Spectrally Efficient Noncoherent Communication

Dilip Warrier, *Member, IEEE*, and Upamanyu Madhow, *Senior Member, IEEE*

***Abstract*—This paper considers noncoherent communication over a frequency-nonselective channel in which the time-varying channel gain is unknown *a priori*, but is approximately constant over a *coherence interval*. Unless the coherence interval is large, coherent communication, which requires explicit channel estimation and tracking prior to detection, incurs training overhead which may be excessive, especially for multiple-antenna communication. In contrast, noncoherent detection may be viewed as a generalized likelihood ratio test (GLRT) which jointly estimates the channel and the data, and hence does not require separate training. The main results in this paper are as follows**

1) **We develop a "signal space" criterion for signal and code design for noncoherent communication, in terms of the distances of signal points from the decision boundaries.**
2) **The noncoherent metric thus obtained is used to guide the design of signals for noncoherent communication that are based on amplitude/phase constellations. These are significantly more efficient than conventional differential phase-shift keying (PSK), especially at high signal-to-noise ratio (SNR). Also, known results on the high-SNR performance of multiple-symbol demodulation of differential PSK are easily inferred from the noncoherent metric.**
3) **The GLRT interpretation is used to obtain near-optimal low-complexity implementations of noncoherent block demodulation. In particular, this gives an implementation of multiple symbol demodulation of differential PSK, which is of linear complexity (in the block length) and whose degradation from the exact, exponential complexity, implementation can be made as small as desired.**

***Index Terms*—Differential phase-shift keying (PSK), differential quadrature amplitude modulation (QAM), generalized likelihood ratio test (GLRT), noncoherent communication, noncoherent distance.**

I. Introduction

THIS paper presents a framework for signal design for noncoherent communication over a frequency-nonselective channel. The channel complex gain is modeled as unknown, but constant over the duration of the transmitted signal. Such a model is well suited to time-varying channels which are difficult to track explicitly, but can be approximated well as piecewise constant over a *coherence interval* that spans, say, several symbol durations. Coherent communication, which requires a good estimate of the channel complex gain over such channels would typically require an overhead (e.g., in terms of unmodulated pilot symbols) that would be excessive unless the coherence interval is large. While noncoherent detection is a classical topic in communication theory, the need for efficient noncoherent coded modulation schemes has become particularly acute in recent years, due to the explosion of interest in high-speed wireless communication systems, and the potential use of multiple antennas to enhance performance [1]–[8]. The overhead required for channel estimation becomes even larger for coherent detection in multiple-antenna systems, since the channel gain between each pair of transmit and receive antenna elements must be measured. Although this paper focuses on frequency-nonselective channels, the results can be applied to frequency-selective channels by converting the latter into a number of frequency-nonselective subchannels by the use of multicarrier modulation, or by the use of equalization.[1] An existing multicarrier system using noncoherent detection is the European Digital Audio Broadcasting (DAB) standard which uses differential 4-PSK modulation on each subcarrier [10]. The results in this paper point the way for potential improvements in such a system, through the use of more power-efficient amplitude/phase constellations (Section IV), and through reduced complexity multiple-symbol demodulation (Section IV-B).

Attention is restricted in this work to signal design over a single coherence interval. Coding over multiple coherence intervals is the subject of future work.

For coherent reception over the additive white Gaussian noise (AWGN) channel, practical channel codes are now available for approaching the Shannon capacity for the entire range of bandwidth efficiencies [11]–[13], and a number of coding techniques have been developed for coherent reception over the Rayleigh faded channel [14], [15]. The state of the art for noncoherent systems lags far behind, consisting mainly of orthogonal modulation (which is not bandwidth efficient) and differential phase-shift keying (PSK) (which is less power efficient than amplitude/phase modulation). As shown later in this paper, noncoherent detection can be viewed as joint estimation of the channel and the data. Hence, if properly optimized, we would expect it to be more efficient, albeit at the expense of higher complexity, than separate channel estimation followed by coherent detection. However, in order to realize the promise of noncoherent communication, much work in signal and code design is needed. This paper takes some preliminary steps in this direction by developing systematic design criteria for noncoherent systems that

Manuscript received October 2, 1999; revised July 1, 2001. This work was supported in part by the National Science Foundation under a CAREER Award NSF NCR96-24008CAR, under Grant CCR99-79381, and by the Army Research Office under Grant DAAD19-00-1-0567. The material in this paper was presented in part at the IEEE Information Theory Workshop on Detection, Estimation and Imaging, Santa Fe, NM, February 1999 and at the Conference on Information Sciences and Systems (CISS), Baltimore, MD, March 1999.

D. Warrier is with Aware, Inc., Bedford, MA 01730 USA (e-mail: dwarrier@aware.com).

U. Madhow is with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA (e-mail: madhow@ece.ucsb.edu).

Communicated by M. L. Honig, Associate Editor for Communications.

Publisher Item Identifier S 0018-9448(02)01306-8.

[1]While traditional equalization methods attempt to track the channel gain and therefore may fail in rapidly time-varying environments, recently developed equalizers based on the differential minimum mean squared error criterion [9], which avoid tracking the channel gain, go naturally with noncoherent detection.

are analogous to the signal space concepts that have been so useful in the design of coherent systems.

Our main results are summarized as follows.

1) Noncoherent detection can be interpreted in terms of the generalized likelihood ratio test (GLRT); that is, as joint channel and data estimation. This leads to a geometric view of the noncoherent decision statistics as the magnitudes of the projection of the received signal on the complex subspaces spanned by each of the possible transmitted signals.

2) The asymptotic rate of decay of the pairwise error probability of noncoherent detection on an AWGN channel is computed based on the preceding view of the decision regions for the noncoherent detector. The noncoherent metric governing the rate of decay is given by the distance of the transmitted signal to the boundary of the decision region, and provides a systematic criterion for signal design, analogous to the notion of Euclidean distance between signal points for coherent systems.

3) It is well known that multiple symbol demodulation of differentially encoded PSK provides much better performance than conventional differential demodulation over two symbols (for example, demodulation over a block of six symbols provides a gain of about 2.1 dB over demodulation over two symbols for differential 8-PSK). Such block demodulation has complexity exponential in the block length. However, the GLRT interpretation of noncoherent detection enables realization of the performance gains of block demodulation at greatly reduced complexity. In particular, near-optimal block demodulation of differentially encoded PSK (DPSK) can be performed at linear complexity.

4) While a straightforward extension of differential encoding to amplitude/phase constellations yields poor performance, a modified block differential encoder that accounts for the noncoherent metric is shown, for relatively small coherence intervals, to achieve better power efficiency than PSK, just as is the case for coherent communication. At high SNRs, a 16-QAM alphabet gives a gain of about 2 dB over 16-PSK for a noncoherent AWGN channel with blocks of six symbols. Further, it is shown that, asymptotically for large coherence intervals, the performance of noncoherent block demodulation of DPSK and amplitude/phase constellations (with the modified block encoding) approaches that of coherent detection.

The performance gains from multiple symbol demodulation of DPSK were pointed out by Divsalar and Simon [16]. They derive a noncoherent metric for DPSK that is a special case of ours. However, their derivation is based on a Bayesian interpretation of noncoherent detection, and involves the asymptotics of the Marcum's Q function, in contrast to the geometric derivation given here. In view of the exponential complexity of multiple symbol demodulation, much work has gone into finding suboptimal procedures for multiple symbol demodulation that are easier to implement [17]–[20]. In particular, feedback of decoded symbols has been studied in great detail, although the error propagation characteristics are not well understood [21]–[23]. In contrast to the somewhat *ad hoc* approaches followed previously, the linear complexity near-optimal detector proposed here follows directly from the GLRT interpretation, and can be designed to have an arbitrarily small level of degradation from the optimal performance.

Coding (over multiple coherence intervals) for DPSK with multiple symbol demodulation has been studied in some detail [24]–[29]. This topic is not addressed in this paper, since our focus here is on signal design within a coherence interval. However, as mentioned in Section V, a systematic approach to coding for noncoherent communication is an important topic for future work.

In contrast to the literature on DPSK, much less is known about amplitude/phase modulation for noncoherent systems, with most work to date concentrating on demodulation based on symbol windows of size two (analogous to conventional DPSK) and on the effects of fading channels and diversity [30]–[35]. This prior work did not have the advantage of designing with a noncoherent metric in mind (which, as we shall see, motivates a modification of the standard differential encoder for amplitude/phase modulation), hence the anticipated gains in power efficiency over PSK were not realized. Coding over quadrature amplitude modulation (QAM) alphabets for fading channels with *coherent* detection has been considered in [36]. Frequency shift keying (FSK) is a popular modulation scheme for noncoherent systems, especially since orthogonal signals can be obtained for particular parameter values [37]–[42].

The properties of the GLRT have been studied from a statistical point of view. In particular, for finite alphabets and for a search set consisting of all memoryless channels, the GLRT can be shown to be equivalent to the maximum mutual information (MMI) estimator [43]. A comparison of the GLRT with the optimal maximum-likelihood (ML) estimator is made in [44]. The GLRT has also been effectively utilized for multiuser detection [45], [46].

Typical information-theoretic models used for noncoherent systems are the compound channel model [43], [47] and the block fading channel model [3], [48], [49], which is considered here as well.

Since the signals in noncoherent detection can be identified with the complex subspaces they span, design of signals for noncoherent detection may be viewed as a packing problem in projective space. Thus, results on the packing of planes in space, optimized according to certain distance criteria, can be applied for noncoherent space-time codes [50], [51]. Work has also been done on finding good packings of planes using gradient search methods [2], [8], [52].

This paper primarily deals with single-antenna systems. However, more recent research [53], [54] has revealed techniques to use one-dimensional (1-D) codes to get good noncoherent space-time codes using orthogonal transformations. These will be considered in detail in a future publication.

Section II contains the interpretation of noncoherent detection as a GLRT, and uses this to obtain reduced complexity near-optimal detectors. An application to multiple symbol demodulation of DPSK is described. The geometry of pairwise

decision making for noncoherent detection is considered in Section III, where the noncoherent metric governing the pairwise error probability is derived. The notion of differential encoding is generalized to amplitude/phase constellations in Section IV, and it is shown that a modification is required to obtain good values of the noncoherent metric. Simulation results for moderate coherence intervals are presented to show the performance of QAM and PSK alphabets in noncoherent systems using the suboptimal algorithm. Asymptotic results are presented in Sections IV-C and IV-D to show that the performance of noncoherent detection can approach that of coherent detection for large coherence intervals. Conclusions are provided and issues for future study identified in Section V.

II. The GLRT Approach and Its Consequences

For single antenna transmission, the received signal over one coherence interval is given by

$$\boldsymbol{y} = \sqrt{\eta}\boldsymbol{x}h + \boldsymbol{n} \tag{1}$$

where

- $\boldsymbol{y}$ is the received vector,[2]
- η is the SNR,
- $\boldsymbol{x}$ is the transmitted vector,
- $\boldsymbol{n}$ is a vector of complex AWGN with $\boldsymbol{E}[\boldsymbol{n}\boldsymbol{n}^H] = \boldsymbol{I}_N$, and
- h is an unknown complex channel coefficient.

All vectors are of size $N \times 1$ where N is the length of the coherence interval.

A. GLRT Detection

Over the AWGN channel, ML estimation corresponds to minimizing the Euclidean distance between the received and the hypothesized signals. The GLRT detector computes the joint ML estimate of the channel and the transmitted signal. Thus, the GLRT estimate for the transmitted signal satisfies

$$\begin{aligned}\hat{\boldsymbol{x}}_{\mathrm{GLRT}} &= \arg\max_{\boldsymbol{x}\in\mathcal{S}}, \sup_h p[\boldsymbol{y}/\boldsymbol{x}, h] \\ &= \arg\min_{\boldsymbol{x}\in\mathcal{S}} \inf_h ||\boldsymbol{y} - \sqrt{\eta}\boldsymbol{x}h||^2\end{aligned} \tag{2}$$

where $\mathcal{S}$ is the family of transmitted signal vectors, referred to as a codebook. The elements of the vectors of $\mathcal{S}$ belong to an alphabet denoted by $\mathcal{A}$. The optimum channel estimate, corresponding to the inner minimization in (2), is achieved by projecting the received vector $\boldsymbol{y}$ onto the 1-D complex subspace spanned by $\boldsymbol{x}$ as follows:

$$\arg\inf_h ||\boldsymbol{y} - \sqrt{\eta}\boldsymbol{x}h||^2 = \frac{1}{\sqrt{\eta}} \frac{\langle \boldsymbol{y}, \boldsymbol{x}\rangle}{||\boldsymbol{x}||^2}.$$

[2]Throughout this paper, boldface notation with lower case letters is used to denote vectors and boldface notation with upper case letters to denote matrices e.g., x is a scalar, $\boldsymbol{x}$ is a vector, and $\boldsymbol{X}$ is a matrix. $\boldsymbol{I}_k$ denotes an identity matrix of size $k \times k$. $\boldsymbol{x}^H$ denotes the conjugate of the transpose of the vector $\boldsymbol{x}$. For a set A, $|A|$ denotes its size.

Substituting back in (2), we obtain

$$\hat{\boldsymbol{x}}_{\mathrm{GLRT}} = \arg\max_{\boldsymbol{x}\in\mathcal{S}} \frac{|\langle \boldsymbol{y}, \boldsymbol{x}\rangle|}{||\boldsymbol{x}||} \tag{3}$$

as compared to the coherent ML estimate given by

$$\hat{\boldsymbol{x}}_{\mathrm{coh}} = \arg\max_{\boldsymbol{x}\in\mathcal{S}} p[\boldsymbol{y}/\boldsymbol{x}, h] = \arg\min_{\boldsymbol{x}\in\mathcal{S}} ||\boldsymbol{y} - \sqrt{\eta}\boldsymbol{x}h||^2. \tag{4}$$

Remark 1: The GLRT decoding rule is independent of the energy of the signal, which is given by $||\boldsymbol{x}||^2$. For large coherence intervals, the information in the energy of the signal is typically insignificant, e.g., see the Shannon-theoretic analysis of a block-fading channel model in [3].

Remark 2: For the metric in (3)

$$\frac{|\langle \boldsymbol{y}, \boldsymbol{x}\rangle|}{||\boldsymbol{x}||} = \frac{|\langle \boldsymbol{y}, \lambda\boldsymbol{x}\rangle|}{||\lambda\boldsymbol{x}||}.$$

Thus, if there exists a complex scalar λ such that both $\boldsymbol{x}$ and $\lambda\boldsymbol{x}$ belong to $\mathcal{S}$, then the GLRT rule cannot differentiate between them. Hence, the information contained in $\boldsymbol{x}$ and $\lambda\boldsymbol{x}$ must be the same.

A distinction has to be made between the coherent ML detector in (4) and the noncoherent ML detector, which assumes no knowledge of the exact value of the channel state, but does assume knowledge about its probability distribution. The noncoherent ML detector works as follows:

$$\hat{\boldsymbol{x}}_{\mathrm{nc,ML}} = \arg\max_{\boldsymbol{x}\in\mathcal{S}} \boldsymbol{E}_h\{p[\boldsymbol{y}/\boldsymbol{x}, h]\} \tag{5}$$

where the expectation is with respect to the assumed distribution of h. For a single transmit antenna, the noncoherent ML detector is identical to the GLRT detector, as long as the channel phase $\arg(h)$ is uniformly distributed over $[0, 2\pi]$ and the signals are of equal energy. Indeed, the noncoherent detector considered here is usually derived under the latter set of assumptions in standard texts on detection [55]. For multiple-antenna transmission, noncoherent ML detection and GLRT detection are identical, for example, for equal energy signaling with the channel gains from the transmit elements to the receive element(s) modeled as independent and identically distributed (i.i.d.) circular Gaussian random variables [3].

B. Overlapped Block Encoding

The original codebook $\mathcal{S}$ can be divided into equivalence classes, each equivalence class consisting of vectors which are scalar multiples of each other. We denote the set of such equivalence classes by $\mathcal{S}_{\mathrm{nc}}$. Any vector in $\mathcal{S}$, say $\boldsymbol{x}$, is uniquely specified by its equivalence class in $\mathcal{S}_{\mathrm{nc}}$, say $\tilde{\boldsymbol{x}}$, and its first element, say $\boldsymbol{x}[1]$. This mapping is denoted by $\boldsymbol{x} = F(\boldsymbol{x}[1], \tilde{\boldsymbol{x}})$. The GLRT decoding rule of (3) can distinguish between the equivalence classes of $\mathcal{S}_{\mathrm{nc}}$, but not between vectors within an equivalence class in $\mathcal{S}_{\mathrm{nc}}$. Hence, (3) can be written as

$$\hat{\boldsymbol{x}}_{\mathrm{GLRT}} = \arg\max_{\boldsymbol{x}\in\mathcal{S}_{\mathrm{nc}}} \frac{|\langle \boldsymbol{y}, \boldsymbol{x}\rangle|}{||\boldsymbol{x}||}. \tag{6}$$

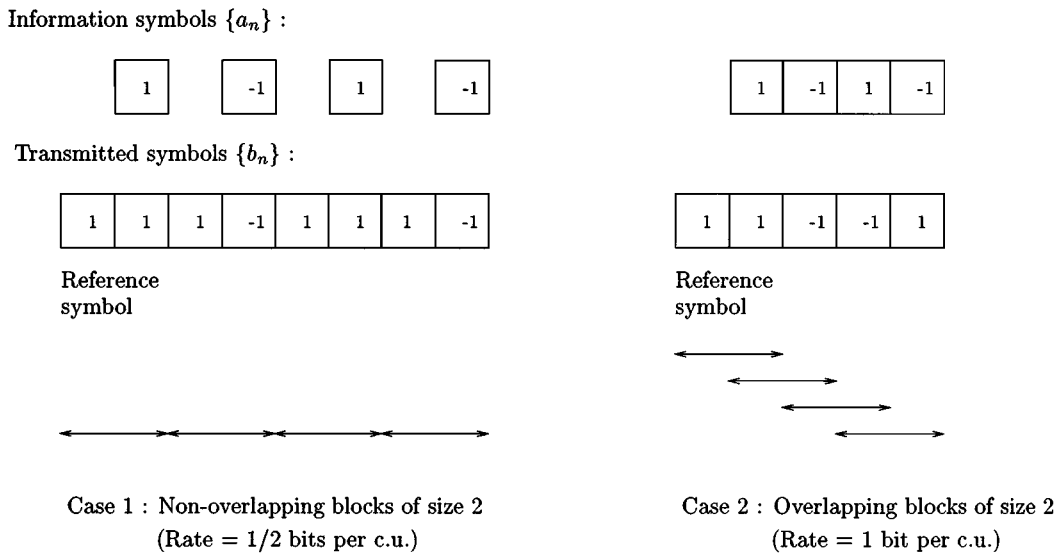

Fig. 1. Noncoherent detection with overlapping blocks.

Example 1: For $N = 2$, $\mathcal{A} = \{-1, 1\}$ and uncoded transmission

$$\mathcal{S} = \{(1, 1), (1, -1), (-1, 1), (-1, -1)\}$$
$$\mathcal{S}_{\rm nc} = \{(1, 1), (1, -1)\}.$$

The vector $\boldsymbol{x} = (-1, 1) = F(-1, (1, -1))$. In this case, $F(a, \boldsymbol{b}) = a\boldsymbol{b}$.

From Remark 2, only $\log_2 |\mathcal{S}_{\rm nc}|$ bits of information can be conveyed in a noncoherent setting for every block of N channel uses. If the channel remains constant over an interval of size N and varies arbitrarily between intervals, then $\frac{\log_2 |\mathcal{S}_{\rm nc}|}{N}$ is the best rate that can be achieved. However, it can be assumed in practice (e.g., in a slow fading channel) that the channel coefficient remains approximately constant over an interval of size N and varies slowly between intervals. Under this assumption, we can consider blocks overlapping by one symbol in the following manner. The source generates $\log_2 |\mathcal{S}_{\rm nc}|$ bits, which specify the equivalence class to be transmitted, say $\tilde{\boldsymbol{x}}$. Now, the last symbol from the previous transmitted block, say b, is used as a reference symbol along with $\tilde{\boldsymbol{x}}$ to get the new vector to be transmitted, i.e., $F(b, \tilde{\boldsymbol{x}})$. The first element of this vector is b, which need not be transmitted again and hence only $(N - 1)$ additional channel uses are required to transmit $\log_2 |\mathcal{S}_{\rm nc}|$ bits. The rate achieved is thus $\frac{\log_2 |\mathcal{S}_{\rm nc}|}{N-1}$, which is an improvement over the rate without overlap of $\frac{\log_2 |\mathcal{S}_{\rm nc}|}{N}$, and can, in certain cases, be equal to the maximum achievable rate using $\mathcal{S}$ of $\frac{\log_2 |\mathcal{S}|}{N}$. (For example, for the sets in Example 1, the maximum achievable rate of 1 bit per channel use is achieved by overlapping as in Fig. 1) This improvement in rate comes at no additional cost in terms of distance properties since the signal set used is the same as the original one. However, it relies on the slowly varying nature of the channel.

At the receiver, overlapping blocks of received vectors are used for detection as follows:

$$(y_0, y_1, \ldots, y_{N-1}) \approx h(b_0, b_1, \ldots, b_{N-1}) + \boldsymbol{n}$$
$$(y_{N-1}, y_N, \ldots, y_{2N-2}) \approx h(b_{N-1}, b_N, \ldots, b_{2N-2}) + \boldsymbol{n}$$

where

$$\{b_0, \ldots, b_{N-1}, b_N, \ldots, b_{2N-2}\}$$

and

$$\{y_0, \ldots, y_{N-1}, y_N, \ldots, y_{2N-2}\}$$

are, respectively, the sequence of transmitted and received symbols. The decoding procedure will not uniquely determine the bit sequences but will determine the equivalence class of $\mathcal{S}_{\rm nc}$ that they belong to.

C. Differential Modulation

As described in Section II-B, a sequence of bits of length $\log_2 |\mathcal{S}_{\rm nc}|$ is used to obtain an equivalence class in $\mathcal{S}_{\rm nc}$, which along with the first element, determines the modulated vector to be transmitted. Differential modulation provides a simple and systematic method for implementing this transformation. The incoming sequence of bits is parsed into $(N - 1)$ blocks of bits $\{\boldsymbol{b}[1], \boldsymbol{b}[2], \ldots, \boldsymbol{b}[N - 1]\}$, each of size $\frac{\log_2 |\mathcal{S}_{\rm nc}|}{N-1}$. In a differential modulation scheme, the ith modulated symbol $\boldsymbol{x}[i]$, $i = 1, 2, \ldots, N-1$ is determined by an operation $f_d(\boldsymbol{b}[i], \boldsymbol{x}[i-1])$ on the previous symbol $\boldsymbol{x}[i-1]$ and the information bits $\boldsymbol{b}[i]$. The reverse operation $\boldsymbol{b}[i] = g_d(\boldsymbol{x}[i], \boldsymbol{x}[i-1])$ is done at the receiver to get back the information bits. From Remark 2, a noncoherent differential modulation scheme must satisfy the following condition.

Condition 1: If $\boldsymbol{x}$ and $\lambda\boldsymbol{x}$ belong to $\mathcal{S}$ for some complex scalar λ, then g_d must satisfy:

$$g_d(\boldsymbol{x}[i], \boldsymbol{x}[i-1]) = g_d(\lambda\boldsymbol{x}[i], \lambda\boldsymbol{x}[i-1]), \quad i = 1, 2, \ldots, N.$$

Example 2: For uncoded MPSK modulation,

$$\mathcal{A} = \left\{1, \exp\left[j\frac{2\pi}{M}\right], \exp\left[j\frac{4\pi}{M}\right], \ldots, \exp\left[j\frac{2\pi(M-1)}{M}\right]\right\},$$
$$|\mathcal{S}| = M^N \text{ and } |\mathcal{S}_{\rm nc}| = M^{N-1}.$$

The members of an equivalence class in $\mathcal{S}_{\rm nc}$ are obtained by taking any one of the vectors in it and multiplying by all values of $\lambda \in \mathcal{A}$. Let $m(\boldsymbol{b})$ denote the mapping from a sequence of

$\log_2 M$ bits, say $\boldsymbol{b}$, to the corresponding symbol in $\mathcal{A}$. Then, conventional differential PSK employs

$$f_{d,\,\mathrm{PSK}}(\boldsymbol{b}[i],\, \boldsymbol{x}[i-1]) = m(\boldsymbol{b}[i])\boldsymbol{x}[i-1]$$
$$g_{d,\,\mathrm{PSK}}(\boldsymbol{x}[i],\, \boldsymbol{x}[i-1]) = m^{-1}(\boldsymbol{x}[i]\boldsymbol{x}^*[i-1]).$$

We can check that $g_{d,\,\mathrm{PSK}}$ satisfies Condition 1 for all values of $\lambda \in \mathcal{A}$.

Differential modulation schemes for M-QAM are discussed in Section IV. We now develop a linear complexity decoding scheme for M-PSK, based on the differential modulation scheme above.

D. Linear Complexity Multiple Symbol Detection of DPSK

Multiple symbol demodulation $(N > 2)$ of standard differential M-PSK gives substantial gains over conventional differential detection with $N = 2$, for $M > 2$ [5]. However, for M-PSK, the cardinality of the search set for the optimization in (6) grows exponentially with N: $|\mathcal{S}_{\mathrm{nc}}| = M^{N-1}$. In the following, we present a near-optimal linear complexity block demodulator for uncoded DPSK. The extension of this concept to more general noncoherent codes is currently under investigation.

In (2) for the GLRT, the exponential complexity occurs because of the maximization over $\mathcal{S}$. However, if the orders of the optimization are interchanged, we have

$$\sup_{h} \max_{\boldsymbol{x} \in \mathcal{S}} p[\boldsymbol{y}/\boldsymbol{x},\, h].$$

The inner maximization is now a coherent detection procedure. For uncoded PSK, coherent detection can be done symbol-by-symbol, and is therefore of linear complexity in N. The complexity of the outer optimization can be reduced by restricting the choice of h to a finite family, incurring a controlled loss in optimality that depends on the granularity of the quantization.

The decision regions for coherent detection of PSK are independent of the amplitude scaling induced by the channel. Thus, it suffices to consider only the phase distortion θ caused by the channel. In this case

$$\boldsymbol{y} = \sqrt{\eta}\boldsymbol{x}\exp(j\theta) + \boldsymbol{n}.$$

Let $\hat{\theta}$ be an estimate of θ from the family Φ, which is obtained by quantizing an interval of candidate phase estimates. For demodulation of M-DPSK, it suffices to quantize the interval $[0,\, \frac{2\pi}{M}]$, because the remainder does not change the estimated equivalence class. In this case, Φ could be chosen as

$$\Phi = \left\{0,\, \frac{2\pi}{ML},\, \ldots,\, \frac{2\pi(L-1)}{ML}\right\}.$$

The size L of Φ can be tuned to approach the optimal performance as closely as required.

The algorithm proceeds in two steps.

1) **Coherent step**:
   For each phase estimate $\hat{\theta} \in \Phi$, perform symbol-by-symbol coherent demodulation based on the phase-corrected received vector $\boldsymbol{y}\exp(-j\hat{\theta})$. This yields an estimate $\hat{\boldsymbol{x}}(\hat{\theta})$ for the transmitted vector $\boldsymbol{x}$.

2) **Noncoherent step**:
   Among the candidates $\{\hat{\boldsymbol{x}}(\hat{\theta}),\, \hat{\theta} \in \Phi\}$ choose the one that yields the largest noncoherent decision metric, that is, the estimate $\hat{\boldsymbol{x}}(\Theta)$ such that

$$\Theta = \arg\max_{\hat{\theta} \in \Phi} \left| \left\langle \boldsymbol{y},\, \frac{\hat{\boldsymbol{x}}\left(\hat{\theta}\right)}{\|\hat{\boldsymbol{x}}\left(\hat{\theta}\right)\|} \right\rangle \right|.$$

The complexity of the preceding detector is linear in the resolution L, which is independent of N. The complexity is linear in N, which is, of course, the minimum possible complexity expected for processing N symbols. Simulation results show that the loss in performance from using this procedure as opposed to the original GLRT is negligible for reasonable values of the phase resolution L (e.g., $L = 16$ for 8-PSK).

In Section IV-B, the notion of differential encoding is generalized to amplitude/phase modulation. In this case, the suboptimal algorithm can no longer ignore the amplitude scaling of the channel. However, as shown in this section, the complexity can be reduced somewhat.

Remark 3: In its pure form, the GLRT requires computations of exponential complexity, since an *implicit* estimate of the channel is made for each possible transmitted signal. We have used *explicit* channel estimates to reduce this complexity, thus reducing noncoherent detection to L coherent detectors in parallel. The key distinction between coherent and noncoherent detection is, however, that we have no reason to have more confidence in any one of these parallel channel estimates. An interesting direction for future research is to attempt to use side information about the channel to reduce the number of parallel coherent detectors, and to adapt the set of channel estimates over time. However, the lack of an absolute phase reference (in the absence of training or pilot symbols) implies that it would still be necessary to employ codes that optimize a noncoherent metric such as the one in Section III.

III. Signal Space Concepts

Every decoding rule partitions the space of all received vectors into decoding regions corresponding to each candidate transmitted vector. The performance of the rule can be characterized by the properties of these decoding regions. We are interested in determining a good measure of the performance of signaling schemes under the GLRT decoding rule. Since the channel is noncoherent Gaussian, conditioned on the amplitude over the coherence interval, we look at the GLRT decoding regions for the AWGN channel to derive such a measure. The received vector in this case is given by $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}$ where $\boldsymbol{n}$ is a vector of AWGN with covariance $E[\boldsymbol{n}\boldsymbol{n}^*] = 2\sigma^2\boldsymbol{I}_N$.

Theorem 1: For an AWGN channel, in the high-SNR regime, the pairwise error probability for the GLRT decoding rule decays exponentially with SNR, that is,

$$\lim_{\sigma^2 \to 0} 2\sigma^2 \log[P(\hat{\boldsymbol{x}} = \boldsymbol{x_2}/\boldsymbol{x} = \boldsymbol{x_1})] = -d^2(\boldsymbol{x_1},\, \boldsymbol{x_2})$$

where $\hat{\boldsymbol{x}}$ is the estimated signal vector and $d(\boldsymbol{x_1}, \boldsymbol{x_2})$ denotes the shortest distance from the transmitted signal $\boldsymbol{x_1}$ to its decoding region boundary and is given by

$$d^2(\boldsymbol{x_1}, \boldsymbol{x_2}) = \frac{||\boldsymbol{x_1}||^2}{2}[1 - |\rho|] \quad \text{and} \quad \rho = \frac{\langle \boldsymbol{x_1}, \boldsymbol{x_2} \rangle}{||\boldsymbol{x_1}||||\boldsymbol{x_2}||}. \quad (7)$$

The proof appears in Appendix A.

Further, for any SNR, the following theorem holds for the Chernoff bound on the error probability.

Theorem 2: For an AWGN channel, under the GLRT decoding rule, the pairwise error probability is bounded by

$$P(\hat{\boldsymbol{x}} = \boldsymbol{x_2}/\boldsymbol{x} = \boldsymbol{x_1}) \leq \frac{1 + |\rho|}{2|\rho|} \exp\left[-\frac{d^2(\boldsymbol{x_1}, \boldsymbol{x_2})}{2\sigma^2}\right]. \quad (8)$$

The proof appears in Appendix B.

These results motivate the consideration of this distance as a measure of the performance of a signaling scheme under the GLRT decoding rule.

An intuitive picture is presented in Fig. 2, which can be explained as follows. Denote the 1-D subspaces spanned by $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ to be $\mathcal{S}_{\boldsymbol{x_1}}$ and $\mathcal{S}_{\boldsymbol{x_2}}$, respectively. Further, let $\theta_{\min}$ be the minimal angle between any two vectors picked from these two subspaces. Then

$$\cos\theta_{\min} = \frac{|\langle \boldsymbol{x_1}, \boldsymbol{x_2} \rangle|}{||\boldsymbol{x_1}||||\boldsymbol{x_2}||}$$

and the square of the distance from $\boldsymbol{x_1}$ to the decoding region boundary is then

$$\begin{aligned} d^2(\boldsymbol{x_1}, \boldsymbol{x_2}) &= ||\boldsymbol{x_1}||^2 \sin^2(\theta_{\min}/2) \\ &= \frac{||\boldsymbol{x_1}||^2}{2}\left[1 - \frac{|\langle \boldsymbol{x_1}, \boldsymbol{x_2} \rangle|}{||\boldsymbol{x_1}||||\boldsymbol{x_2}||}\right]. \end{aligned}$$

When $\boldsymbol{x_2}$ is sent, the distance is $d(\boldsymbol{x_2}, \boldsymbol{x_1})$. In general, the goal of signal design is to control the minimum of the two distances (normalized by the bit energy E_b) given by

$$d^2_{\text{nc}}(\boldsymbol{x_1}, \boldsymbol{x_2}) := \frac{\min[||\boldsymbol{x_1}||^2, ||\boldsymbol{x_2}||^2]}{2E_b}\left(1 - \frac{|\langle \boldsymbol{x_1}, \boldsymbol{x_2} \rangle|}{||\boldsymbol{x_1}||||\boldsymbol{x_2}||}\right). \quad (9)$$

For a given codebook $\mathcal{S}$ and the corresponding set of equivalence classes $\mathcal{S}_{\text{nc}}$, the performance measure in a noncoherent setting is given by the worst case distance

$$\begin{aligned} &\min_{\boldsymbol{x_1}, \boldsymbol{x_2} \in \mathcal{S}_{\text{nc}}} d^2_{\text{nc}}(\boldsymbol{x_1}, \boldsymbol{x_2}) \\ &\quad = \min_{\boldsymbol{x_1}, \boldsymbol{x_2} \in \mathcal{S}_{\text{nc}}} \left\{\frac{\min[||\boldsymbol{x_1}||^2, ||\boldsymbol{x_2}||^2]}{2E_b}\left(1 - \frac{|\langle \boldsymbol{x_1}, \boldsymbol{x_2} \rangle|}{||\boldsymbol{x_1}||||\boldsymbol{x_2}||}\right)\right\}. \quad (10) \end{aligned}$$

The distance corresponding to (9), for coherent detection, is given by

$$d^2_{\text{coh}}(\boldsymbol{x_1}, \boldsymbol{x_2}) := \frac{1}{4E_b}(||\boldsymbol{x_1}||^2 + ||\boldsymbol{x_2}||^2 - 2\text{Re}[\langle \boldsymbol{x_1}, \boldsymbol{x_2} \rangle]). \quad (11)$$

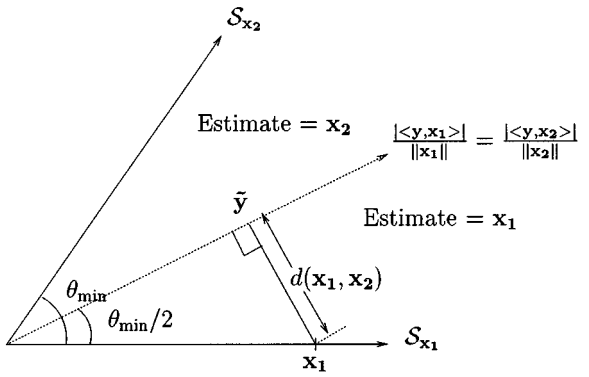

Fig. 2. Signal space geometry.

The following result holds for the coherent distance, in relation to the noncoherent distance.

Proposition 1: For any two vectors $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$

$$d_{\text{nc}}(\boldsymbol{x_1}, \boldsymbol{x_2}) \leq d_{\text{coh}}(\boldsymbol{x_1}, \boldsymbol{x}_2).$$

Proof: Equation (11) can be expressed as

$$\begin{aligned} d^2_{\text{coh}}(\boldsymbol{x_1}, \boldsymbol{x_2}) &= \frac{1}{2E_b} \\ &\quad \times \frac{||\boldsymbol{x_1}||^2 + ||\boldsymbol{x_2}||^2}{2}\left(1 - \frac{\text{Re},[\langle \boldsymbol{x_1}, \boldsymbol{x_2} \rangle]}{(||\boldsymbol{x_1}||^2 + ||\boldsymbol{x_2}||^2)/2}\right). \end{aligned}$$

The desired result is obtained by comparing individual terms above with those in (9) and using the following relations:

$$\frac{||\boldsymbol{x_1}||^2 + ||\boldsymbol{x_2}||^2}{2} \geq \min[||\boldsymbol{x_1}||^2, ||\boldsymbol{x_2}||^2]$$

$$\frac{||\boldsymbol{x_1}||^2 + ||\boldsymbol{x_2}||^2}{2} \geq ||\boldsymbol{x_1}||||\boldsymbol{x_2}||$$

and

$$\text{Re}[\langle \boldsymbol{x_1}, \boldsymbol{x_2} \rangle] \leq |\langle \boldsymbol{x_1}, \boldsymbol{x_2} \rangle|.$$

The two distances are equal only when the inner product $\langle \boldsymbol{x_1}, \boldsymbol{x_2} \rangle$ is real and the signals are of equal energy. Hence, noncoherent signal design can yield quite different results from those obtained by coherent signal design. However, from the proposition above, it is clear that signals designed using the noncoherent metric will provide the desired performance even when the channel is known and coherent detection is employed. The converse does not hold, i.e., signal designs based on the coherent metric are not necessarily amenable to noncoherent detection.

A. Applications to DPSK

The distance measures obtained in the previous section can be used to evaluate the performance of multiple symbol demodulation of DPSK. The results in this section have been previously derived in [16] and [56] using a more complex approach involving the asymptotics of the Marcum's Q-function. Similarly, results for the noncoherent distance for continuous phase modulation (CPM) signals have been derived in [34]. Our purpose here is to demonstrate that the results for equal energy signals can be evaluated as special cases of (10).

In Appendix C, it is shown that the minimum in (10) for an M-PSK alphabet is achieved by the pair

$$\tilde{\boldsymbol{x}}_{\boldsymbol{1}} = \{\underbrace{1, 1, \ldots, 1}_{(N-1) \text{ times}}, 1\} \tag{12}$$

and

$$\tilde{\boldsymbol{x}}_{\boldsymbol{2}} = \left\{\underbrace{1, 1, \ldots, 1}_{(N-1) \text{ times}}, \exp\left[j\frac{2\pi}{M}\right]\right\}. \tag{13}$$

The following conclusions result from this observation.

1) For $M = 2$ (differential BPSK)

$$d^2_{\text{nc}}(\tilde{\boldsymbol{x}}_{\boldsymbol{1}}, \tilde{\boldsymbol{x}}_{\boldsymbol{2}}) = 1$$

independent of N for all $N \geq 2$.

Remark 4: In terms of error exponents, there are no gains in increasing the window size beyond $N = 2$ for differential BPSK.

2) For large values of N

$$\lim_{N\to\infty} d^2_{\text{nc}}(\tilde{\boldsymbol{x}}_{\boldsymbol{1}}, \tilde{\boldsymbol{x}}_{\boldsymbol{2}}) = \log_2[M]\sin^2\left[\frac{\pi}{M}\right]$$

which is equal to the distance obtained by coherent detection using an M-PSK alphabet. Meanwhile, for $N = 2$ (conventional differential detection)

$$d^2_{\text{nc}}(\tilde{\boldsymbol{x}}_{\boldsymbol{1}}, \tilde{\boldsymbol{x}}_{\boldsymbol{2}}) = 2\log_2[M]\sin^2\left[\frac{\pi}{2M}\right].$$

Since $\sin[\frac{\pi}{M}] \approx 2\sin[\frac{\pi}{2M}]$ for large M

$$\lim_{N\to\infty} d^2_{\text{nc}}(\tilde{\boldsymbol{x}}_{\boldsymbol{1}}, \tilde{\boldsymbol{x}}_{\boldsymbol{2}}) \approx 2d^2_{\text{nc}}(\tilde{\boldsymbol{x}}_{\boldsymbol{1}}, \tilde{\boldsymbol{x}}_{\boldsymbol{2}})\Big|_{N=2}.$$

Remark 5: There is a gain of about 3 dB in using large values of N over conventional differential detection of M-PSK for large values of M.

In the next section, the noncoherent metric is applied to obtain new results on the design and evaluation of signals based on amplitude/phase constellations.

IV. Differential Amplitude/Phase Modulation

QAM alphabets[3] are known to be more energy efficient than PSK for coherent systems, especially at high SNR. In this section, we demonstrate that this advantage is applicable to noncoherent systems as well, by using the noncoherent metric in Theorem 1 as a guide to signal design based on QAM constellations. QAM alphabets for systems with unknown phase have been considered before [22], [23], [30], [31]. However, our use of the noncoherent metric enables us to identify constraints necessary for QAM constellations and for the partitioning thereof. We demonstrate that the use of QAM constellations in this manner helps realize the potential performance gains of QAM over PSK.[4]

First, a differential modulation scheme, analogous to that for M-PSK in Section II-C, is proposed for M-QAM alphabets as follows. The M-ary constellation $\mathcal{A}$ is divided into N_a subconstellations, such that each subconstellation is a (possibly offset) N_p-PSK alphabet and $M = N_aN_p$. Out of the $\log_2 M$ bits used to label a symbol, the first $\log_2 N_a$ bits are used to label the subconstellation and the remaining $\log_2 N_p$ bits are used to label each point within it. This labeling is shown for 8-QAM ($N_a = 2$, $N_p = 4$) and 16-QAM ($N_a = 4$, $N_p = 4$) in Figs. 3 and 4, respectively.

Information is encoded in the transitions between symbols as follows. Let the previous symbol have a bit labeling $\boldsymbol{s}$ of length $\log_2 M$ bits parsed into two parts $\boldsymbol{s_1}$ and $\boldsymbol{s_2}$ of length $\log_2 N_a$ and $\log_2 N_p$, respectively. Also, let the information bit sequence be $\boldsymbol{b}$ of length $\log_2 M$ bits parsed into two parts $\boldsymbol{b_1}$ and $\boldsymbol{b_2}$. Then, the next symbol is given by

$$\begin{aligned}\boldsymbol{r} &= f_{d,\,\text{QAM}}(\boldsymbol{b}, \boldsymbol{s}) \\ &:= (\boldsymbol{b_1} \oplus \boldsymbol{s_1}, (\boldsymbol{b_2} + \boldsymbol{s_2}) \bmod N_p)\end{aligned}$$

where $\oplus$ denotes the bitwise XOR operation. Some examples of differential modulation using this scheme are presented in Fig. 5. This provides a natural extension of differential PSK to amplitude/phase modulation. It can be checked that $f_{d,\,\text{QAM}}$ satisfies Condition 1.

We wish to generate a codebook consisting of vectors with elements drawn from this QAM alphabet. For PSK alphabets, it was assumed that all possible vectors are included in the codebook (hence, $|\mathcal{S}| = M^N$), but the following example illustrates that a codebook with all the possible vectors from a QAM alphabet has poor noncoherent performance. For the rest of this section, we focus attention on an 8-QAM alphabet as in Fig. 6.

Example 3: Consider $N = 4$ and an 8-QAM alphabet with $r_0 = 1$, $r_1 = 2$. Let

$$\tilde{\boldsymbol{x}}_{\boldsymbol{1}} = \{1, 1, 1, 1\} \quad \text{and} \quad \tilde{\boldsymbol{x}}_{\boldsymbol{2}} = \{1, 1, 1, 2\exp(j\pi/4)\}.$$

Then $d^2_{\text{nc}}(\tilde{\boldsymbol{x}}_{\boldsymbol{1}}, \tilde{\boldsymbol{x}}_{\boldsymbol{2}}) = 0.2443$, which compares poorly with

$$\min_{\boldsymbol{x_1}, \boldsymbol{x_2} \in \mathcal{S}_{\text{nc}}} d^2_{\text{nc}}(\boldsymbol{x_1}, \boldsymbol{x_2}) = 0.3391$$

where $\mathcal{S}_{\text{nc}}$ consists of vectors of 8-PSK symbols.

An intuitive explanation for the poor performance of unconstrained QAM is as follows. The signal $\tilde{\boldsymbol{x}}_{\boldsymbol{1}}$ is of low energy compared to the other signals, so that it has a very high probability of being decoded wrongly, irrespective of its correlations with other signals. This suggests that the signals in the codebook have

[3]While the term QAM is generally reserved for alphabets with signal points at the corners of squares, the term is used in this paper to mean any modulation scheme with both amplitude and phase information.

[4]An approach similar to our Gray labeling based approach for partitions of QAM constellations is used by Weber [33]. However, in Weber's work, the nearest neighbors are selected based on the Euclidean distance. This leads to the division of the signal space into pie-shaped sectors, where each sector is labeled using a subset of the information bits and signal points within each sector are labeled using the remaining bits. Our identification of the worst case vectors based on the noncoherent metric gives us a better division of the signal space. Note, for instance, in Fig. 8, that the first two bits are common for the outermost circle, the two middle semicircles, and the innermost circle. Such a division is not possible using pie-shaped sectors.

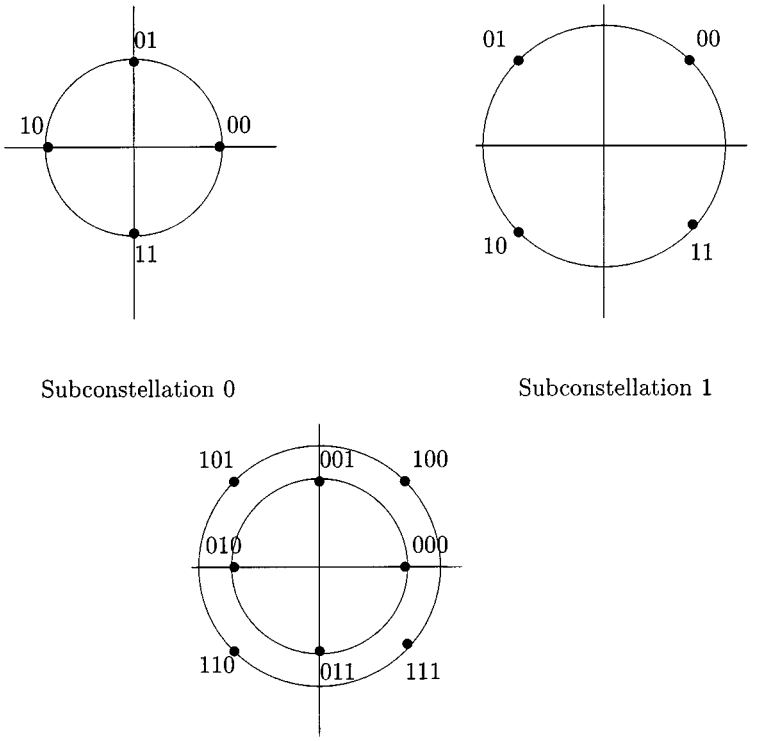

Fig. 3. Bit labeling for an 8-QAM alphabet.

to satisfy certain minimum energy constraints. The need for energy constraints is quantified in Section IV-C using asymptotic results.

We consider the following method for implementing the energy constraint. A vector of t symbols is constrained to have at least $\lceil \frac{t}{2} \rceil$ symbols from the higher amplitude level. This constraint is easily implemented since, if a given vector does not satisfy the constraint, one that satisfies the constraint can be obtained by simply inverting the amplitude bit.

Example 4: Consider two blocks of $(N-1)$ information symbols with $N = 4$ and an 8-QAM alphabet with $r_0 = 1$ and $r_1 = 2$, as follows:

$$\{001, 110, 001\} \quad \text{and} \quad \{100, 001, 010\}.$$

Let the reference symbol for the first block be $\{000\}$. By the differential modulation procedure above

$$\begin{array}{ccccccc} 000 & \xrightarrow{001} & 001 & \xrightarrow{110} & 111 & \xrightarrow{001} & 100 \\ & & \downarrow & & \downarrow & & \downarrow \\ & & 1\exp(j\pi/2) & & 2\exp(j7\pi/4) & & 2\exp(j\pi/4.) \end{array}$$

The resulting signal vector

$$\{1\exp(j\pi/2),\, 2\exp(j7\pi/4),\, 2\exp(j\pi/4)\}$$

satisfies the minimum energy criterion and can be transmitted as is. (Note that overlapping blocks can be used as described in Section II-B. Hence, only $(N-1)$ symbols are transmitted and the last symbol of the previous block is prefixed to each transmitted block for detection purposes.)

For the second block, the reference symbol is $\{100\}$. By the encoding procedure

$$\begin{array}{ccccccc} 100 & \xrightarrow{100} & 000 & \xrightarrow{001} & 001 & \xrightarrow{010} & 011 \\ & & \downarrow & & \downarrow & & \downarrow \\ & & 1 & & 1\exp(j\pi/2) & & 1\exp(j3\pi/2). \end{array}$$

This block does not satisfy the minimum energy criterion and, hence, all the amplitude bits are flipped to give

$$\{100, 101, 111\} \iff \{2\exp(j\pi/4),\, 2\exp(j3\pi/4),\, 2\exp(j7\pi/4)\}$$

as the signal vector to be transmitted.

If the reference signal for the first block and all the transmitted signals are decoded correctly, the decoded information bits will be

$$\{001, 110, 001, 000, 001, 010\}$$

Comparing with the original information bits, it is observed that one bit is incorrectly decoded because the energy constraint had to be satisfied at the encoder. Hence, the first amplitude bit of every block has to be ignored, resulting in a loss of rate of $\frac{1}{N}$ bit per channel use, as a result of the energy constraint.

The rate obtained for an M-QAM alphabet with overlapping blocks and a coherence interval length of N, in terms of number of bits per channel use, is given by

$$R = \begin{cases} \log_2(M) - \frac{1}{N}, & N \text{ odd} \\ \frac{1}{N}\left[\log_2\left(M^N + 2^{-N}\binom{N}{N/2}\right) - 1\right], & N \text{ even} \end{cases}$$

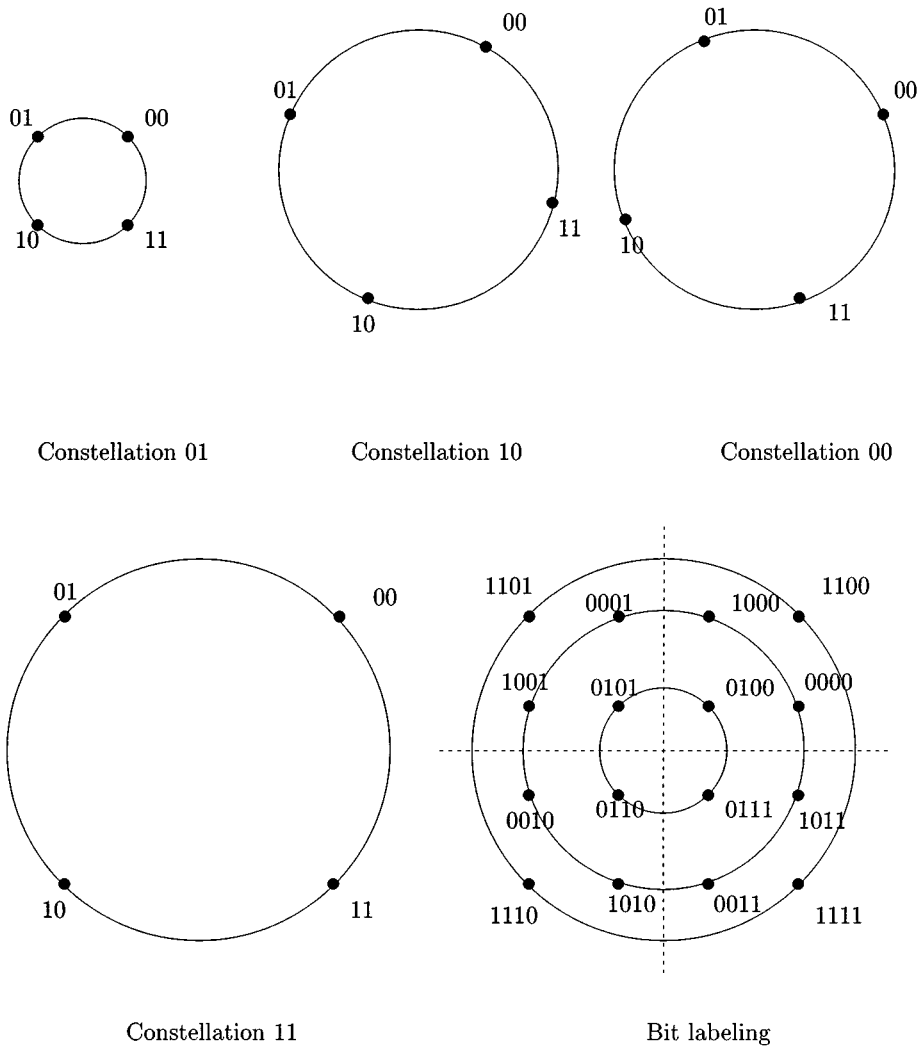

Fig. 4. Bit labeling for a 16-QAM alphabet.

compared to a rate of $\log_2(M)$ for an uncoded M-PSK constellation. The loss in rate (approximately $\frac{1}{N}$ as is seen in the example above and in the expression for the rate) due to the energy constraints is, therefore, negligible for large values of N.

For $N = 6$ and an 8-QAM alphabet, the worst case vectors, in terms of the noncoherent metric (found by exhaustive search) are

$$\{1,\, 1,\, 1,\, 2\exp(j\pi/4),\, 2\exp(j\pi/4),\, 2\exp(j\pi/4)\} \tag{14}$$

and

$$\{1, 1, 2\exp(j\pi/4), 2\exp(j\pi/4), 2\exp(j\pi/4), 2\exp(j\pi/4)\}. \tag{15}$$

The distance corresponding to this pair affords a further improvement of about 0.65 dB over 8-PSK with $N = 6$. (For a 16-QAM alphabet, finding the worst case vectors involves a computationally intensive search. We rely on actual channel simulations, instead, to evaluate performance.)

Thus, QAM alphabets can be used with certain power constraints to improve upon the performance of PSK alphabets. The theoretical results obtained using distance arguments in this section are substantiated using simulation results in the following section.

A. Gray Bit Labeling

So far, we have considered strategies to minimize the probability that one codevector is mistaken for another. However, given that a demodulation error occurs, the number of bit errors can be minimized using Gray bit labeling. Examples of Gray bit labeling are provided for PSK and QAM constellations in Figs. 7 and 8.

Note that the Gray bit labeling gives a different result from the bit labeling for differential modulation (called "differential labeling" for short). (Compare, for instance, the bit labeling for 8-QAM in Figs. 3 and 7.) The difference is owing to the different philosophies associated with the two schemes.

1) Gray labeling is intended to reduce the difference in the number of bits between the representations of two equivalence classes in $\mathcal{S}_{\mathrm{nc}}$ that are close by the noncoherent metric. For example, for 8-QAM and $N = 4$, a worst case vector pair is given by:

$$\left\{1,\, 1,\, 2\exp\left(-j\frac{\pi}{4}\right),\, 2\exp\left(-j\frac{\pi}{4}\right)\right\} \iff \{000,\, 000,\, 100,\, 100\}$$

and

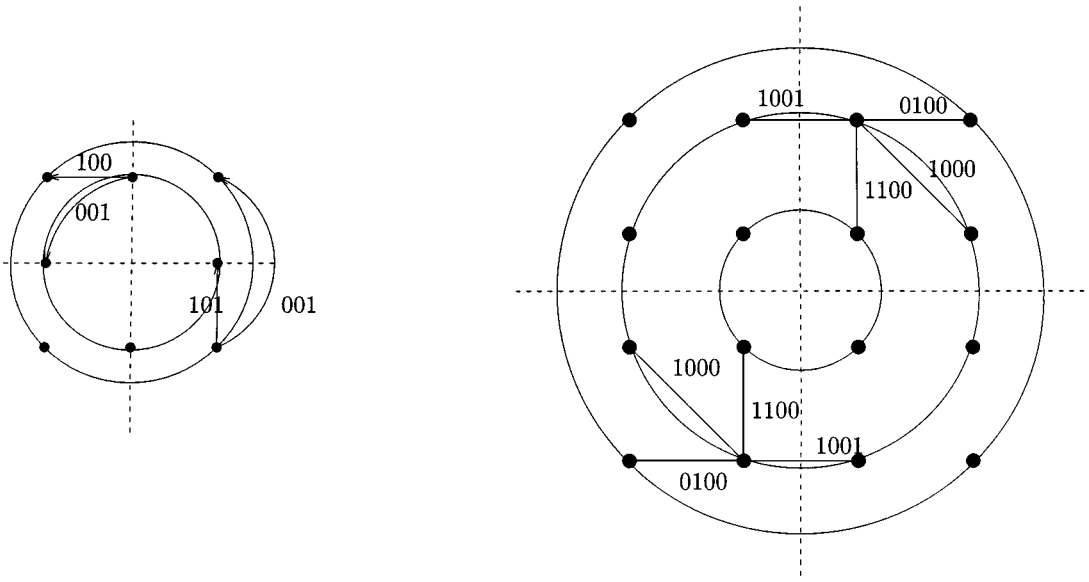

Fig. 5. Differential modulation for 8-QAM and 16-QAM.

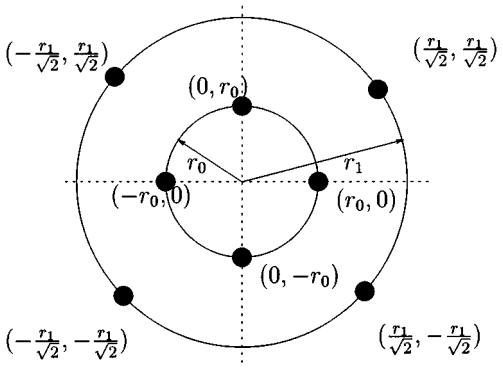

Fig. 6. 8-QAM alphabet.

$$\left\{1,\, 2\exp\left(-j\frac{\pi}{4}\right),\, 2\exp\left(-j\frac{\pi}{4}\right),\, 2\exp\left(-j\frac{\pi}{4}\right)\right\} \Longleftrightarrow \{000,\, 100,\, 100,\, 100\}.$$

Thus, the difference in representation is in only 1 bit for the worst case vectors. However, vectors within the same equivalence class may have bit representations by Gray labeling that are quite different and a connection between them is not obvious. For example, for 8-PSK and $N = 3$

$$\left\{1,\, \exp\left(j\frac{\pi}{2}\right),\, \exp\left(j\frac{3\pi}{2}\right)\right\} \Longleftrightarrow \{000,\, 001,\, 101\}$$

and

$$\left\{\exp\left(j\frac{\pi}{2}\right),\, \exp(j\pi),\, 1\right\} \Longleftrightarrow \{001,\, 110,\, 000\}$$

belong to the same equivalence class, but the relation between the bit representations for the two vectors is not clear.

2) Differential labeling is purely a matter of convenience, used to get simple implementations of the ideas of overlapped block encoding, differential modulation, and energy constraints. (Indeed, if simplicity of implementation is not an issue, the ideas of overlapped block encoding, differential modulation, and energy constraints can be thought of just in terms of the complex symbols and the modulation labeling scheme can be dispensed with.) Vectors within the same equivalence class have very similar representations by this labeling. For example, for 8-PSK and $N = 3$

$$\left\{1,\, \exp\left(j\frac{\pi}{2}\right),\, \exp\left(j\frac{3\pi}{2}\right)\right\} \Longleftrightarrow \{000,\, 010,\, 110\}$$

and

$$\left\{\exp\left(j\frac{\pi}{2}\right),\, \exp(j\pi),\, 1\right\} \Longleftrightarrow \{010,\, 100,\, 000\}$$

belong to the same equivalence class and adding $010 (\mathrm{mod}\, 8)$ to each of the bit representations in the first vector gives the second vector. However, equivalence classes that are close in the noncoherent metric may be very disparate in bit representations, thus causing a large number of potential bit errors for a worst case symbol vector pair. For example, for 8-QAM and $N = 4$, a worst case vector pair is given by

$$\left\{1,\, 1,\, 2\exp\left(-j\frac{\pi}{4}\right),\, 2\exp\left(-j\frac{\pi}{4}\right)\right\} \Longleftrightarrow \{000,\, 000,\, 111,\, 111\}$$

and

$$\left\{1,\, 2\exp\left(-j\frac{\pi}{4}\right),\, 2\exp\left(-j\frac{\pi}{4}\right),\, 2\exp\left(-j\frac{\pi}{4}\right)\right\} \Longleftrightarrow \{000,\, 111,\, 111,\, 111\}.$$

Thus, there is a difference of 3 bits in the representation as opposed to just 1 bit for Gray labeling.

In order to gain the benefits of Gray labeling, while preserving the simplicity of implementation due to differential labeling, we consider a communication system with a block diagram representation as in Fig. 9 for our simulations. Raw information bits

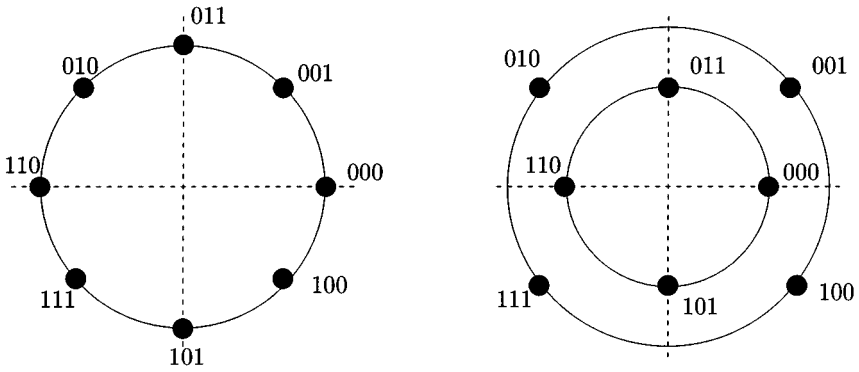

Fig. 7. Gray bit labeling for 8-PSK and 8-QAM alphabets.

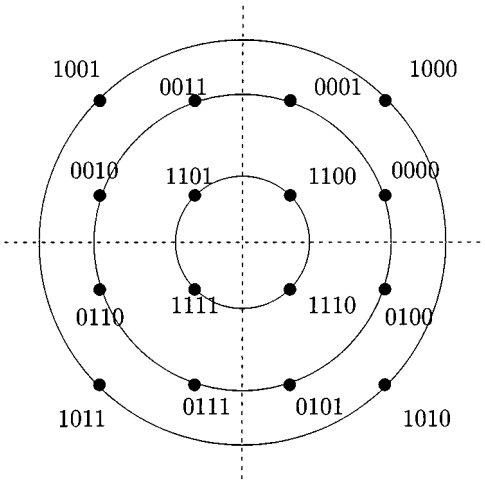

Fig. 8. A 16-QAM alphabet with Gray bit labeling.

are first parsed and reverse Gray coded.[5] Then, we obtain the differential labeling, which is used to implement energy constraints and to do overlapped block encoding. At the receiver, after demodulation, we get the differential labeling of the equivalence class, which is Gray coded to get the estimates of the information bits.

While the preceding scheme is general enough to apply to both QAM and PSK constellations, it reduces to the following standard implementation for DPSK.

1) Map the information bits to a sequence of PSK information symbols $\{a[i]\}$ using Gray coding.

2) Generate the transmitted symbols using differential modulation as $x[i] = x[i-1]a[i]$.

The multiplication operation involved in the differential modulation performed in Step 2) is equivalent to EXCLUSIVE-OR of the bitwise representation of the PSK symbols using a natural labeling. Step 1) can, therefore, be interpreted as reverse Gray coding, mapping from the information symbols to the natural bitwise representation of the PSK alphabet.

[5]Traditionally, the Gray coding procedure refers to the conversion from the natural bit labeling to the Gray bit labeling, e.g., $\exp(j\frac{\pi}{2})$ or $010 \rightarrow 011$ and $\exp(j\pi)$ or $100 \rightarrow 110$ for 8-PSK. So, the reverse procedure, which is done at the encoder, is referred to as reverse Gray coding.

B. Simulation Results

The suboptimal scheme of Section II-D can also be extended for the case of amplitude distortion introduced by the channel for QAM-type alphabets. However, the unbounded set of all possible amplitudes cannot be substituted by a set of finite amplitude estimates without significant loss of optimality, unless some estimate of the SNR over the block is available. Hence, we consider a decoder that works in two steps and is very similar to the decoding procedure for PSK in Section II-D.

1) **Coherent step**:
Choose a possible realization of the amplitude bits and a phase estimate. For the chosen amplitude realization and phase estimate, perform symbol-by-symbol coherent demodulation on the received vector to obtain an estimate of the transmitted vector.

2) **Noncoherent step**:
From among the candidate estimates (one for each phase estimate and amplitude realization), choose the one that yields the largest noncoherent decision metric.

The complexity of the algorithm is exponential in the number of subconstellations N_a (but not in the total alphabet size M). A comparison of the performance of 8-QAM and 8-PSK in simulations using the suboptimal algorithm (with a resolution of $L = 16$ for the channel phase shift) is now presented. The ratio of the amplitudes $(\frac{r_1}{r_0})$ used for the 8-QAM alphabet can be tuned to provide the optimal performance for a given coherence interval. For the simulations presented here, the ratio was maintained at $\frac{r_1}{r_0} = 2$, which is near-optimal for a large range of coherence intervals.

For the AWGN channel, the bit-error probability exhibits the following behavior:

$$\lim_{\eta \rightarrow \infty} \frac{1}{\eta} \log P_{b,\,\mathrm{AWGN}}(\eta) = -d_{\mathrm{nc}}^2 \tag{16}$$

where η is the SNR. Thus, the distance in (10) determines the asymptotic slopes of the curves in Fig. 10. The probability of a bit error for 8-QAM is noted to decay faster with SNR than that for 8-PSK, as predicted by the distance values.

The disparity in performance between QAM and PSK alphabets can be seen more clearly from the case of 16 signal points. A standard 16-QAM constellation as shown in Fig. 8 is compared with a 16-PSK constellation. An energy constraint is imposed upon the vectors from the QAM alphabet that ensures that

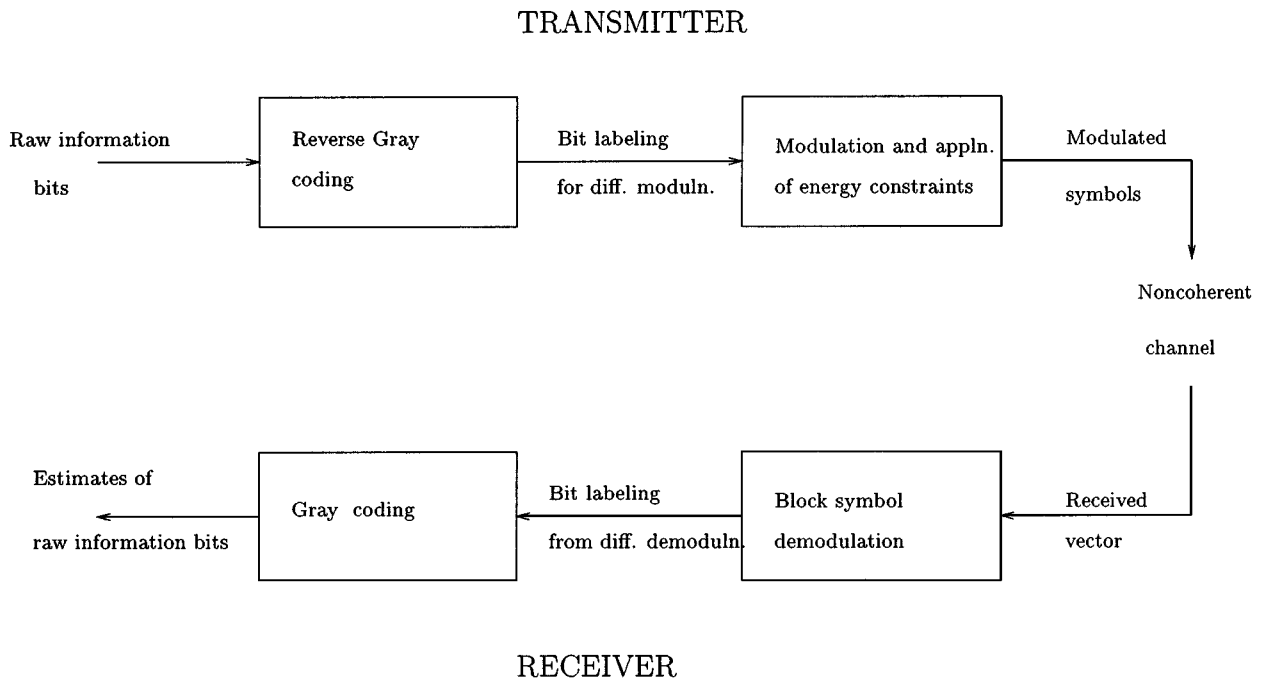

Fig. 9. Block diagram for a system implementation.

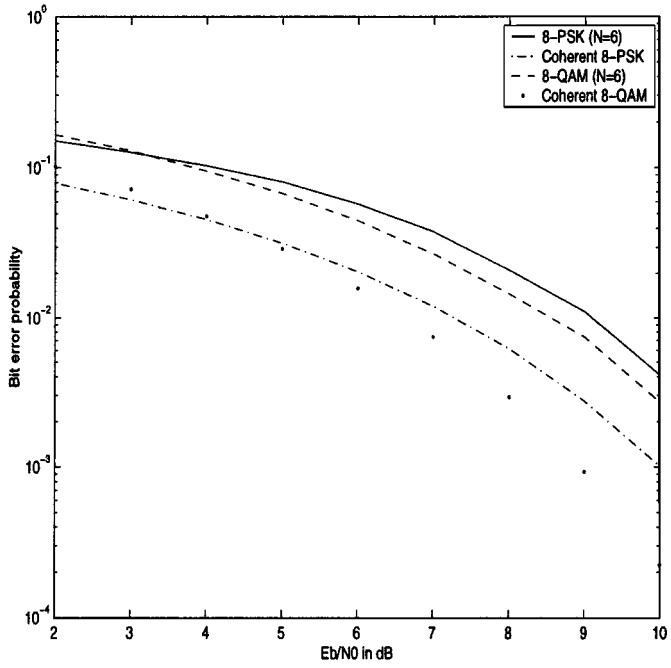

Fig. 10. Bit-error probabilities for 8-QAM and 8-PSK alphabets (AWGN).

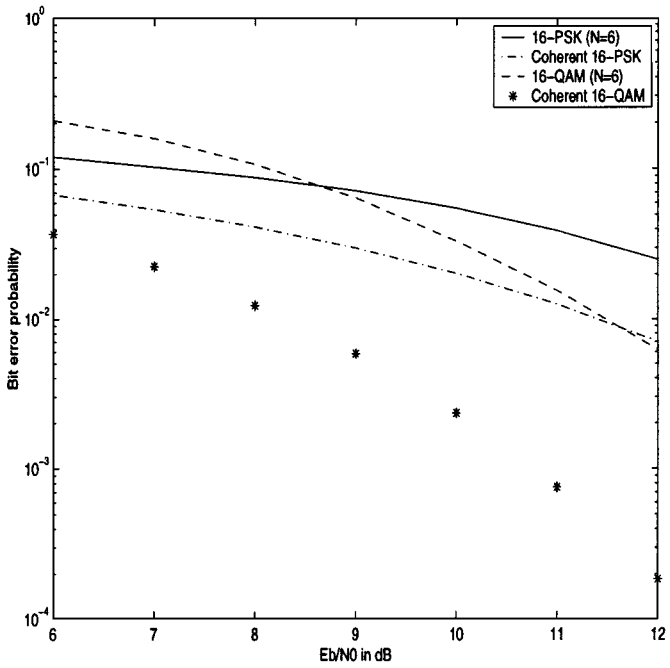

Fig. 11. Bit-error probabilities for 16-PSK and 16-QAM alphabets (AWGN, $N = 6$).

the number of elements from the lowest amplitude level in any vector is no more than the number of elements from the highest amplitude level. A valid codeword can be obtained from an invalid one by simply inverting the first bit in all of them, as was done in the 8-QAM case. Gray labeling is used to minimize the bit-error probability for both 16-QAM and 16-PSK. From the curves in Fig. 11, it is observed that 16-QAM affords an improvement of 2 dB over 16-PSK at high SNR, which is the regime of interest for such large constellations.

In the next section, theoretical results are presented to show that the coherent gain of QAM over PSK can be realized even in the noncoherent case as the coherence interval size gets large.

C. Asymptotic Results for the Noncoherent Distance

We present asymptotic results in this subsection and the next that prove that, as the coherence interval gets large, the performance of noncoherent detection can get arbitrarily close to that

of coherent detection. In particular, it is shown that the noncoherent distance (which, in general, is less than the coherent distance by Proposition 1) approaches the coherent distance for large N. In the next subsection, it is shown that, for the block fading channel model [3], the information rate for noncoherent detection approaches that for coherent detection for large coherence intervals. Together, the communication-theoretic and information-theoretic arguments imply that even for large coherence intervals (for which separate channel estimation followed by coherent detection does not require excessive overhead), signal designs based on the noncoherent metric should give performance competitive with those designed for coherent communication.

The worst case vector pair for PSK (as given by (12) and (13)) and those for QAM (as given by (14) and (15), for instance) have the following common property. They can be parsed into smaller parts which are scalar multiples of each other, i.e., they can be written (after a rearrangement of elements, if necessary) as

$$\boldsymbol{x_N} = [\hat{\boldsymbol{x}}_{\boldsymbol{k}};\ \tilde{\boldsymbol{x}}_{\boldsymbol{N-k}}] \qquad \boldsymbol{y_N} = [\hat{\boldsymbol{y}}_{\boldsymbol{k}};\ \tilde{\boldsymbol{y}}_{\boldsymbol{N-k}}]$$

where

$$\tilde{\boldsymbol{x}}_{\boldsymbol{N-k}} = \alpha \tilde{\boldsymbol{y}}_{\boldsymbol{N-k}} \quad \text{and} \quad \hat{\boldsymbol{x}}_{\boldsymbol{k}} = \beta \hat{\boldsymbol{y}}_{\boldsymbol{k}}$$

for some complex scalars α and β independent of k such that $|\alpha| \geq 1$ and $|\beta| \geq 1$. The vectors $\hat{\boldsymbol{x}}_{\boldsymbol{k}}, \hat{\boldsymbol{y}}_{\boldsymbol{k}}$ are of length k and $\tilde{\boldsymbol{x}}_{\boldsymbol{N-k}}, \tilde{\boldsymbol{y}}_{\boldsymbol{N-k}}$ are of length $(N-k)$, where the length k remains constant while N increases.

For such vectors, the following connection exists between the Euclidean distance and the GLRT metric.

Theorem 3: If

$$\lim_{N\to\infty} \frac{k}{N} = 0, \qquad \text{i.e., } k = o(N)$$

then

$$\lim_{N\to\infty} [d_{\text{nc}}(\boldsymbol{x_N}, \boldsymbol{y_N})^2 \times E_b] = \frac{\|\hat{\boldsymbol{x}}_{\boldsymbol{k}} - \alpha \hat{\boldsymbol{y}}_{\boldsymbol{k}}\|^2}{4}.$$

Thus, in the asymptotic limit, the noncoherent metric and the Euclidean distance are equivalent for such vector pairs. The proof is presented in Appendix D.

The following remarks follow from this result:

1) For an M-PSK alphabet, the asymptotic distance is given by
$$\frac{1}{4}\left|1 - \exp\left(j\frac{2\pi}{M}\right)\right|^2 = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi}{M}\right)\right]$$
which is the same as the coherent distance for M-PSK.

2) For an 8-QAM alphabet without any constraints, the vector pair of
$$\{1,\ 1,\ \ldots\} \quad \text{and} \quad \{1,\ 2\exp(j\pi/4),\ 2\exp(j\pi/4),\ \ldots\}$$
gives a normalized asymptotic distance of
$$\frac{1}{4E_b}\left|1 - \frac{1}{2\exp(j\pi/4)}\right|^2$$
which evaluates to a lower value than the corresponding distance for 8-PSK. This reiterates the need for energy constraints on signals constructed using the QAM alphabet.

With the application of our minimum energy criterion, the all-ones vector is removed. Now, for an 8-QAM alphabet, the minimum asymptotic distance (normalized) can be shown to be
$$\frac{1}{4E_b}|1 - \exp(j4\pi/M)|^2.$$
The evaluation of this minimum distance shows that the asymptotic gain obtained by using 8-QAM over 8-PSK is about 1.35 dB.

Thus, for an entire class of alphabets, the noncoherent distance asymptotically approaches the coherent distance. This generalizes earlier results [16] on the relation between the noncoherent and coherent metrics for PSK alphabets.

D. Asymptotic Results for the Block Fading Channel

In this subsection, we investigate the block fading channel model for slow-fading channels and show that the information rates per channel use are asymptotically the same for coherent and noncoherent detection as the coherence interval becomes large.

Consider a system with N_t transmitter antennae and one receiver antenna. (The case of multiple receiver antennas can be treated easily as an extension of the theory developed here.) Under the block fading channel model, the transmitted signal from each antenna is multiplied by a fading coefficient and, then, summed up at the receiver antenna along with AWGN. The fading coefficients are mutually independent circular Gaussian random variables. Further, they are assumed to remain constant for the coherence interval N, and, then, change to independent new values.

Mathematically

$$\boldsymbol{y} = \sqrt{\frac{\eta}{N_t}}\,\boldsymbol{X}\boldsymbol{h} + \boldsymbol{n} \tag{17}$$

where

- $\boldsymbol{X}$ is the transmitted signal matrix of size $N \times N_t$ with the energy constraint
$$\text{Trace}\left(\frac{\boldsymbol{E}[\boldsymbol{X}^H\boldsymbol{X}]}{N \times N_t}\right) \leq 1;$$
- $\boldsymbol{h}$ is the vector of channel coefficients of size $N_t \times 1$ which are zero mean, circular Gaussian with $\boldsymbol{E}[\boldsymbol{h}\boldsymbol{h}^H] = I_{N_t}$;
- $\boldsymbol{n}$ is a vector of AWGN of size $N \times 1$ with $\boldsymbol{E}[\boldsymbol{n}\boldsymbol{n}^H] = I_N$;
- $\boldsymbol{y}$ is the received vector of size $N \times 1$; and
- η is the SNR.

The mutual information between $\boldsymbol{X}$ and $\boldsymbol{y}$ for the coherent and noncoherent cases are respectively given by

$$I_c = I(\boldsymbol{X}; \boldsymbol{y}/\boldsymbol{h})$$
$$I_{\text{nc}} = I(\boldsymbol{X};\ \boldsymbol{y})$$

and the corresponding achievable rates per channel use are I_c/N and I_{nc}/N, respectively. We now compare these quantities for large values of the coherence interval N (see also [48]). The following relations hold:

$$\begin{aligned} I(\boldsymbol{X};\ \boldsymbol{y}/\boldsymbol{h}) &= I(\boldsymbol{X},\ \boldsymbol{h};\ \boldsymbol{y}) - I(\boldsymbol{h};\ \boldsymbol{y}) \\ &= I(\boldsymbol{X};\ \boldsymbol{y}) + I(\boldsymbol{y};\ \boldsymbol{h}/\boldsymbol{X}) - I(\boldsymbol{h};\ \boldsymbol{y}) \end{aligned}$$

Thus,

$$I_c - I_{\text{nc}} \leq I(\boldsymbol{y};\ \boldsymbol{h}/\boldsymbol{X}). \tag{18}$$

The left-hand side of the inequality in (18) gives the loss in information from using noncoherent detection as opposed to coherent detection. The inequality says that this loss is no more than the amount of information carried by the channel state ($\boldsymbol{h}$) about the received signal ($\boldsymbol{y}$) that cannot be obtained from the transmitted signal ($\boldsymbol{X}$). Our next step is to prove that this extra information becomes negligible as the coherence interval N gets large.

The term on the right-hand side in the inequality in (18) can also be viewed as the mutual information of a coherent "dual" channel, which has $\boldsymbol{h}$ as the input and $\boldsymbol{X}$ as the channel response. This mutual information is maximized when $\boldsymbol{y}$ is special complex Gaussian (see [1] for details), so that

$$I_c - I_{\mathrm{nc}} \leq \boldsymbol{E}\left\{\log\left[\det\left(I_{N_t} + \frac{\eta}{N_t}\boldsymbol{X}^H\boldsymbol{X}\right)\right]\right\}. \qquad (19)$$

Denoting the eigenvalues of $\boldsymbol{X}^H\boldsymbol{X}$ by $\{\lambda_i;\ i = 1, \ldots, N_t\}$ and using Jensen's inequality

$$\begin{aligned}\boldsymbol{E}&\left\{\log\left[\det\left(I_{N_t} + \frac{\eta}{N_t}\boldsymbol{X}\boldsymbol{X}^H\right)\right]\right\}\\ &= N_t\sum_{i=1}^{N_t}\frac{1}{N_t}\boldsymbol{E}\left[\log\left(1+\frac{\eta}{N_t}\lambda_i\right)\right]\\ &\leq N_t\log\left(1+\frac{\eta}{N_t}\sum_{i=1}^{N_t}\frac{1}{N_t}\boldsymbol{E}[\lambda_i]\right).\end{aligned}$$

By the energy constraint on $\boldsymbol{X}$

$$\mathrm{Trace}\left(\frac{\boldsymbol{E}[\boldsymbol{X}^H\boldsymbol{X}]}{N\times N_t}\right) = \frac{1}{N\times N_t}\sum_{i=1}^{N_t}\boldsymbol{E}[\lambda_i] \leq 1.$$

Thus,

$$\boldsymbol{E}\left\{\log\left[\det\left(I_{N_t} + \frac{\eta}{N_t}\boldsymbol{X}\boldsymbol{X}^H\right)\right]\right\} \leq N_t\log\left(1+\frac{\eta}{N_t}N\right).$$

Substituting in (19), we have

$$I_c - I_{nc} \leq N_t\log\left(1+\frac{\eta}{N_t}N\right)$$

which yields

$$\lim_{N\to\infty}\frac{I_{\mathrm{c}} - I_{\mathrm{nc}}}{N} = 0.$$

This implies that, while noncoherent detection may be most useful in practice for small coherence intervals, codes designed for noncoherent detection will perform well even for large coherence intervals.

V. Conclusion

The geometric interpretation of noncoherent detection via the GLRT gives insight into the form of the decision regions, and the noncoherent metric identified thereby provides a systematic framework for signal design within a coherence interval. In particular, the generalization of the notion of differential encoding to QAM constellations enables power- and bandwidth-efficient transmission for time-varying channels with high SNR. It was shown that the complexity of noncoherent detection could be reduced by approximating it by a number of parallel coherent detectors. Extensions of these ideas that take into account channel side information and specific signal structure are subjects for further investigation. Noncoherent detection is of particular interest in the multiple-antenna case, where channel estimation may incur a significant overhead. Extensions of the techniques considered here to the multiple-antenna case will be presented separately.

While noncoherent communication may well be the only practical option for channels with rapid time variations, we speculate that, in general, it is a good approach to detection even for channels with slow or no time variations. In particular, the results of this paper show that design based on a noncoherent metric provides good performance even when (complete or partial) channel information is available. For example, Proposition 1 implies that the pairwise error probability for a given pair of signals can only improve if channel information is available. Further, it is shown in Sections IV-C and IV-D that, for large coherence intervals, the performance and capacity of noncoherent communication approaches that of coherent systems. However, there are two main obstacles to realizing the promise of noncoherent communication:

1) Noncoherent detection is typically more complex than coherent detection (e.g., for linear modulation, block noncoherent demodulation is more complex than symbol-by-symbol coherent demodulation).
2) In coherent communication, systematic design principles that attain information-theoretic limits are now available [57], at least for the classical AWGN channel. However, progress on coding for noncoherent communication has been largely *ad hoc* in its nature, typically based on standard differential demodulation followed by standard differential decoding.

Thus, important topics for future investigation include reduction of detector and decoder complexity in noncoherent systems, and coding (over multiple coherence intervals) for noncoherent communication that approaches the Shannon capacity of important channel models, such as the block fading channel considered in Section IV-D. An approach that may be promising in this regard is the use of turbo-like codes, in conjunction with joint noncoherent demodulation and decoding using iterative techniques, as in [26] and [28].

Appendix A
Bounds on the Pairwise Error Probability for an AWGN Channel

The proof for Theorem 1 is provided here.

We first evaluate the distance from the signal point $\boldsymbol{x_1}$ to its decoding region boundary. This distance is given by

$$\inf_{\boldsymbol{y}\in T}\|\boldsymbol{y}-\boldsymbol{x_1}\|^2, \qquad \text{where } T = \left\{\boldsymbol{y}\colon \frac{|\langle\boldsymbol{y},\boldsymbol{x_2}\rangle|}{\|\boldsymbol{x_2}\|} \geq \frac{|\langle\boldsymbol{y},\boldsymbol{x_1}\rangle|}{\|\boldsymbol{x_1}\|}\right\}.$$

If $\boldsymbol{y} = \boldsymbol{s} + \tilde{\boldsymbol{s}}$ where $\boldsymbol{s}$ is in the span of $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ and $\tilde{\boldsymbol{s}}$ is perpendicular to it, then

$$\|\boldsymbol{y}-\boldsymbol{x_1}\|^2 = \|\boldsymbol{s}-\boldsymbol{x_1}\|^2 + \|\tilde{\boldsymbol{s}}\|^2.$$

Hence, $\tilde{\boldsymbol{s}} = 0$ for the optimal y.

Let $\boldsymbol{e_i}$ represent the unit vector in the direction of $\boldsymbol{x_i}$ for $i = 1, 2$. Now, if $\boldsymbol{y} = \alpha\boldsymbol{e_1} + \beta\boldsymbol{e_2}$, where α and β are complex scalars, then, after some algebra, it can be shown that the following relation holds:

$$|\langle \boldsymbol{y}, \boldsymbol{e_2}\rangle|^2 - |\langle \boldsymbol{y}, \boldsymbol{e_1}\rangle|^2 = (1-|r|^2)(|\beta|^2 - |\alpha|^2) \tag{20}$$

where $\rho = \langle \boldsymbol{e_1}, \boldsymbol{e_2}\rangle$. Thus, the original problem reduces to

$$\inf_{|\alpha|\le|\beta|} ||\alpha\boldsymbol{e_1} + \beta\boldsymbol{e_2} - ||\boldsymbol{x_1}||\boldsymbol{e_1}||^2.$$

Solving, using Lagrange multipliers, the optimal values are

$$\alpha = \tfrac{1}{2}||\boldsymbol{x_1}|| \tag{21}$$

$$\beta = \frac{1}{2}||\boldsymbol{x_1}||\frac{\rho}{|\rho|} \tag{22}$$

and

$$\tilde{\boldsymbol{y}} = \frac{1}{2}||\boldsymbol{x_1}||\left(\boldsymbol{e_1} + \frac{\rho}{|\rho|}\boldsymbol{e_2}\right). \tag{23}$$

The minimum distance is given by

$$d^2(\boldsymbol{x_1}, \boldsymbol{x_2}) = \frac{||\boldsymbol{x_1}||^2}{2}\left(1 - \frac{|\langle \boldsymbol{x_1}, \boldsymbol{x_2}\rangle|}{||\boldsymbol{x_1}||||\boldsymbol{x_2}||}\right) \tag{24}$$

as required.

We can derive an upper bound on the error probability (and the error rate) by noting that an error will only occur if the noise vector has energy larger than the squared minimum distance obtained in (24). Accordingly

$$\begin{aligned} P[(\hat{\boldsymbol{x}} = \boldsymbol{x_2}/\boldsymbol{x} = \boldsymbol{x_1})] &\le P[||\boldsymbol{y} - \boldsymbol{x_1}||^2 \ge d^2(\boldsymbol{x_1}, \boldsymbol{x_2})] \\ &= \exp\left(-\frac{d^2(\boldsymbol{x_1}, \boldsymbol{x_2})}{2\sigma^2}\right)\sum_{i=0}^{N-1}\frac{[d^2(\boldsymbol{x_1}, \boldsymbol{x_2})/(2\sigma^2)]^i}{i!}. \end{aligned} \tag{25}$$

Evaluating the error rate

$$\limsup_{\sigma\to 0} 2\sigma^2 \log[P[(\hat{\boldsymbol{x}} = \boldsymbol{x_2}/\boldsymbol{x} = \boldsymbol{x_1})]] \le -d^2(\boldsymbol{x_1}, \boldsymbol{x_2}). \tag{26}$$

For a lower bound on the error probability, we use a particular error event (noise along the minimum distance direction) which is most likely to happen in the asymptotic regime and has an error rate close to the squared minimum distance.

Let $\boldsymbol{w}$ denote a unit vector in the direction $(\tilde{\boldsymbol{y}} - \boldsymbol{x_1})$, as in Fig. 12, given by

$$\begin{aligned} \boldsymbol{w} &= \frac{(\tilde{\boldsymbol{y}} - \boldsymbol{x_1})}{d(\boldsymbol{x_1}, \boldsymbol{x_2})} \\ &= \frac{-||\boldsymbol{x_1}||}{2d(\boldsymbol{x_1}, \boldsymbol{x_2})}\boldsymbol{e_1} + \frac{||\boldsymbol{x_1}||}{2d(\boldsymbol{x_1}, \boldsymbol{x_2})}\frac{\rho}{|\rho|}\boldsymbol{e_2} \end{aligned} \tag{27}$$

and $\boldsymbol{w}^{\perp}$ denote the vector perpendicular to it in the span of $\boldsymbol{e_1}$ and $\boldsymbol{e_2}$. Only the noise components in the span of $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ need to be considered, and, hence, we can assume, without loss of generality (w.l.o.g.), that the received vector is of the form

$$\boldsymbol{y}(\alpha_1, \alpha_2) := \tilde{\boldsymbol{y}} + \alpha_1\boldsymbol{w} + \alpha_2\boldsymbol{w}^{\perp}$$

for some values of α_1 and α_2. We now consider a vector obtained by a slight perturbation from $\tilde{\boldsymbol{y}}$ into the decoding region for $\boldsymbol{x_2}$, given by

$$\begin{aligned} \boldsymbol{y}(\epsilon_1, 0) &= \tilde{\boldsymbol{y}} + \epsilon_1\boldsymbol{w} \\ &= \frac{||\boldsymbol{x_1}||}{2}\left(1 - \frac{\epsilon_1}{d(\boldsymbol{x_1}, \boldsymbol{x_2})}\right)\boldsymbol{e_1} \\ &\quad + \frac{||\boldsymbol{x_1}||}{2}\frac{\rho}{|\rho|}\left(1 + \frac{\epsilon_1}{d(\boldsymbol{x_1}, \boldsymbol{x_2})}\right)\boldsymbol{e_2} \end{aligned}$$

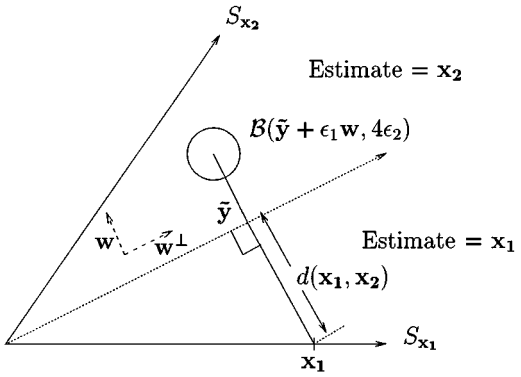

Fig. 12. Signal space geometry.

where $\epsilon_1 > 0$. Then, from (20), the following strict inequality holds:

$$|\langle \boldsymbol{y}(\epsilon_1, 0), \boldsymbol{e_1}\rangle| < |\langle \boldsymbol{y}(\epsilon_1, 0), \boldsymbol{e_2}\rangle|.$$

By the continuity of the inner product in its arguments, we can find a ball of radius $4\epsilon_2$ (where $\epsilon_2 < \epsilon_1$) around $\boldsymbol{y}(\epsilon_1, 0)$, denoted by $\mathcal{B}(\boldsymbol{y}(\epsilon_1, 0), 4\epsilon_2)$, that is inside the decoding region for $\boldsymbol{x_2}$, or

$$\mathcal{B}(\boldsymbol{y}(\epsilon_1, 0), 4\epsilon_2) \subseteq \{\boldsymbol{y}\colon |\langle \boldsymbol{y}, \boldsymbol{e_1}\rangle| \le |\langle \boldsymbol{y}, \boldsymbol{e_2}\rangle|\}.$$

Also, since

$$||\boldsymbol{y}(\alpha_1, \alpha_2) - \boldsymbol{y}(\epsilon_1, 0)||^2 = |\alpha_1 - \epsilon_1|^2 + |\alpha_2|^2$$

we have

$$\begin{aligned} &\{\boldsymbol{y}(\alpha_1, \alpha_2)\colon |\alpha_1 - \epsilon_1| \le 2\epsilon_2 \text{ and } |\alpha_2| \le 2\epsilon_2\} \\ &\qquad \subseteq \mathcal{B}(\boldsymbol{y}(\epsilon_1, 0), 4\epsilon_2) \subseteq \{\boldsymbol{y}\colon |\langle \boldsymbol{y}, \boldsymbol{e_1}\rangle| \le |\langle \boldsymbol{y}, \boldsymbol{e_2}\rangle|\}. \end{aligned}$$

Finally, for a received vector $\boldsymbol{y}(\alpha_1, \alpha_2)$, the components of noise along $\boldsymbol{w}$ and $\boldsymbol{w}^{\perp}$ are, respectively, $d(\boldsymbol{x_1}, \boldsymbol{x_2}) + \alpha_1$ and α_2. Thus,

$$\begin{aligned} &P[(\hat{\boldsymbol{x}} = \boldsymbol{x_2}/\boldsymbol{x} = \boldsymbol{x_1})] \\ &\quad \ge P[|\alpha_1 - \epsilon_1| \le 2\epsilon_2 \text{ and } |\alpha_2| \le 2\epsilon_2] \\ &\quad = P[|\langle \boldsymbol{n}, \boldsymbol{w}\rangle - d(\boldsymbol{x_1}, \boldsymbol{x_2}) - \epsilon_1| \le 2\epsilon_2 \text{ and } |\langle \boldsymbol{n}, \boldsymbol{w}^{\perp}\rangle| \le 2\epsilon_2] \\ &\quad \ge P[|\mathrm{Re}(\langle \boldsymbol{n}, \boldsymbol{w}\rangle) - d(\boldsymbol{x_1}, \boldsymbol{x_2}) - \epsilon_1| \le \epsilon_2 \\ &\qquad \text{and } |\mathrm{Im}(\langle \boldsymbol{n}, \boldsymbol{w}\rangle)| \le \epsilon_2 \text{ and } |\langle \boldsymbol{n}, \boldsymbol{w}^{\perp}\rangle| \le 2\epsilon_2] \\ &\quad = P[|\mathrm{Re}(\langle \boldsymbol{n}, \boldsymbol{w}\rangle) - d(\boldsymbol{x_1}, \boldsymbol{x_2}) - \epsilon_1| \le \epsilon_2] \\ &\qquad \times P[|\mathrm{Im}(\langle \boldsymbol{n}, \boldsymbol{w}\rangle)| \le \epsilon_2] \times P[|\langle \boldsymbol{n}, \boldsymbol{w}^{\perp}\rangle| \le 2\epsilon_2] \\ &\quad = \left(Q\left[\frac{d(\boldsymbol{x_1}, \boldsymbol{x_2}) + \epsilon_1 - \epsilon_2}{\sigma}\right] - Q\left[\frac{d(\boldsymbol{x_1}, \boldsymbol{x_2}) + \epsilon_1 + \epsilon_2}{\sigma}\right]\right) \\ &\qquad \times \left(1 - 2Q\left[\frac{\epsilon_2}{\sigma}\right]\right) \times \left(1 - 2Q\left[\frac{\epsilon_2}{\sigma}\right]\right) \end{aligned}$$

where $Q[x] = \int_x^{\infty}\frac{1}{\sqrt{2\pi}}\exp(-\frac{t^2}{2})\,dt$.

Evaluating the error rate

$$\begin{aligned} &\liminf_{\sigma\to 0} 2\sigma^2 \log P[(\hat{\boldsymbol{x}} = \boldsymbol{x_2}/\boldsymbol{x} = \boldsymbol{x_1})] \\ &\quad \ge \liminf_{\sigma\to 0} 2\sigma^2 \log\left(Q\left[\frac{d(\boldsymbol{x_1}, \boldsymbol{x_2}) + \epsilon_1 - \epsilon_2}{\sigma}\right]\right. \\ &\qquad \left. - Q\left[\frac{d(\boldsymbol{x_1}, \boldsymbol{x_2}) + \epsilon_1 + \epsilon_2}{\sigma}\right]\right) \\ &\quad = -(d(\boldsymbol{x_1}, \boldsymbol{x_2}) + \epsilon_1 - \epsilon_2)^2 \end{aligned}$$

using $\lim_{x\to\infty} -\frac{2}{x^2}\log Q[x] = 1$. Since ϵ_1 and ϵ_2 can be made arbitrarily small

$$\liminf_{\sigma\to 0} 2\sigma^2 \log P[(\hat{\boldsymbol{x}} = \boldsymbol{x_2}/\boldsymbol{x} = \boldsymbol{x_1})] \geq -d(\boldsymbol{x_1}, \boldsymbol{x_2})^2. \quad (28)$$

Combining (26) and (28)

$$\lim_{\sigma\to 0} 2\sigma^2 \log P[(\hat{\boldsymbol{x}} = \boldsymbol{x_2}/\boldsymbol{x} = \boldsymbol{x_1})] = -d(\boldsymbol{x_1}, \boldsymbol{x_2})^2$$

as required.

Appendix B
An Upper Bound to the Pairwise Error Probability Using the Chernoff Bound

The proof of Theorem 2 is provided here.

If $\boldsymbol{x_1}$ is the transmitted signal, a decoding error is made if there exists $\boldsymbol{x_2} \in \mathcal{S}$ such that $\boldsymbol{x_2} \neq \boldsymbol{x_1}$ and

$$\frac{|\langle \boldsymbol{y}, \boldsymbol{x_2}\rangle|}{||\boldsymbol{x_2}||} \geq \frac{|\langle \boldsymbol{y}, \boldsymbol{x_1}\rangle|}{||\boldsymbol{x_1}||}.$$

The pairwise probability of error is, then, given by

$$P\left[\left|\left\langle \boldsymbol{y}, \frac{\boldsymbol{x_2}}{||\boldsymbol{x_2}||}\right\rangle\right|^2 - \left|\left\langle \boldsymbol{y}, \frac{\boldsymbol{x_1}}{||\boldsymbol{x_1}||}\right\rangle\right|^2 \geq 0 \middle/ \boldsymbol{y} = \boldsymbol{x_1} + \boldsymbol{n}\right].$$

By the Chernoff bound, an upper bound to this probability is given by

$$\inf_{\lambda\geq 0} E\left[\exp\left\{\lambda\left(\left|\left\langle \boldsymbol{x_1}+\boldsymbol{n}, \frac{\boldsymbol{x_2}}{||\boldsymbol{x_2}||}\right\rangle\right|^2 - \left|\left\langle \boldsymbol{x_1}+\boldsymbol{n}, \frac{\boldsymbol{x_1}}{||\boldsymbol{x_1}||}\right\rangle\right|^2\right)\right\}\right].$$

For a fixed value of λ, the expectation above gives

$$\frac{1}{1-4\lambda^2\sigma^4+4\lambda^2\sigma^4|\rho|^2}\exp\left[-\frac{||\boldsymbol{x_1}||^2\lambda(1-2\lambda\sigma^2)(1-|\rho|^2)}{1-4\lambda^2\sigma^4+4\lambda^2\sigma^4|\rho|^2}\right]$$

if $(1-4\lambda^2\sigma^4+4\lambda^2\sigma^4|\rho|^2) \geq 0$. Using

$$\lambda = \frac{1}{2\sigma^2(1+|\rho|)}$$

an upper bound to the probability of error is given by

$$\frac{1+|\rho|}{2|\rho|}\exp\left[-\frac{1}{2\sigma^2}\frac{||x_1||^2(1-|\rho|)}{2}\right].$$

Thus, the same exponential factor as is given by the GLRT in (7) is obtained here.

Appendix C
Nearest Neighbors for Differential M-PSK

Consider vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ from two different equivalence classes in $\mathcal{S}_{\rm nc}$, given by

$$\boldsymbol{x}_1 = \{1, \exp[j\phi_2], \ldots, \exp[j\phi_N]\}$$

and

$$\boldsymbol{x}_2 = \{1, \exp[j\theta_2], \ldots, \exp[j\theta_N]\}$$

where θ_i and ϕ_i are of the form $\frac{2\pi k}{M}$, $k = 0, \ldots, M-1$ for all i. From (10), since

$$||\boldsymbol{x}_i||^2 = N, \qquad \forall\, \boldsymbol{x}_i \in \mathcal{S}$$

we only need to find a pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ such that $|\langle \boldsymbol{x}_1, \boldsymbol{x}_2\rangle|$ is maximized

$$\begin{aligned}|\langle \boldsymbol{x}_1, \boldsymbol{x}_2\rangle|^2 &= |\exp[j(\theta_1-\phi_1)] + \cdots + \exp[j(\theta_N-\phi_N)]|^2 \\ &= N + 2\sum_{i=1}^{N}\sum_{j=i+1}^{N}\cos[\alpha_i - \alpha_j]\end{aligned}$$

where $\alpha_i = \theta_i - \phi_i$ is also of the form $\frac{2\pi k}{M}$, $k = 0, \ldots, M-1$ for all i.

Since $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are not equivalent, not all α_i are equal to 0. In order to maximize the inner product under this condition, it suffices to have

$$\alpha_1 = \alpha_2 = \cdots = \alpha_{N-1} = 0 \quad \text{and} \quad \alpha_N = \frac{2\pi}{M}.$$

Thus,

$$\tilde{\boldsymbol{x}}_{\boldsymbol{1}} = \{\underbrace{1, 1, \ldots, 1}_{(N-1)\text{ times}}, 1\}$$

and

$$\tilde{\boldsymbol{x}}_{\boldsymbol{2}} = \left\{\underbrace{1, 1, \ldots, 1}_{(N-1)\text{ times}}, \exp\left[j\frac{2\pi}{M}\right]\right\}$$

achieve the minimum value of $d_{\rm nc}(\boldsymbol{x_1}, \boldsymbol{x_2})$ for $\boldsymbol{x_1}, \boldsymbol{x_2} \in \mathcal{S}_{\rm nc}$.

Appendix D
Proof of the Asymptotic Equivalence of the GLRT Metric and the Euclidean Distance

The proof of Theorem 3 is provided here.

Assume, w.l.o.g., that

$$||\boldsymbol{x}_N|| \leq ||\boldsymbol{y}_N||.$$

The GLRT metric of (9) then simplifies as follows:

$$d_{\rm nc}(\boldsymbol{x_N}, \boldsymbol{y_N})^2 \times E_b = \frac{1}{2||\boldsymbol{y_N}||^2}\frac{||\boldsymbol{x_N}||^2||\boldsymbol{y_N}||^2 - |\langle \boldsymbol{x_N}, \boldsymbol{y_N}\rangle|^2}{1+\frac{|\langle \boldsymbol{x_N}, \boldsymbol{y_N}\rangle|}{||\boldsymbol{x_N}||||\boldsymbol{y_N}||}}.$$

Simplifying individual terms

$$\begin{aligned}|\langle \boldsymbol{x_N}, \boldsymbol{y_N}\rangle|^2 &= |\alpha|^2||\hat{\boldsymbol{y}}_{\boldsymbol{k}}||^4 + |\beta|^2||\tilde{\boldsymbol{y}}_{\boldsymbol{N-k}}||^4 \\ &\quad + 2\mathrm{Re}[\alpha\beta^*]||\hat{\boldsymbol{y}}_{\boldsymbol{k}}||^2||\tilde{\boldsymbol{y}}_{\boldsymbol{N-k}}||^2 \\ ||\boldsymbol{x_N}||^2||\boldsymbol{y_N}||^2 &= |\alpha|^2||\hat{\boldsymbol{y}}_{\boldsymbol{k}}||^4 + |\beta|^2||\tilde{\boldsymbol{y}}_{\boldsymbol{N-k}}||^4 \\ &\quad + ||\hat{\boldsymbol{x}}_{\boldsymbol{k}}||^2||\tilde{\boldsymbol{y}}_{\boldsymbol{N-k}}||^2 + ||\hat{\boldsymbol{y}}_{\boldsymbol{k}}||^2||\tilde{\boldsymbol{x}}_{\boldsymbol{N-k}}||^2.\end{aligned}$$

Thus,

$$d_{\rm nc}(\boldsymbol{x_N}, \boldsymbol{y_N})^2 \times E_b = \frac{||\tilde{\boldsymbol{y}}_{\boldsymbol{N-k}}||^2}{2||\boldsymbol{y_N}||^2}\frac{||\hat{\boldsymbol{x}}_k - \alpha\hat{\boldsymbol{y}}_k||^2}{1+|s_N|} \quad (29)$$

where $s_N = \frac{\langle \boldsymbol{x_N}, \boldsymbol{y_N}\rangle}{||\boldsymbol{x_N}||||\boldsymbol{y_N}||}$. Now

$$\frac{||\hat{\boldsymbol{y}}_{\boldsymbol{k}}||^2}{||\tilde{\boldsymbol{y}}_{\boldsymbol{N-k}}||^2} \leq \frac{kr_1^2}{(N-k)r_0^2} \overset{N\to\infty}{\longrightarrow} 0$$

and, hence,

$$\frac{||\tilde{\boldsymbol{y}}_{\boldsymbol{N-k}}||^2}{||\boldsymbol{y_N}||^2} \overset{N\to\infty}{\longrightarrow} 1.$$

Using these

$$\lim_{N\to\infty} |s_N|^2 = 1.$$

Substituting in (29), the required result is obtained.

ACKNOWLEDGMENT

The authors are grateful to the anonymous reviewers for their valuable comments.

REFERENCES

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Labs, Tech. Rep. BL0112170-950615-07TM, 1995.
[2] B. Hochwald and T. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," in *Proc. 36th Annu. Allerton Conf. Communication, Control and Computing*, Monticello, IL, 1998.
[3] T. L. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, Jan. 1999.
[4] B. Hochwald and W. Sweldens, "Differential unitary space-time modulation," IEEE Trans. Commun., submitted for publication.
[5] B. Hassibi, B. M. Hochwald, A. Shokrollahi, and W. Sweldens, "Multiple antennas and representation theory," in *Proc. 2000 Int. Symp. Information Theory (ISIT 2000)*, Sorrento, Italy, June 2000, p. 337.
[6] B. L. Hughes, "Extensions to the theory of differential space–time modulation," in *Proc. 2000 Int. Symp. Information Theory (ISIT 2000)*, Sorrento, Italy, June 2000, p. 285.
[7] ——, "Differential space-time modulation," in *Proc. IEEE Wireless Networking and Communications Conf. (WCNC'99)*, Sept. 1999, p. 285.
[8] B. Hochwald, T. Marzetta, T. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space–time constellations," *IEEE Trans. Inform. Theory*, submitted for publication.
[9] U. Madhow, L. J. Zhu, and L. Galup, "Differential MMSE: New adaptive algorithms for equalization, interference suppression, and beamforming," in *Proc. 32nd Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Oct. 1998, pp. 640–644.
[10] European Telecommunication Standard, "Radio broadcasting systems: Digital audio broadcasting (DAB) to mobile, portable and fixed receivers," ETSI, Tech. Rep. ETS 300 401, May 1997.
[11] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55–67, Jan. 1982.
[12] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Trans. Commun.*, vol. 44, pp. 1261–1271, Oct. 1996.
[13] J. Huber, U. Wachsmann, and R. Fischer, "Coded modulation by multilevel-codes: Overview and state of the art," *IEEE Trans. Inform. Theory*, vol. 44, pp. 927–946, May 1998.
[14] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.
[15] E. Biglieri, D. Divsalar, P. J. McLane, and M. Simon, *Introduction to Trellis-Coded Modulation with Applications*. New York: MacMillan, 1991.
[16] D. Divsalar and M. K. Simon, "Multiple-symbol differential detection of MPSK," *IEEE Trans. Commun.*, vol. 38, pp. 300–308, Mar. 1990.
[17] T. Giallorenzi and S. Wilson, "Noncoherent demodulation techniques for trellis coded M-DPSK signals," *IEEE Trans. Commun.*, vol. 43, pp. 2370–2380, Aug. 1995.
[18] D. Raphaeli, "Noncoherent coded modulation," *IEEE Trans. Commun.*, vol. 44, pp. 172–183, Feb. 1996.
[19] F. Adachi and M. Sawahashi, "Viterbi-decoding differential detection of DPSK," *Electron. Lett.*, vol. 28, no. 23, pp. 2196–2198, Nov. 1992.
[20] F. Adachi, "Reduced state transition Viterbi differential detection of M-ary DPSK signals," *Electron. Lett.*, vol. 32, no. 12, pp. 1064–1066, June 1996.
[21] F. Edbauer, "Bit error rate of binary and quaternary DPSK signals with multiple differential feedback detection," *IEEE Trans. Commun.*, vol. 40, pp. 457–460, Mar. 1992.
[22] F. Adachi and M. Sawahashi, "Decision feedback differential detection of differentially encoded 16APSK signals," *IEEE Trans. Commun.*, vol. 44, pp. 416–418, Apr. 1996.
[23] R. Schober, W. Gerstacker, and J. Huber, "Decision-feedback differential detection scheme for 16-DAPSK," *Electron. Lett.*, vol. 34, no. 19, pp. 1812–1813, Sept. 1998.
[24] D. Divsalar, M. K. Simon, and M. Shahshahani, "The performance of trellis-coded MDPSK with multiple symbol detection," *IEEE Trans. Commun.*, vol. 38, pp. 1391–1403, Sept. 1990.
[25] R. Van Nobelen and D. Taylor, "Multiple symbol differentially detected multilevel codes for the Rayleigh fading channel," *IEEE Trans. Commun.*, vol. 45, pp. 1529–1537, Dec. 1997.
[26] M. Peleg and S. Shamai, "Iterative decoding of coded and interleaved noncoherent multiple symbol detected DPSK," *Electron. Lett.*, vol. 33, no. 12, pp. 1018–1020, June 1997.
[27] J. Ventura-Traveset, G. Caire, E. Biglieri, and G. Taricco, "Impact of diversity reception on fading channels with coded modulation—part II: Differential block detection," *IEEE Trans. Commun.*, vol. 45, pp. 676–686, June 1997.
[28] R.-R. Chen, D. Agrawal, and U. Madhow, "Noncoherent detection of factor-graph codes over fading channels," in *Proc. 2000 Conf. Information Sciences and Systems (CISS 2000)*, Princeton, NJ, Mar. 2000.
[29] F. W. Sun and H. Leib, "Multiple-phase codes for detection without carrier phase reference," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1477–1491, July 1998.
[30] W. Webb, L. Hanzo, and R. Steele, "Bandwidth efficient QAM schemes for Rayleigh fading channels," *Proc. Inst. Elec. Eng.*, vol. 138, no. 3, pp. 169–175, June 1991.
[31] N. Svensson, "On differentially encoded star 16QAM with differential detection and diversity," *IEEE Trans. Veh. Technol.*, vol. 44, pp. 586–593, Aug. 1995.
[32] M. Shensa, "Quotient coding for fading channels," *IEEE Trans. Veh. Technol.*, vol. 47, pp. 499–505, May 1998.
[33] W. Weber, III, "Differential encoding for multiple amplitude and phase shift keying systems," *IEEE Trans. Commun.*, vol. COM-26, pp. 385–391, Mar. 1978.
[34] J. Anderson, T. Aulin, and C.-E. Sundberg, *Digital Phase Modulation*, 1st ed. New York: Plenum, 1986.
[35] D. Bouras, P. Mathiopoulos, and D. Makrakis, "Optimal detection of coded differentially encoded QAM and PSK signals with diversity reception in correlated fast Rician fading channels," *IEEE Trans. Veh. Technol.*, vol. 42, pp. 245–258, Aug. 1993.
[36] D. Divsalar and M. K. Simon, "Maximum-likelihood differential detection of uncoded and trellis coded amplitude phase modulation over AWGN and fading channels—Metrics and performance," *IEEE Trans. Commun.*, vol. 42, pp. 76–89, Jan. 1994.
[37] G. Vitetta, U. Mengali, and D. Taylor, "Optimal noncoherent detection of FSK signals transmitted over linearly time-selective Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 45, pp. 1417–1425, Nov. 1997.
[38] ——, "Error probability with incoherent diversity reception of FSK signals transmitted over fast Rician fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 1443–1447, Nov. 1998.
[39] I. Korn, "Error probability of M-ary FSK with differential phase detection in satellite mobile channel," *IEEE Trans. Veh. Technol.*, vol. 38, pp. 76–85, May 1989.
[40] G. Corazza and R. De Gaudenzi, "Analysis of coded noncoherent transmission in DS-CDMA mobile satellite communications," *IEEE Trans. Commun.*, vol. 46, pp. 1525–1535, Nov. 1998.
[41] M. Simon and A. Alouini, "A unified approach to the probability of error for noncoherent and differentially coherent modulations over generalized fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 1625–1638, Dec. 1998.
[42] C. Chang and P. McLane, "Bit-error-probability for noncoherent orthogonal signals in fading with optimum combining for correlated branch diversity," *IEEE Trans. Inform. Theory*, vol. 43, pp. 263–274, Jan. 1997.
[43] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2148–2177, Oct. 1998.
[44] O. Zeitouni, J. Ziv, and N. Merhav, "When is the generalized likelihood ratio test optimal?," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1597–1602, Sept. 1992.
[45] E. Visotsky and U. Madhow, "Multiuser detection for CDMA systems with nonlinear modulation," *IEEE Trans. Inform. Theory*, submitted for publication.
[46] M. K. Varanasi and A. Russ, "Noncoherent decorrelative detection for nonorthogonal multipulse modulation over the multiuser Gaussian channel," *IEEE Trans. Commun.*, vol. 46, pp. 1675–1684, Dec. 1998.
[47] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed. Berlin, Germany: Springer-Verlag, 1978.

[48] M. Peleg and S. Shamai (Shitz), "On the capacity of the blockwise incoherent MPSK channel," *IEEE Trans. Commun.*, vol. 46, pp. 603–609, May 1998.

[49] G. Taricco and M. Elia, "Capacity of fading channels with no side information," *Electron. Lett.*, vol. 33, no. 16, pp. 1368–1370, July 1997.

[50] N. J. A. Sloane. Packing planes in four dimensions and other mysteries. [Online]. Available: http://www.research.att.com/~njas

[51] G. Kabatyanskii and V. Levenshtein, "Bounds for packings on a sphere and in space," *Probl. Inform. Transm.*, vol. 14, no. 1, pp. 1–17, Jan.–Mar. 1978.

[52] D. Agrawal, T. Richardson, and R. Urbanke, "Multiple antenna signal constellations for fading channels," *IEEE Trans. Inform. Theory*, submitted for publication.

[53] D. Warrier, "A framework for spectrally efficient noncoherent communication," Ph.D. dissertation, Univ. Illinois, Urbana-Champaign, 2000.

[54] D. Warrier, U. Madhow, and R. Koetter, "Coding for noncoherent communication," in *Proc. Int. Symp. Information Theory (ISIT'00)*, Sorrento, Italy, June 2000, p. 76.

[55] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 1st ed. New York: Springer-Verlag, 1988.

[56] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[57] D. Costello, J. Hagenauer, H. Imai, and S. Wicker, "Applications of error-control coding," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2531–2560, Oct 1998.