLBE, the asymmetric coding gain of (24) is determined by \( q_e \), the minimum square circular Euclidean distance over the code or the null space of \( H \).

**Example 1:** \((7,4,3)\) Hamming code.

- a) The parity check matrix is given by
  \[
  H = \begin{bmatrix}
  1101100 \\
  1010101 \\
  0111001 
  \end{bmatrix}
  \]
  Let \( b \in GF^m(2) \) denote an MSK error pattern from Table I. It is easily verified that if \( d_{EC}^2(b) = 1 \) or 2 then \( H \cdot b \neq 0 \), where \( b \) is a vector with all its components zero. If \( b = [1110000]^T \) which gives \( d_{EC}^2(b) = 3 \), then \( H \cdot b = 0 \), and hence \( q_e = w_e = 3 \). b) Let us now consider an equivalent \((7,4,3)\) Hamming code which is obtained by interchanging the third and fourth bit in a codeword of the previous code. The parity check matrix of this code is 
  \[
  H = \begin{bmatrix}
  1110100 \\
  1011010 \\
  0111001 
  \end{bmatrix}
  \]
  It is easy to verify that \( H \cdot b \neq 0 \) if \( d_{EC}^2(b) = 1 \) or 2 or 3. But if \( b = [1100011]^T \) which gives \( d_{EC}^2(b) = 4 \), then \( H \cdot b = 0 \) and hence \( q_e = 4 > w_e \).

From Example 1 we have the following.

**Corollary:** Equivalent codes do not necessarily give the same coding gain when combined with MSK and N-MLBE.

This is another feature of the memory of MSK since with a memoryless modulation \( q_e = w_e \) and equivalent codes will always give the same coding gain.

Let us consider now linear block codes for MSK with N-MLBE which exploit the inherent memory of this modulation format. We fix the block-length \( n_e \) and find a high rate code that ensures a given \( q_e \). This is equivalent to finding the binary matrix \( H \) of \( n_e \) columns and small number of rows such that the set of elements of \( GF(q_e)(2) \) which correspond to error patterns with \( d_{EC}^2 < q_e \) are outside the null space.

**Example 2:** A code for \( q_e = 2 \).

From Table I we have that the error patterns with \( d_{EC}^2 = 1 \) are given by \( b_1 = [100\cdots000]^T \) and \( b_2 = [000\cdots001]^T \). A proper parity matrix is \( H = [100\cdots001] \), since \( Hb_1 = Hb_2 = 0 \). This parity check matrix defines the rate \( (n_e-1)/n_e \) encoding procedure: \( u_1 u_2 \cdots u_{n_e-1} u_1 \), which consists of adding one parity bit only on the first information bit (or alternatively, repeating the first information bit in the last codeword position). From (24) we have that the asymptotic coding gain of this code approaches 3 \( \text{dB} \) when it is used with MSK and N-MLBE, and the codeword length \( n_e \) is increased. The bandwidth expansion factor \( 1/R \) approaches 1 in this situation. Therefore this system is equivalent in performance to MSK with RBNE as \( n_e = N \rightarrow \infty \). Table V shows the asymptotic coding gain of this simple code for small block-lengths \( n_e \) and presents a comparison with uncoded MSK with RBNE for fixed demodulation complexity \( k_e \). It is seen that at \( k_e = 3 \), 4 this code has a larger coding gain than RBNE, however, there is also bandwidth expansion. For larger \( k_e \) RBNE demodulation gives a larger gain. In general, the differences in coding gains are small, and as \( k_e \rightarrow \infty \) the two systems yield identical performances.

A more powerful linear code requires that larger \( d_{EC}^2 \) error patterns to be absent from the null space of its parity check matrix. This usually requires more parity bits. Table I presents the error patterns up to \( d_{EC}^2 = 4 \). It is seen that the majority of error patterns with \( d_{EC}^2 \leq 3 \) are characterized by a double error cluster that can start at any location in the block. This error structure can be exploited by treating separately the even and odd bits. In other words we use two different interleaved codes. In this way the double error cluster will be reflected as a single error to each of the codes. If the minimum Hamming distance of at least one code is larger than one, then all these patterns of double errors will be outside the null space of the composed parity check matrix.

**Example 3:** A code for \( q_e = 4 \). Assume that \( n_e \) is even. Let us add one parity for the odd bits and another parity
for the even bits. The encoding rule for \( n_c = 8 \) is given by
\[
\begin{array}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
\end{array}
\]
The corresponding parity check matrix is \( H = \left[ 1010 \ldots 01 \right] \). From Table I we have that \( Hb \neq 0 \) if \( b \)
corresponds to an error pattern with \( d_{EC}^2(b) \leq 3 \). The rate of this code is \( (n_c - 2)/n_c \), which tends to one as \( n_c \to \infty \). Table V shows the asymptotic coding gain for small block length \( n_c \). When compared with RBNE of MSK, this code gives a gain of 1.1 dB at \( k_c = 6 \), and the gain increases to 2 dB when \( k_c = 8 \).

From Table I we see that the error patterns with \( q_c = 4 \) are characterized by two clusters of double errors which can appear anywhere in the block. By using the interleaved coding technique where one of the codes has a minimum Hamming distance of three we can construct a code for \( q_c = 5 \).

Example 4: A code for \( q_c = 5 \).

Let us consider a \((6,3,3)\) linear binary code with parity check matrix \( H_e = \left[ \begin{array}{cccccccc}
0111 & 0010 & 0011 & 0100 & 0101 & 1000 & 0000 & 1000 \\
1010 & 0010 & 0011 & 0100 & 0101 & 1000 & 0000 & 1000 \\
0010 & 0011 & 0100 & 0101 & 1000 & 0000 & 1000 & 0000 \\
\end{array} \right] \). If this code is used as one of the interleaved codes for a total block length of 11 bits, we have the following parity check matrix for the composed code \( H = \left[ \begin{array}{cccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\end{array} \right] \) where \( 0 \) are all zero columns and \( z \) denotes additional columns which depend on the second interleaved code. The first interleaved code checks the odd bits and the second interleaved code checks the even bits. Since the first code has a minimum Hamming distance of three we can construct a code for \( q_c = 5 \).

From Table V it is seen that the \((6,3,3)\) interleaved code gives a good tradeoff between complexity and gain, however the bandwidth expansion factor is relatively large. This is because of the low rate of the initial \((6,3,3)\) code. With a higher rate initial code we would end with a higher rate interleaved code which will result in less bandwidth expansion and larger gain as well; however the complexity \( k_c \) of decoding would also increase. If we start with a \((2^m - 1, 2^m - m - 1, 3)\) Hamming code for the odd bits and a single parity code as in Example 4 for the even bits, then the composed code will have a rate of \( R = \frac{2^m + 1 - m - 3}{2^m + 1 - 2} \).

<table>
<thead>
<tr>
<th>( m )</th>
<th>( n_c = N )</th>
<th>( k_c )</th>
<th>( 1/R = n_c/k_c )</th>
<th>( G_{\infty}[dB] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>14</td>
<td>10</td>
<td>1.4</td>
<td>5.1</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>25</td>
<td>1.2</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>62</td>
<td>56</td>
<td>1.1</td>
<td>6.4</td>
</tr>
<tr>
<td>( \infty )</td>
<td>( \infty )</td>
<td>( \infty )</td>
<td>1</td>
<td>6.98</td>
</tr>
</tbody>
</table>

The asymptotic gain of this code for \( m \geq 3 \) will be
\[
G_{\infty} = \frac{5}{2} \frac{2^{m+1} - m - 3}{2^m - 1} \frac{D_5^6(1,2(2^m-1)/5)}{D_5^6(1,2^{m+1} - m - 3)}
\]
and \( k_c = 2^{m+1} - m - 3 \). Some numerical results for these codes are presented in Table VI. It is seen that large coding gains with low bandwidth expansion can be achieved; however the required \( k_c \) is large and hence the decoding complexity is high. Similar methods can be used to construct codes for \( q_c \geq 5 \). The interleaved codes must be of larger Hamming distance and thus of larger block length.

VI. CONCLUSION

This paper deals with noncoherent demodulation of MSK. It has been shown that the optimal noncoherent MSK receiver which is based on the N-MLBE strategy is composed of matched filters and a combination of multi-differential detectors as well as envelope detectors. A key feature of MSK is its phase continuity which has to be exploited in order to have good demodulation performance. A method that exploits the inherent coding features induced by the MSK phase continuum, RBNE, has been introduced and analyzed. With RBNE, the performance of noncoherent MSK is close to antipodal signaling. The performance of noncoherent MSK can be further improved by enhancing its coding properties. This can be done by the use of an additional code. The use of high rate binary block coding with N-MLBE of MSK has been considered. It has been found that simple codes which exploit the error pattern structure of noncoherent MSK can provide significant coding gains.

APPENDIX A

In this Appendix we derive an asymptotic expression for the pairwise probability of error \((8)\).

From [14, p. 293] we have, \( P_e(i/m) = Q(a, b) - (1/2)e^{-(a^2+b^2)/2}L_0(ab) \) where \( Q() \) is the Marcum Q-function, \( L_0() \) is the modified Bessel function of order zero, and
\[
a = \sqrt{N(\gamma/2) [1 - \sqrt{1 - |\rho_{cm}|^2}]},
b = \sqrt{N(\gamma/2) [1 + \sqrt{1 - |\rho_{cm}|^2}]}.\]
The Marcum Q-function can be expressed as [14, p. 28]
\[ Q(a, b) = e^{-\left(\pi^2 + b^2\right)/2} \sum_{k=0}^{\infty} \frac{(a/b)^k I_k(ab)}, \quad b > a > 0 \]

where, \(I_k(ab)\) are the modified Bessel function of order \(k\). Using this expression yields

\[ P_r(i/m) = e^{-\left(\pi^2 + b^2\right)/2} \cdot \left[ \left(1/2\right) I_0(ab) + \sum_{k=1}^{\infty} \frac{(a/b)^k I_k(ab)}{2^k} \right]. \]

For large SNR \((\gamma \to \infty)\) and \(|\rho_{im}| > 0\) we have \(a, b \to \infty\), and the asymptotic form \(I_0(ab) \approx (2\pi/ab)^{-1/2} e^{-ab}\). Therefore,

\[ P_r(i/m) \approx (2\pi ab)^{-1/2} e^{-ab} \cdot \left[ \left(1/2\right) + \sum_{k=1}^{\infty} \frac{(a/b)^k}{2^k} \right]. \]

Now since \(a < b\) we have \(\sum_{k=1}^{\infty} (a/b)^k = (a/b)/(1 - (a/b))\), and, \(P_r(i/m) \approx 2^{-3/2}(\pi ab)^{-1/2}(1 + (a/b))/(1 - (a/b)) \cdot e^{-ab}\). Using the expressions for \(a\) and \(b\) yields \((a/b)^2 = \gamma N(1 - |\rho_{im}|), ab = (1/2)\gamma N|\rho_{im}|(1 + a/b)/(1 - a/b)\), \(1 - a/b = [1 + |\rho_{im}|]/[1 - |\rho_{im}|]\) and finally \(P_r(i/m) \approx (1/2)\gamma N(1 - |\rho_{im}|)^{-1/2} \cdot e^{-\pi\gamma N(1 - |\rho_{im}|)^{-1/2}}\).

Now we can express \(P_r(i/m)\) in terms of the \(Q\)-function associated with the Gaussian density, \(Q(x) = \int_{x}^{\infty} e^{-y^2/2} dy\). By using the asymptotic relation \(Q(x) \approx (2\pi)^{-1/2} \int_{x}^{\infty} e^{-y^2/2} dy\). We have \(P_r(i/m) \approx \sqrt{1 + |\rho_{im}|/[2|\rho_{im}|]} Q\left[\sqrt{\pi\gamma N(1 - |\rho_{im}|)^{-1/2}}\right].\)

**APPENDIX B**

In this Appendix we investigate the lower bound to the noncoherent distance of (13b). From (13b) we have:

\[ D_0^2(q, N) = N - \sqrt{(N - q)^2 + q^22/\pi^2}. \]

Taking the derivative of \(D_0^2(q, N)\) with respect to \(N\) gives \(\frac{\partial D_0^2(q, N)}{\partial N} = 1 - \left\{N - q \left\{ (N - q)^2 + q^22/\pi^2 \right\} \right\}^{-1/2}\). It is seen that \(\frac{\partial D_0^2(q, N)}{\partial N} > 0\), \(N \geq q\) which shows that \(D_0^2(q, N)\) increases monotonically with \(N(\geq N)\). Taking the derivative of \(D_0^2(q, N)\) with respect to \(q\) gives \(\frac{\partial D_0^2(q, N)}{\partial q} = \left\{N - q \left\{ (N - q)^2 + q^22/\pi^2 \right\} \right\}^{-1/2} \cdot \frac{1 - q}{\sqrt{(N - q)^2 + q^2(2/\pi^2)}}\). It is seen that at \(q = N\), \(D_0^2(q, N) \leq D_0^2(N, N) < N/2\). Therefore, for \(1 \leq q < q = N\), \(D_0^2(q, N)\) increases with \(q\) and for \(q < q \leq N\), \(D_0^2(q, N)\) decreases with \(q\) (see Fig. 2a). A simple algebraic manipulation gives

\[ D_0^2(q, N) = q\left[2 - (N/q) \left\{ (N - q)^2 + q^2(2/\pi^2) \right\} \right] \cdot \frac{1 + \sqrt{(N - q)^2 + q^2(2/\pi^2)}}{(N - q)^2 + q^2(2/\pi^2)} \]

which shows that \(\lim_{N \to \infty} D_0^2(q, N) = q\).\]

**ACKNOWLEDGMENT**

The authors wish to thank the reviewers of this paper for their careful review, which helped to improve the original manuscript.

**REFERENCES**


**Harry Leib** (M’84–SM’97) for a photograph and biography, see the January 1992 issue of this TRANSACTIONS, see p. 50.

**Subbarayan Pasupathy** (M’73–SM’81–F’91) for a photograph and biography, see the January 1992 issue of this TRANSACTIONS, see p. 50.