Noncoherent Block Detection in the Presence of DC Offset

Char-Dir Chung, *Senior Member, IEEE*

Abstract— In this paper, we deal with noncoherent detection of a digitally phasor block-modulated signal in the additive white Gaussian noise channel when a direct-current (DC) offset is present in the receiver baseband processing. By processing the received baseband signal block by block, a generalized linear transform is used to remove the offset prior to data detection, thereby releasing the succeeding detection process from the threat of DC offset. Operating on transform output blocks, a generalized maximum-likelihood scheme is developed for noncoherent data detection without *a priori* knowledge of channel amplitude and phase. When all the signaling blocks are confined within the space expanded by the basis vectors obtained from the offset-removal transform matrix, the proposed detection scheme can exploit the advantage of performing data detection and estimation on channel amplitude and phase jointly in the maximum-likelihood sense. It is analytically shown that the block detection scheme provides the bit error performance asymptotically approaching that of the corresponding ideal coherent phase-shift-keyed (PSK) detection in the absence of DC offset when the block length is increased. An iterative detection scheme is also modified from the block detection scheme to simplify the realization complexity. Both block and iterative detection schemes are shown to outperform the conventional training-sequence-aided PSK detection scheme under the same transmission throughput efficiency.

Index Terms— DC offset, block modulation, noncoherent block detection, iterative detection.

I. INTRODUCTION

IN certain wireless communication circumstances, the receiver is disturbed by the presence of a direct-current (DC) offset in the baseband processing when an unwanted constant or slowly time-varying offset is generated after converting the received waveform down to baseband. For most of cases, a DC offset is present at baseband when a tone on carrier frequency is processed in association with the desired waveform during down conversion. This may occur in applications where a pilot tone is purposely transmitted with the modulation waveform to enable tone calibration and thus assist signal reception at the receiver [1]-[4], where an interfering or jamming tone is unfriendly received with the desired signal [5]-[7], and where a carrier tone signal is leaked from the receiver local oscillator to the receiver frontend and reflected off the antenna when the direct-conversion receiver architecture is used [8]-[14]. In some communication circuits, a DC offset may also be induced when there exists transistor mismatch in the signal path following the down converter [8].

In practical applications, the redundant DC offset is detrimental to signal demodulation and has to be compensated for or removed by appropriate designs. By shaping the transmitted signaling waveform spectrally (e.g., using a signal spectrum with a notch on carrier frequency [1], [2]) or temporally (e.g., using a pulse with zero DC [3]), the DC offset can be completely eliminated by deliberatively filtering the received waveform in the receiver. Because the signaling waveform is shaped with constraints, spectral efficiency is inevitably traded off by this stream of approaches. Another stream of approaches is to eliminate or reduce to some extent the DC offset by careful frontend design [8], [9], e.g., using alternating-current (AC) coupling filters in the down conversion process. The trade-off is that the receiver complexity is raised by using more analog circuits. In addition, the use of AC coupling places a constraint on nulling the DC of the signal spectrum and hence trades off the spectral efficiency. Except waveform shaping or frontend processing, the DC offset may also be taken care of in the baseband signal processing. Most of approaches in this stream are to estimate and then compensate for the DC offset by virtue of adaptive filtering or least square approaches [10]-[14]. In particular, the training-sequence-aided (TSA) approaches in [13]-[14], which estimate DC offset and channel response in the prefix training sequence and then make data detection with the aid of the estimates in the suffix data sequence, are shown to provide satisfactory error performance in various fading environments. Since training sequences are used, loss in both spectral and power efficiencies results. To reduce the loss, each suffix data sequence has to be made as long as possible. This requires that both DC offset and channel response remain constant over a long processing window.

This paper is concerned with a new baseband processing approach that can be used for broader applications where both DC offset and channel response are only required to remain constant over a short processing window. The proposed approach manipulates the signal block by block and works under the additive white Gaussian noise (AWGN) channel and the environment that both DC offset and channel response remain constant over each block.[1] Section II models the received baseband signal in conjunction with DC offset. By processing

[1]Although not treated explicitly, the proposed detection approaches can be used for block fading channels, as long as DC offset and channel response remain constant over a block.

the received signal block by block, a generalized transform matrix is used to remove the offset completely and, thereby, releases the succeeding data detection from the threat of DC offset. Applying the principle of the generalized likelihood ratio test [15] on the transform output, a noncoherent block detection scheme is designed in conjunction with a new block modulation format to perform data detection and channel estimation jointly. The design process consists of two steps, namely implicit channel estimation (Section III) and noncoherent data detection using simultaneous channel estimation (Section IV). Assuming that the signaling block is known *a priori* to the receiver, the maximum-likelihood (ML) scheme operating on the transform output is developed in Section III to provide the joint estimates of channel amplitude and phase. It is analytically found that the estimation performance prevalence can be achieved when both real and imaginary parts of the signaling block belong to a vector space $\mathcal{S}_M$ expanded by M basis vectors obtained from the transform matrix. Taking advantage of this property, a new digital phasor block modulation format is proposed in Section IV to allow for noncoherent data detection using simultaneous channel estimation. The basic idea is to modulate at the transmitter multiple L-ary constant-modulus data phasors with the basis vectors in $\mathcal{S}_M$ and demodulate at the receiver the block-modulated signal by the proposed block detection scheme. In the block detection scheme, the ML estimates of channel amplitude and phase are made implicitly for each candidate signaling block and used as the *tentative* channel values to assist the data detection. Since the real and imaginary parts of all the signaling blocks belong to $\mathcal{S}_M$, the amplitude and phase estimates corresponding to the correct signaling block are optimized and thereby benefit the detection performance. The proposed block detection scheme is analyzed in terms of bit error probability (BEP). An iterative detection scheme is also modified from the block detection scheme to simplify the realization complexity. Both block and iterative schemes are shown to outperform the conventional TSA scheme in [13]. Section V concludes the paper.

II. DC Offset Elimination

Consider the baseband communication system where the filtering is evenly split between transmitter and receiver so that the overall channel satisfies the first Nyquist condition. Filtering the received waveform through a matched filter and sampling at the right time instants yields the received block $\mathbf{r} \triangleq [r_1, r_2, ..., r_K]^t$ with[2]

$$r_k = \rho a_k \exp\{j\theta\} + \xi_k + n_k. \tag{1}$$

Here, ρ and θ are the channel amplitude and phase which are assumed to be constant but otherwise unknown, with $\rho \geq 0$ and $\theta \in [-\pi, \pi)$. The samples ξ_k's model the DC

offset emerging from oscillator leakage, tone interference, tone jamming, pilot tone, and/or transistor mismatch. The offset is assumed to remain constant over K consecutively received samples so that $\xi_k = \xi$. n_k's describe the noise samples which are independent and identically distributed, circularly symmetric complex Gaussian random variables with mean zero and variance $E\{|n_k|^2\} = 1$. The K-tuple $\mathbf{a} \triangleq [a_1, a_2, ..., a_K]^t$ represents the signaling block of complex symbols that are sent by the transmitter. With $W \triangleq \mathbf{a}^h \mathbf{a}$ denoting the total energy of $\mathbf{a}$, the ratio of the average signal energy per received complex symbol to noise density level (i.e., average symbol SNR) is given by $\gamma = W \rho^2 / K$.

To remove ξ, a block transform $\mathbf{G}$ is applied to convert $\mathbf{r}$ to a block $\mathbf{z} \triangleq [z_1, z_2, ..., z_M]^t$ by $\mathbf{z} = \mathbf{G} \mathbf{r}$. Here, the transform is characterized by a real-valued matrix $\mathbf{G} \triangleq [g_{m,k}]_{M \times K}$ satisfying the constraint that all rows sum to zero, i.e.,

$$\sum_{k=1}^{K} g_{m,k} = 0 \tag{2}$$

for all m.[3] Equivalently, the transform output can be written as

$$\mathbf{z} = \rho \exp\{j\theta\} \mathbf{G} \mathbf{a} + \mathbf{u} \tag{3}$$

where the resultant $M \times 1$ noise vector $\mathbf{u}$ is a complex-valued Gaussian random vector [16] which has zero mean vector and covariance matrices $E\{\mathbf{u}\mathbf{u}^t\} = \mathbf{O}_{M \times M}$ and $E\{\mathbf{u}\mathbf{u}^h\} = \mathbf{\Lambda}$ with $\mathbf{O}_{M \times M}$ the $M \times M$ null matrix and $\mathbf{\Lambda} \triangleq \mathbf{G}\mathbf{G}^t$ a Hermitian symmetric $M \times M$ covariance matrix. It is further constrained that $\mathbf{G}\mathbf{a}$ is not a null vector so that the necessary signaling information is preserved with $\mathbf{z}$. Evidently, the succeeding processes operating on $\mathbf{z}$ can be conducted without being affected by DC offset.

In the paper, $\mathbf{\Lambda}$ is assumed to be positive definite and have positive eigenvalues $\lambda_1, \lambda_2, ..., \lambda_M$ in decreasing order (i.e., $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_M > 0$). This is possible only when $M < K$ because the rank of $\mathbf{G}$ is less than K due to the imposed constraint (2). We thus consider $M < K$ throughout. Within this assumption, $\mathbf{\Lambda}$, its inverse $\mathbf{\Lambda}^{-1}$, and $\mathbf{G}$ can be expressed by taking advantage of eigenvalue and singular value decompositions [16] as

$$\mathbf{\Lambda} = \mathbf{E}\mathbf{D}\mathbf{E}^t, \quad \mathbf{\Lambda}^{-1} = \mathbf{E}\mathbf{D}^{-1}\mathbf{E}^t, \quad \mathbf{G} = \mathbf{E}\mathbf{F}\mathbf{Q}^t \tag{4}$$

where $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_M]$ is the $M \times M$ real-valued unitary matrix with the mth column $\mathbf{e}_m$ being the eigenvector of $\mathbf{\Lambda}$ corresponding to eigenvalue λ_m, $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_K]$ is a $K \times K$ real-valued unitary matrix, $\mathbf{D} = diag(\lambda_1, \lambda_2, ..., \lambda_M)$, $\mathbf{D}^{-1} = diag(1/\lambda_1, 1/\lambda_2, ..., 1/\lambda_M)$, and $\mathbf{F}$ is an $M \times K$ rectangular matrix given by $\mathbf{F} = [diag(\sqrt{\lambda_1}, \sqrt{\lambda_2}, ..., \sqrt{\lambda_M}), \mathbf{O}_{M \times (K-M)}]$. We shall proceed our discussion hereinafter using these notations.

III. Implicit Channel Estimation

Operating on $\mathbf{z}$, we are interested in the design of the signaling block $\mathbf{a}$ which allows for noncoherent data detection

---

[2]In the paper, the boldface lower letter denotes a column vector and the boldface upper letter a matrix. Superscripts t, $*$ and h denote the transpose, the complex conjugate and the conjugate transpose, respectively. $\|\mathbf{x}\|^2$ denotes the squared norm of vector $\mathbf{x}$, $|\mathbf{A}|$ the determinant of matrix $\mathbf{A}$, and $\mathbf{A}^{-1}$ the inverse of matrix $\mathbf{A}$. For convenience, we also use $diag(x_1, x_2, ..., x_M)$ to represent an $M \times M$ diagonal matrix with x_m denoting the mth diagonal entry, and $\mathbf{A} = [A_{m,k}]_{M \times K}$ an $M \times K$ matrix with $A_{m,k}$ denoting the entry corresponding to the mth row and the kth column.

[3]The constraint (2) is exactly the constraint on the training sequence that was found in [13] and [14] to enable a complete removal of DC offset and, moreover, optimize the channel estimation performance results of the conventional TSA least-square approaches for jointly estimating DC offset and channel response.

performing simultaneous implicit channel estimation. To begin with, we consider here the ML estimation of ρ and θ based on $\mathbf{z}$, assuming that $\mathbf{a}$ is known *a priori*. The resultant estimates will be then used implicitly in the noncoherent detection scheme designed in next section.[4]

Given ρ, θ and $\mathbf{a}$, $\mathbf{z}$ is a complex Gaussian random vector [16] with the probability density

$$
p(\mathbf{z}|\rho,\theta,\mathbf{a}) = \pi^{-M}|\mathbf{\Lambda}|^{-1}\exp\{-(\mathbf{z}-\rho e^{j\theta}\mathbf{Ga})^h\mathbf{\Lambda}^{-1}
$$
$$
\cdot(\mathbf{z}-\rho e^{j\theta}\mathbf{Ga})\}. \tag{5}
$$

The ML estimates $\widehat{\rho}$ and $\widehat{\theta}$ are obtained by finding the arguments of ρ and θ that maximize $p(\mathbf{z}|\rho,\theta,\mathbf{a})$ jointly, as

$$
\widehat{\rho} = \frac{|V|}{U} \quad and \quad \widehat{\theta} = -\tan^{-1}\left(\frac{Im\{V\}}{Re\{V\}}\right) \tag{6}
$$

for $\widehat{\theta} \in [-\pi,\pi)$ where U and V are defined by

$$
U \triangleq (\mathbf{Ga})^h\mathbf{\Lambda}^{-1}(\mathbf{Ga}) \quad and \quad V \triangleq (\mathbf{Ga})^t\mathbf{\Lambda}^{-1}\mathbf{z}^*. \tag{7}
$$

Note that U is strictly positive since $\mathbf{Ga}$ is not a null vector. As implied in (6), the performance of $\widehat{\rho}$ and $\widehat{\theta}$ depends on the formats for $\mathbf{a}$ and $\mathbf{G}$ and deserves further investigation.

A) Estimation Performance Evaluation: The estimation performance is characterized in terms of the estimation mean and the mean square estimation error (MSE). Because V is a linear transform of the complex Gaussian vector $\mathbf{z}$, it is a complex Gaussian vector. It thus follows that the joint probability density of $\widehat{\rho}$ and $\widehat{\theta}$ is given by

$$
p(\widehat{\rho},\widehat{\theta}|\rho,\theta,\mathbf{a}) = \frac{1}{\pi}U\widehat{\rho}\exp\{-U[\widehat{\rho}^2+\rho^2-2\widehat{\rho}\rho\cos(\widehat{\theta}-\theta)]\} \tag{8}
$$

for $\widehat{\rho} \geq 0$ and $\widehat{\theta} \in [-\pi,\pi)$. This density is exactly the joint probability density of the amplitude and phase (in the polar form) of a circularly symmetric complex-valued Gaussian random variable with mean $\rho e^{-j\theta}$ and variance $1/U$ [17, eq. 7.2.26]. Quoting [17, eqs. 7.2.27 and 7.2.33] and [18, eq. 2-1-146], the first two moments of $\widehat{\theta}$ and $\widehat{\rho}$ are given by $E\{\widehat{\theta}\} = \theta$,

$$
Var\{\widehat{\theta}\} = \int_{-\pi}^{\pi}x^2\left\{\frac{1}{2\pi}e^{-\rho^2U}+\sqrt{\frac{\rho^2U}{\pi}}e^{-\rho^2U\sin^2x}\cos x\right.
$$
$$
\left.\cdot\left[1-Q\left(\sqrt{2\rho^2U}\cos x\right)\right]\right\}dx \tag{9}
$$

$$
E\{\widehat{\rho}\} = \frac{1}{2}\sqrt{\frac{\pi}{U}}e^{-\frac{1}{2}\rho^2U}\left[(1+\rho^2U)I_0(\frac{1}{2}\rho^2U)\right.
$$
$$
\left.+(\rho^2U)I_1(\frac{1}{2}\rho^2U)\right] \tag{10}
$$

$$
E\{\widehat{\rho}^2\} = \rho^2 + \frac{1}{U} \tag{11}
$$

where $Q(x) \triangleq (1/\sqrt{2\pi})\int_x^{\infty}\exp\{-z^2/2\}dz$ is the Gaussian tail integral and $I_n(x)$ is the nth-order modified Bessel function of the first kind [18, eq. 2-1-120].

---

[4]The proposed scheme can also be used to provide the ML estimates of ρ and θ based on $\mathbf{z}$, when $\mathbf{a}$ bears the training symbols known to the receiver. Such data-aided estimates are useful for coherently demodulating the phase-shift-keyed or quadrature-amplitude-modulated signals based on $\mathbf{z}$, if DC offset and channel response remain constant over a long processing window. This content is, however, beyond the scope of the paper.
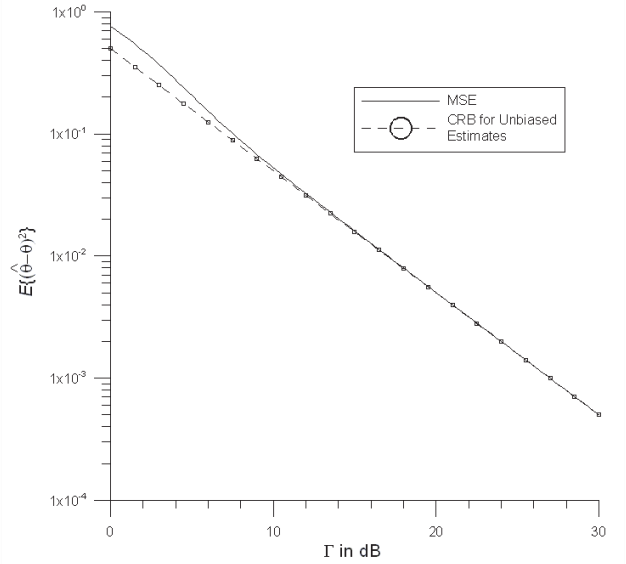

Fig. 1. Mean squared error of phase estimate.

Since $\widehat{\theta}$ is unbiased, its MSE is given by $MSE_{\widehat{\theta}} \triangleq E\{(\widehat{\theta}-\theta)^2\} = Var\{\widehat{\theta}\}$. The Cramér-Rao bound (CRB) that provides a fundamental lower bound (for an unbiased estimate) to $MSE_{\widehat{\theta}}$ is given by $MSE_{\widehat{\theta}} \geq CRB_{\widehat{\theta}}$ with[5]

$$
CRB_{\widehat{\theta}} \triangleq \left[-E\left\{\frac{\partial^2}{\partial\theta^2}\ln p(\mathbf{z}|\rho,\theta,\mathbf{a})\right\}\right]^{-1} = \frac{1}{2\rho^2U} \tag{12}
$$

From (9) and (12), both $MSE_{\widehat{\theta}}$ and $CRB_{\widehat{\theta}}$ are found to depend only on the sole measure

$$
\Gamma \triangleq \rho^2U = K\gamma U/W.
$$

Fig. 1 gives $MSE_{\widehat{\theta}}$ and $CRB_{\widehat{\theta}}$ with respect to Γ. As shown, $MSE_{\widehat{\theta}}$ decreases monotonically when Γ increases and approaches to $CRB_{\widehat{\theta}}$ when Γ is above 10 dB. Since $\widehat{\theta}$ is unbiased and $MSE_{\widehat{\theta}}$ approaches to zero as Γ goes to infinity, $\widehat{\theta}$ is asymptotically consistent [15] with Γ.

It is seen from (10) and (11) that the performance of $\widehat{\rho}$ depends strongly on the true amplitude value ρ. To compare the relative estimation accuracies for different amplitude levels, we consider here the performance characteristics of $\widehat{\rho}$ normalized by ρ. Specifically, the normalized mean $E\{\widehat{\rho}\}/\rho$ and the normalized MSE $MSE_{\widehat{\rho}}/\rho^2 = E\{(\widehat{\rho}-\rho)^2\}/\rho^2 = E\{\widehat{\rho}^2\}/\rho^2 - 2E\{\widehat{\rho}\}/\rho+1$ are taken into account. Investigating (10) and (11), both $E\{\widehat{\rho}\}/\rho$ and $MSE_{\widehat{\rho}}/\rho^2$ are found to depend only on Γ, as is the case with $MSE_{\widehat{\theta}}$ and $CRB_{\widehat{\theta}}$. Fig. 2 shows the results of $E\{\widehat{\rho}\}/\rho$ with respect to Γ. Evidently, $\widehat{\rho}$ is biased and it becomes asymptotically unbiased when Γ gets larger. If $\widehat{\rho}$ was unbiased, $MSE_{\widehat{\rho}}/\rho^2$ could be lower bounded by the CRB (for an unbiased estimate) as $MSE_{\widehat{\rho}}/\rho^2 \geq CRB_{\widehat{\rho}}/\rho^2$ with

$$
CRB_{\widehat{\rho}}/\rho^2 \triangleq \left[-\rho^2E\left\{\frac{\partial^2}{\partial\rho^2}\ln p(\mathbf{z}|\rho,\theta,\mathbf{a})\right\}\right]^{-1} = \frac{1}{2\rho^2U} \tag{13}
$$

---

[5]The Fisher information matrix of the estimation problem with $p(\mathbf{z}|\rho,\theta,\mathbf{a})$ is diagonal so that the CRB's for unbiased estimates of ρ and θ are separable [19], [20].
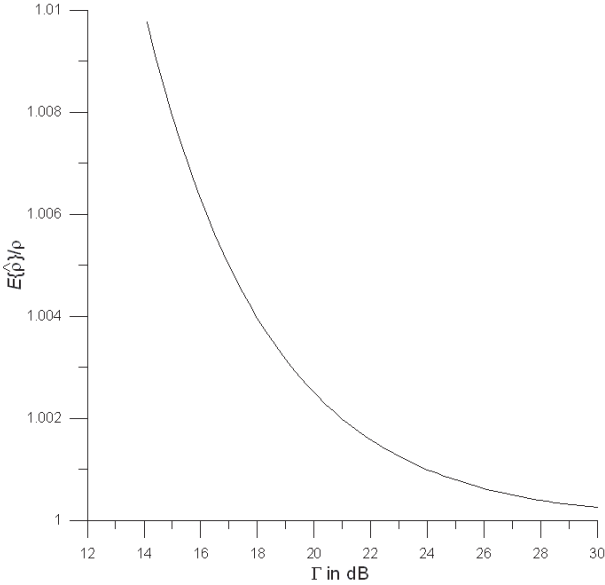
Fig. 2. Normalized mean of amplitude estimate.

Fig. 3. Normalized mean squared error of amplitude estimate.

which also depends only on Γ. Since $\widehat{\rho}$ is asymptotically unbiased, $CRB_{\widehat{\rho}}/\rho^2$ provides a good approximation of the lower bound for $MSE_{\widehat{\rho}}/\rho^2$ when Γ is large. Fig. 3 shows $MSE_{\widehat{\rho}}/\rho^2$ and $CRB_{\widehat{\rho}}/\rho^2$ with respect to Γ. $MSE_{\widehat{\rho}}/\rho^2$ is found to decrease monotonically when Γ increases and approaches to $CRB_{\widehat{\rho}}/\rho^2$ when Γ is above 10 dB. Because $E\{\widehat{\rho}\}/\rho$ approaches to one and $MSE_{\widehat{\rho}}/\rho^2$ approaches to zero as Γ goes to infinity, $\widehat{\rho}$ is asymptotically consistent with Γ.

It is interesting to compare the performance of $\widehat{\rho}$ and $\widehat{\theta}$ with the CRB performance for unbiased amplitude and phase estimates, denoted by $\widehat{\rho}_u$ and $\widehat{\theta}_u$, of an unmodulated signal received in the ideal AWGN channel without DC offset. Because neither data symbols nor DC offset is involved in the estimation process, the CRB's for $\widehat{\rho}_u$ and $\widehat{\theta}_u$ are regarded as the benchmark results of the MSE's that can be achieved by $\widehat{\rho}$ and $\widehat{\theta}$. Based on the signal model in (1) with signaling symbols and DC offset removed, i.e., $a_k = 1$ and $\xi_k = 0$ for all k, the CRB's for $\widehat{\rho}_u$ and $\widehat{\theta}_u$ are obtained as

$$CRB_{\widehat{\rho}_u}/\rho^2 = CRB_{\widehat{\theta}_u} = \frac{1}{2K\gamma} \tag{14}$$

which depend only on the average block SNR $K\gamma$. Note that the value $1/(2K\gamma)$ differs from the common CRB value $1/(2\Gamma)$ in (12) and (13) exactly by the factor W/U which depends on $\mathbf{a}$ and $\mathbf{G}$. Therefore, when Γ is large, the asymptotic MSE characteristics of $\widehat{\rho}/\rho$ and $\widehat{\theta}$ approach to (W/U)-fold the common CRB of $\widehat{\rho}_u/\rho$ and $\widehat{\theta}_u$. It is noted that the factor W/U is not smaller than one for all permissible forms of $\mathbf{a}$ and $\mathbf{G}$. When $\mathbf{a}$ and $\mathbf{G}$ are optimized, W/U is shown in the following to reach its minimum value and thereby the benchmark CRB results in (14) are asymptotically achieved by $\widehat{\rho}/\rho$ and $\widehat{\theta}$.

B) Design of Signaling Block $\mathbf{a}$ *and Transform* $\mathbf{G}$*:* As shown previously, the performance characteristics of $\widehat{\theta}$ and $\widehat{\rho}/\rho$ depend only on the sole measure $\Gamma = (K\gamma)\cdot(U/W)$. The larger the value of Γ, the better performance both estimates provide. Therefore, for a fixed block SNR $K\gamma$, the estimation performances improve proportionally to U/W. Since the ratio
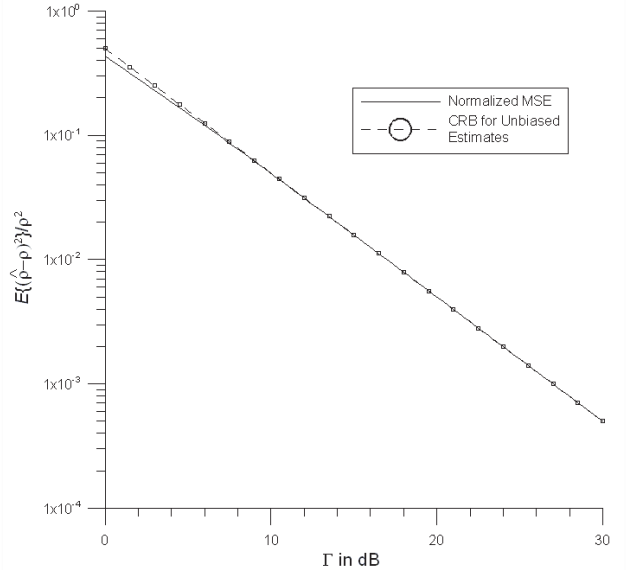
U/W depends only on $\mathbf{a}$ and $\mathbf{G}$, both $\mathbf{a}$ and $\mathbf{G}$ can be optimized in the sense that U/W is maximized. Such an optimization is given below.

Rewrite U and W in terms of the rectangular form $\mathbf{a} = Re\{\mathbf{a}\} + jIm\{\mathbf{a}\}$ as

$$
\begin{aligned}
U &= Re\{\mathbf{a}\}^t\mathbf{G}^t\mathbf{\Lambda}^{-1}\mathbf{G}Re\{\mathbf{a}\} \\
&\quad +Im\{\mathbf{a}\}^t\mathbf{G}^t\mathbf{\Lambda}^{-1}\mathbf{G}Im\{\mathbf{a}\} \tag{15} \\
W &= Re\{\mathbf{a}\}^tRe\{\mathbf{a}\} + Im\{\mathbf{a}\}^tIm\{\mathbf{a}\}. \tag{16}
\end{aligned}
$$

Here, $\mathbf{G}^t\mathbf{\Lambda}^{-1}\mathbf{G}$ is a $K \times K$ symmetric real matrix and can be simplified from (4) as

$$\mathbf{G}^t\mathbf{\Lambda}^{-1}\mathbf{G} = \mathbf{Q}\mathbf{F}^t\mathbf{D}^{-1}\mathbf{F}\mathbf{Q}^t = \mathbf{Q}\mathbf{I}_M\mathbf{Q}^t \tag{17}$$

where $\mathbf{I}_M$ is a $K \times K$ diagonal matrix given by $\mathbf{I}_M = diag(1, 1, ..., 1, 0, 0, ..., 0)$ with the preceding M diagonal elements being all ones and the succeeding $K - M$ diagonal elements all zeros. It follows from (17) that the symmetric matrix $\mathbf{G}^t\mathbf{\Lambda}^{-1}\mathbf{G}$ is positive semidefinite with two distinct eigenvalues given by one, which has multiplicity M and eigenvectors $\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_M$, and zero, which has multiplicity $K - M$ and eigenvectors $\mathbf{q}_{M+1}, \mathbf{q}_{M+2}, ..., \mathbf{q}_K$. Quoting the theorem of Rayleigh quotient [16, Theorem A.2.4], we have the bounds

$$
\begin{aligned}
0 &\leq Re\{\mathbf{a}\}^t\mathbf{G}^t\mathbf{\Lambda}^{-1}\mathbf{G}Re\{\mathbf{a}\} \leq Re\{\mathbf{a}\}^tRe\{\mathbf{a}\} \tag{18} \\
0 &\leq Im\{\mathbf{a}\}^t\mathbf{G}^t\mathbf{\Lambda}^{-1}\mathbf{G}Im\{\mathbf{a}\} \leq Im\{\mathbf{a}\}^tIm\{\mathbf{a}\} \tag{19}
\end{aligned}
$$

where the upperbounds are achieved when $Re\{\mathbf{a}\}$ and $Im\{\mathbf{a}\}$ are linear combinations of $\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_M$, and the lowerbounds are achieved when $Re\{\mathbf{a}\}$ and $Im\{\mathbf{a}\}$ are linear combinations of $\mathbf{q}_{M+1}, \mathbf{q}_{M+2}, ..., \mathbf{q}_K$. Putting these results into (15) and (16) shows that the maximum value for U/W is one and it occurs when both $Re\{\mathbf{a}\}$ and $Im\{\mathbf{a}\}$ belong to the M-dimensional space $\mathcal{S}_M$ expanded by the basis vectors $\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_M$. Let the $K \times M$ matrix $\mathbf{Q}_M$ be the matrix composed by the first M columns of $\mathbf{Q}$, namely $\mathbf{Q}_M \triangleq$

$[\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_M]$. It is readily obtained from the decomposition $\mathbf{G} = \mathbf{EFQ}^t$ that

$$\mathbf{Q}_M = \mathbf{G}^t \mathbf{E} \mathbf{D}^{-1/2} \qquad (20)$$

with $\mathbf{D}^{-1/2} = diag(\lambda_1^{-1/2}, \lambda_2^{-1/2}, ..., \lambda_M^{-1/2})$. As a result, the optimized performance characteristics of $\widehat{\theta}$ and $\widehat{\rho}/\rho$ are found to depend only on $\Gamma_{\max} = K\gamma$ when the signaling block $\mathbf{a}$ is limited to the form

$$\mathbf{a} = \mathbf{Q}_M \mathbf{d} = \mathbf{G}^t \mathbf{E} \mathbf{D}^{-1/2} \mathbf{d} \qquad (21)$$

where $\mathbf{d} \triangleq [d_1, d_2, ..., d_M]^t$ is a complex vector with any nonzero energy $\mathbf{d}^h \mathbf{d} = W$, as long as the linear transform $\mathbf{G}$ satisfies the offset-removal constraint (2) and the constraint that $\Lambda = \mathbf{GG}^t$ is positive definite. Combining this with the results in Figs. 2 and 3 reveals that the asymptotic MSE characteristics of $\widehat{\rho}/\rho$ and $\widehat{\theta}$ corresponding to large γ approach to the benchmark CRB's (14) when $\mathbf{a}$ is optimized for a prescribed transform $\mathbf{G}$.

To realize the estimation mechanism, the transform $\mathbf{G}$ needs to be properly designed. In addition to the offset-removal constraint (2) and the constraint that $\mathbf{GG}^t$ is positive definite, the third constraint that none of the columns in $\mathbf{G}$ is a null vector is also pragmatically imposed on the permissible forms of $\mathbf{G}$ so that all received samples in $\mathbf{r}$ are processed after the offset-removal transform. In general, there are infinitely many forms of $\mathbf{G}$ satisfying these constraints. Two examples with $M = K - 1$ are given here. Table I lists the transform $\mathbf{G}$ and the associated matrices Λ, Λ^{-1}, $\mathbf{E}$, $\mathbf{D}$ and $\mathbf{Q}_M$ for both examples. These associated matrices were obtained after quoting the properties of Green, Jacobi and inter-class correlation matrices in [21] and are useful in realizing the signaling block $\mathbf{a}$ in (21) and the estimates $\widehat{\theta}$ and $\widehat{\rho}$ in (6). Specifically, $\mathbf{z}$ and $\mathbf{r}$ are related by

$$z_m = r_{m+1} - r_m, \quad m = 1, 2, ..., M \qquad (22)$$

in *Example One*, and

$$z_m = r_{m+1} - r_1, \quad m = 1, 2, ..., M \qquad (23)$$

in *Example Two*. Because the differences of the received samples in the same block are taken, both transforms (22) and (23) are obviously capable of completely eliminating the DC offset when the offset remains constant over K consecutive samples. Since the differences of two consecutive samples are used, transform (22) can be further applied to other scenarios where the DC offset remains constant over an interval of two or slightly more samples and varies gradually between intervals.

IV. NONCOHERENT BLOCK DETECTION USING SIMULTANEOUS CHANNEL ESTIMATION

In the section, a noncoherent block detection scheme operating on $\mathbf{z}$ is designed in conjunction with a digital phasor block modulation format to perform data detection and implicit channel estimation jointly. The modulation format (21) is adopted here to optimize the performance of the implicit channel estimation corresponding to a correct signaling block. Specifically, with a prescribed $\mathbf{G}$, both real and imaginary parts of the signaling block $\mathbf{a}$ belong to $\mathcal{S}_M$ and are jointly

modulated in the form $\mathbf{a} = \mathbf{Q}_M \phi$ where $\phi \triangleq [\phi_1, \phi_2, ..., \phi_M]^t$ is the data phasor block to be detected, with $\phi_m \in \mathcal{A} \triangleq \{\exp\{j2\pi l/L\}\}_{l=0}^{L-1}$. By signaling so, the received amplitude ρ is related to the average symbol SNR γ by $\rho^2 = K\gamma/M$. Further, it is assumed that $K > M \geq 2$ and L is an integer power of two.

Operating on $\mathbf{z}$, the decision $\widehat{\phi}$ is determined by the argument of ϕ which maximizes $\max_{\rho,\theta} p(\mathbf{z}|\rho, \theta, \mathbf{a})$. Note that this rule is an application of the generalized likelihood ratio test argument in [15]. Structurally, the rule can be regarded as the ML decision of ϕ using the *tentative* ML estimates of ρ and θ which are made for all candidate signaling blocks. The tentative ML estimates of ρ and θ have exactly the same form as in (6). Since the real and imaginary parts of all candidate signaling blocks belong to $\mathcal{S}_M$, the tentative estimates of ρ and θ corresponding to the correct signaling block can provide the best estimation in forming the test metric in the case, thereby benefiting the detection performance. This justifies the use of modulation format $\mathbf{Q}_M \phi$ for constructing $\mathbf{a}$.

Because $-\ln(x)$ decreases with increasing x, the decision rule is tantamount to minimizing $\min_{\rho,\theta}\{-\ln p(\mathbf{z}|\rho, \theta, \mathbf{a})\}$. Discarding terms that are irrelevant to the decision, the rule simplifies to

$$\widehat{\phi} = \arg \max_{\phi \in \mathcal{B}} |V(\phi)|^2. \qquad (24)$$

Here, $\mathcal{B}$ is the family of all candidate data phasor blocks. The metric $V(\phi)$ is as defined by (7) and can be simplified by using (20) to

$$V(\phi) = (\mathbf{GQ}_M \phi)^t \Lambda^{-1} \mathbf{z}^* \qquad (25)$$
$$= \phi^t \mathbf{D}^{-1/2} \mathbf{E}^t \mathbf{z}^*. \qquad (26)$$

Note that the rule (24) can not differentiate between ϕ and $\chi\phi$ with χ a complex scalar. Thus, if both ϕ and $\chi\phi$ belong to $\mathcal{B}$, the rule (24) will result in decision ambiguity. To avoid the ambiguity, we fix the first phasor ϕ_1 in all data phasor blocks to a constant phasor. Without loss of generality, the constant phasor is set to one since it does not influence the decision operation nor the error performance. In the case, the family $\mathcal{B}$ consists of L^{M-1} distinguishable data phasor blocks of the form $[1, \phi_2, ..., \phi_M]^t$ with $\phi_m \in \mathcal{A}$ for $m \geq 2$. By this setting, only $(M - 1)\log_2 L$ bits of information is conveyed in every block of K channel symbol uses. Thus, the average bit SNR γ_b that is commonly used as the SNR measure when comparing among different detection schemes is related to the average symbol SNR γ by $\gamma_b = K\gamma/[(M - 1)\log_2 L]$ for a generic form of $\mathbf{G}$. For example, if the transforms for $\mathbf{G}$ in Table I are used, κ bits of information are conveyed every block and $\gamma_b = K\gamma/\kappa$, with $\kappa \triangleq (K - 2)\log_2 L$. If DC offset and channel response remain constant over an interval of K received samples and vary arbitrarily between intervals,[6] then κ/K bits/symbol is the best transmission rate that can be achieved. When K is fixed, $M = K - 1$ is the best choice for setting M in view of transmission throughput efficiency. In the case, the maximum rate of $\log_2 L$ bits/symbol can be asymptotically achieved if a large K is permitted of use.

---

[6]This may occur in slow frequency hopping or time division multiple access applications.

Table I. Two design examples for the transform $\mathbf{G}$ with $M = K - 1$ and its associated matrices.

	Example One	Example Two
$\mathbf{G} = [g_{m,k}]_{(K-1)\times K}$	$g_{m,k} = \begin{cases} -1, & m = k \\ 1, & m = k-1 \\ 0, & otherwise \end{cases}$	$g_{m,k} = \begin{cases} -1, & k = 1 \\ 1, & k = m+1 \\ 0, & otherwise \end{cases}$
$\mathbf{\Lambda} = [\Lambda_{m,k}]_{(K-1)\times(K-1)}$	$\Lambda_{m,k} = \begin{cases} 2, & m = k \\ -1, & \|m-k\| = 1 \\ 0, & otherwise \end{cases}$	$\Lambda_{m,k} = \begin{cases} 2, & m = k \\ 1, & otherwise \end{cases}$
$\mathbf{\Lambda}^{-1} = [(\Lambda^{-1})_{m,k}]_{(K-1)\times(K-1)}$	$(\Lambda^{-1})_{m,k} = \begin{cases} m(1-\frac{m}{K}), & m = k \\ k(1-\frac{m}{K}), & m > k \\ m(1-\frac{k}{K}), & m < k \end{cases}$	$(\Lambda^{-1})_{m,k} = \begin{cases} 1 - \frac{1}{K}, & m = k \\ \frac{-1}{K}, & otherwise \end{cases}$
$\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_{K-1}]$ $\mathbf{e}_m = [e_{m,1}, e_{m,2}, ..., e_{m,K-1}]^t$	$e_{m,k} = \sqrt{\frac{2}{K}} \sin\left(\frac{mk\pi}{K}\right)$	$e_{m,k} = \begin{cases} \frac{1}{\sqrt{K-1}}, & m = 1 \\ 0, & m > 1 \text{ and } m < k \\ -\sqrt{1-\frac{1}{m}}, & m > 1 \text{ and } m = k \\ \frac{1}{\sqrt{m(m-1)}}, & m > 1 \text{ and } m > k \end{cases}$
$\mathbf{D} = diag(\lambda_1, \lambda_2, ..., \lambda_{K-1})$	$\lambda_m = 4\sin^2\left(\frac{m\pi}{2K}\right)$	$\lambda_m = \begin{cases} K, & m = 1 \\ 1, & otherwise \end{cases}$
$\mathbf{Q}_{K-1} = [\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_{K-1}]$ $\mathbf{q}_m = [q_{m,1}, q_{m,2}, ..., q_{m,K}]^t$	$q_{m,k} = \begin{cases} -\sqrt{\frac{2}{K}}\cos\left(\frac{m\pi}{2K}\right), & k = 1 \\ -\sqrt{\frac{2}{K}}\cos\left(\frac{m(2k-1)\pi}{2K}\right), & 1 < k < K \\ (-1)^{m+1}\sqrt{\frac{2}{K}}\cos\left(\frac{m\pi}{2K}\right), & k = K \end{cases}$	$q_{1,k} = \begin{cases} -\sqrt{1-\frac{1}{K}}, & k = 1 \\ \frac{1}{\sqrt{K(K-1)}}, & otherwise \end{cases}$ $q_{m,k} = \begin{cases} 0, & k = 1 \\ \frac{1}{\sqrt{m(m-1)}}, & 1 < k < m+1 \\ -\sqrt{1-\frac{1}{m}}, & k = m+1 \\ 0, & k > m+1 \\ & for\ m > 1 \end{cases}$

A) BEP Analysis on Block Detection Scheme: The exact BEP expression for rule (24) with general values of L and M is difficult, if not impossible, to obtain. An upper bound is, however, achievable based on the same union bound argument of [22, eq. 7.67] and is given here. Let ϕ and $\widehat{\phi}$ be the transmitted and detected phasor blocks, respectively. From the union bound argument, an upper bound to the BEP of rule (24) is given by

$$P_b \leq \sum_{\phi} L^{1-M} \sum_{\widehat{\phi} \neq \phi} \frac{D_H(\phi, \widehat{\phi})}{(M-1)\log_2 L}$$
$$\cdot \Pr\{|V(\widehat{\phi})|^2 > |V(\phi)|^2 | \phi\} \qquad (27)$$

where $D_H(\phi, \widehat{\phi})$ represents the Hamming distance between the binary representations of blocks ϕ and $\widehat{\phi}$, and $\Pr\{|V(\widehat{\phi})|^2 > |V(\phi)|^2 | \phi\}$ denotes the pairwise error probability that $|V(\widehat{\phi})|^2$ is greater than $|V(\phi)|^2$ when ϕ was indeed transmitted. Given that ϕ was transmitted, since $V(\phi)$ and $V(\widehat{\phi})$ are both linear functions of complex Gaussian vector $\mathbf{z}$, they are jointly complex Gaussian random variables. It thus follows that $|V(\widehat{\phi})|^2 - |V(\phi)|^2$, when ϕ was transmitted, is a special case of the general quadratic form of [18, Appendix B, eq. (B-1)] and the pairwise error probability can be derived accordingly as

$$\Pr\{|V(\widehat{\phi})|^2 > |V(\phi)|^2 | \phi\} = \frac{1}{2}[1 - Q(\beta, \alpha) + Q(\alpha, \beta)] \quad (28)$$

where $Q(x, y)$ is Marcum's Q function [18, eq. 2.1.123] and

$$\left\{ \begin{array}{c} \alpha \\ \beta \end{array} \right\} = \sqrt{\frac{[(M-1)\log_2 L]\gamma_b}{2}} \left\{ 1 \left\{ \begin{array}{c} - \\ + \end{array} \right\} \left[1 - \frac{|\delta|^2}{M^2} \right]^{\frac{1}{2}} \right\}^{\frac{1}{2}} . \qquad (29)$$

The parameter δ is related to decision error by

$$\delta \triangleq \phi^h \widehat{\phi} = 1 + \sum_{m=2}^{M} \phi_m^* \widehat{\phi}_m. \qquad (30)$$

Note that the upper bound (27) depends on the parameters M, L and γ_b, but has nothing to do with the block length K.

When the family $\mathcal{B}$ is binary (i.e., $M = L = 2$), the expression in (27) is indeed equal to the exact BEP. We obtain from (30) that $\delta = 0$ for all pairs $(\phi, \widehat{\phi})$ with $\widehat{\phi} \neq \phi$, and thus from (29) that $\alpha = 0$ and $\beta = \sqrt{[(M-1)\log_2 L]\gamma_b}$. Since $Q(\alpha, 0) = 1$ and $Q(0, \beta) = \exp\{-\beta^2/2\}$, we have from (27) and (28) that

$$P_b = \frac{1}{2}\exp\left\{-\frac{1}{2}\gamma_b\right\} \qquad (31)$$

which is inferior exactly by 3.01 dB to the well-known BEP result [18, eq. 5-2-69] of the two-symbol differential detection of the differentially encoded binary phase-shift-keyed (PSK) signal in the ideal AWGN channel (without DC offset).

When $\mathcal{B}$ is nonbinary, the bound (27) can be used to evaluate the BEP characteristics of rule (24) for arbitrary bits-to-blocks code mapping scheme for ϕ. If phasors $\phi_2, ..., \phi_M$ are assumed to be encoded independently by the Gray code mapping scheme of bits to phasors [18], the bound can be further simplified for the large SNR condition, as follows.

It is well known that the probability of the form (28) decreases as α decreases and β increases [22, eqs. 7.68 and 7.69]. Since α and β vary with $|\delta|$ in reverse trends, the pairwise error probability decreases when $|\delta|$ decreases. Thus, as γ_b is large enough, the bound in (27) is asymptotically dominated by those terms with the largest $|\delta|$ value, denoted by $|\delta|_{\max}$. Since $0 < |\delta|_{\max} < M$, $\beta > \alpha$ is guaranteed for those pairs $(\phi, \widehat{\phi})$ yielding $|\delta|_{\max}$. In the case, if we use the
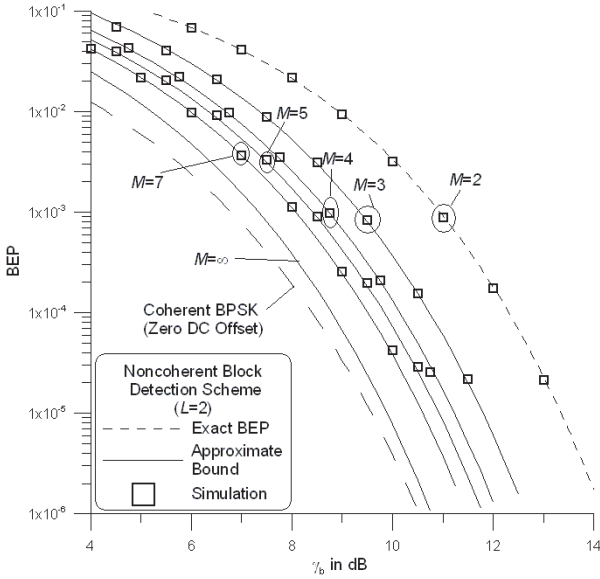
Fig. 4. BEP results for the block detection scheme with $L = 2$.
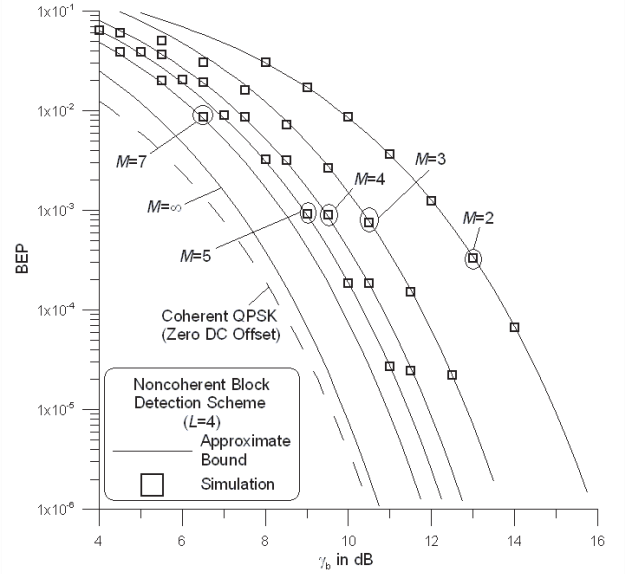
Fig. 5. BEP results for the block detection scheme with $L = 4$.

asymptotic expressions [22, eqs. 7.76 and 7.87]

$$1 - Q(\beta, \alpha) + Q(\alpha, \beta) \tag{32}$$

$$\cong \frac{\exp\left\{-\frac{(\beta-\alpha)^2}{2}\right\}}{\sqrt{2\pi}(\beta-\alpha)}\left(\sqrt{\frac{\beta}{\alpha}} + \sqrt{\frac{\alpha}{\beta}}\right), \beta \gg \alpha \gg 1$$

$$Q(x) \cong \frac{1}{\sqrt{2\pi}x}\exp\left\{-\frac{x^2}{2}\right\}, \qquad x \gg 1 \tag{33}$$

the pairwise error probability (28) corresponding to the pairs $(\phi, \widehat{\phi})$ yielding $|\delta|_{\max}$ can be approximated for a large γ_b by $\mathcal{P}(|\delta|_{\max})$, with $\mathcal{P}(x)$ defined by

$$\mathcal{P}(x) = \sqrt{\frac{M+x}{2x}}Q\left(\sqrt{[(M-1)\log_2 L]\gamma_b(1-\frac{x}{M})}\right). \tag{34}$$

When γ_b is large, $\mathcal{P}(|\delta|_{\max})$ decreases more rapidly with γ_b increased if a larger $M - |\delta|_{\max}$ is used, and *vice versa*. Thus, when γ_b is large, (27) can be asymptotically approximated by

$$P_b \lesssim \frac{D_{H,\max}}{(M-1)L^{M-1}\log_2 L}\mathcal{P}(|\delta|_{\max}) \tag{35}$$

where $D_{H,\max}$ is the accumulated Hamming distance corresponding to the set of the pairs $(\phi, \widehat{\phi})$ yielding $|\delta|_{\max}$. Observing (30), the largest $|\delta|$ value is obviously given by

$$|\delta|_{\max} = |(M-1) + \exp\{j2\pi/L\}|. \tag{36}$$

$|\delta|_{\max}$ occurs in two cases. In the first case, there exists an $l \in \{2, 3, ..., M\}$ so that $\phi_l^* \widehat{\phi}_l \in \{\exp\{j2\pi/L\}, \exp\{-j2\pi/L\}\}$ and $\phi_m = \widehat{\phi}_m$ for all other $m \neq l$. The Hamming distance between any pair $(\phi, \widehat{\phi})$ in the case is equal to one and thus the corresponding accumulated Hamming distance is equivalent to the number of such pairs $(\phi, \widehat{\phi})$, given by $(M-1)L^{M-1}\varpi$ with $\varpi = 1$ if $L = 2$ and $\varpi = 2$ if $L \geq 4$. In the second case, $\phi_m^* \widehat{\phi}_m = \exp\{j2\pi/L\}$ for all $m \in \{2, 3, ..., M\}$ or $\phi_m^* \widehat{\phi}_m = \exp\{-j2\pi/L\}$ for all $m \in \{2, 3, ..., M\}$. In the case, the Hamming distance between any pair $(\phi, \widehat{\phi})$ is equal

to $M - 1$ and there are a total of $L^{M-1}\varpi$ pairs. Thus, the corresponding accumulated Hamming distance is also given by $(M-1)L^{M-1}\varpi$. It should be noted that these two cases are the same when $M = 2$ but are mutually exclusive when $M > 2$. Combining the results in both cases yields

$$D_{H,\max} = (M-1)L^{M-1}\varpi\zeta \tag{37}$$

where $\zeta = 1$ if $M = 2$ and $\zeta = 2$ if $M > 2$. Using (36) and (37) in (35) yields an approximate BEP bound to rule (24). This approximate expression is especially useful when the size of family $\mathcal{B}$ is very large and thenceforth the bound (27) becomes too cumbersome to compute.

When γ_b is large and L is fixed, the approximate BEP bound (35) decreases as M increases. For a fixed block length K, it is thus natural to set M by its largest possible value, i.e., $M = K - 1$, in that both the highest transmission rate κ/K bits/symbol and the prevalence in asymptotic error performance can be simultaneously achieved. This further implies that the transforms with $M = K - 1$, as given in Table I, are prevailing in practical use.

Under the ideal circumstance that the DC offset remains constant for an infinitely long time, the asymptotic BEP bound is obtained by setting $M \to \infty$ in (35) as

$$P_b|_{M\to\infty} \lesssim \frac{2\varpi}{\log_2 L}Q\left(\sqrt{2\gamma_b \log_2 L}\sin(\frac{\pi}{L})\right) \tag{38}$$

which is exactly two-fold the approximate BEP of Gray-labeled LPSK with coherent detection in the absence of DC offset, given by [18, eq. 5-2-62]. This indicates that the proposed block detection scheme can provide the BEP performance approaching that of Gray-labeled LPSK with ideal coherent detection by increasing M.

Figs. 4-6 give the BEP results obtained by use of the exact BEP (31), the approximate bounds (35) and (38), and also from simulation for the proposed block detection scheme operating in the presence of constant DC offset with $L = 2, 4$ and 8, respectively. The BEP results of ideal coherent LPSK

Table II. Complexity comparison between block decision rule and iterative decision rule based on the metric $V(\phi) = \phi^t \mathbf{D}^{-1/2} \mathbf{E}^t \mathbf{z}^*$. All matrices $\phi^t \mathbf{D}^{-1/2} \mathbf{E}^t$ in the block decision rule and the matrix $\mathbf{D}^{-1/2} \mathbf{E}^t$ in the iterative decision rule, that are independent of the received vector $\mathbf{z}$, are assumed to be pre-stored and not counted in complexity evaluation.

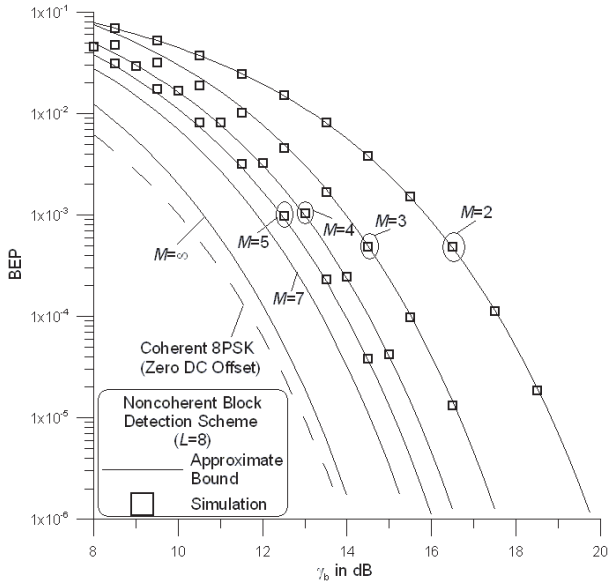|	Number of Real Multiplications Required per KT Seconds	Number of Real Additions Required per KT Seconds	Number of Real Comparisons Required per KT Seconds
Block Decision Rule	$2L^{M-1}(MK+1)$	$L^{M-1}(2MK-2M+1)$	$\leq \binom{L^{M-1}}{2}$
Iterative Decision Rule with I Iterations	$(2MK+6LM-6L)$ $+(I-1)(2LM-2L+4M-4)$	$(2MK-2M+5LM-5L)$ $+(I-L)(M-1)(2M+3L-2)$	$\leq I(M-1)\binom{L}{2}$

Fig. 6. BEP results for the block detection scheme with $L = 8$.

detection are also plotted. All data phasors are assumed to be independently Gray labeled. In the simulation, the transform $\mathbf{G}$ in *Example One* of Table I is considered. As shown in the figures, the simulation results verify clearly the tightness of the approximate bound (35). It is also demonstrated that significant performance improvement can be achieved by increasing M. In particular, extending M from $M = 2$ to $M = 4$ recovers more than half of the SNR loss of the worst-case detection (with $M = 2$) versus the optimum detection (with $M \to \infty$). Moreover, the block detection scheme is shown to provide the BEP performance asymptotically approaching that of the corresponding ideal coherent LPSK detection when M is increased.

B) Iterative Detection Scheme: In implementing the block decision rule (24), the metric $|V(\phi)|^2$ has to be computed for all L^{M-1} possible blocks and thus the required complexity grows exponentially with M. The implementation load may become unwieldy when the block length is large. To reduce complexity, an iterative decision rule is modified from the block decision rule here.

Let $\phi_m^{(J)}$ and $\widehat{\phi}_m^{(J)}$ be the dummy phasor and the tentative decision, respectively, for ϕ_m at the Jth iteration. Also, denote $\phi_m^{(J)}$'s and $\psi_m^{(J)}$'s as the $M \times 1$ dummy phasor vectors and the $M \times 1$ tentative decision vectors, respectively, at the Jth iteration. For $m = 2, 3, ..., M$ and $J = 1, 2, 3, ...$, we define $\phi_m^{(J)} \triangleq [0, ..., 0, \phi_m^{(J)}, 0, ..., 0]^t$ and $\psi_m^{(J+1)} \triangleq [0, ..., 0, \widehat{\phi}_m^{(J)}, 0, ..., 0]^t$

where only the mth element is nonzero in each vector. By default, we set that $\psi_1^{(J)} = [1, 0, ..., 0]^t$ for all J and $\psi_m^{(1)}$ is a null vector for all $m > 1$. Using these notations, the iterative rule can be formulated as follows.

The same test metric $|V(\phi)|^2$ of the block decision rule is adopted for the iterative decision rule. However, unlike the block rule by which a single decision is made on ϕ, the proposed iterative rule makes $I(M - 1)$ tentative decisions before the final decision on ϕ is settled after I iterations, with each iteration consisting of $M - 1$ individual tests for making tentative decisions on $\phi_2, \phi_3, ..., \phi_M$. Specifically, the $M - 1$ tentative decisions at the Jth iteration are given by $\{\widehat{\phi}_2^{(J)}, \widehat{\phi}_3^{(J)}, ..., \widehat{\phi}_M^{(J)}\}$ with[7]

$$\widehat{\phi}_m^{(J)} = \arg \max_{\phi_m^{(J)} \in \mathcal{A}} \left| V\left(\phi_m^{(J)} + \sum_{k \neq m} \psi_k^{(J)}\right) \right|^2 \quad (39)$$

$$= \arg \max_{\phi_m^{(J)} \in \mathcal{A}} \left| V\left(\phi_m^{(J)}\right) + \sum_{k \neq m} V\left(\psi_k^{(J)}\right) \right|^2 \quad (40)$$

for $J = 1, 2, ..., I$, and the final decision on ϕ is then made at the end of the Ith iteration as $\widehat{\phi}_m = \widehat{\phi}_m^{(I)}$ for $m = 2, 3, ..., M$. In (39), the argument in $V(\cdot)$ represents the dummy phasor vector for testing ϕ_m, given by $[1, \widehat{\phi}_2^{(J-1)}, ..., \widehat{\phi}_{m-1}^{(J-1)}, \phi_m^{(J)}, \widehat{\phi}_{m+1}^{(J-1)}, ..., \widehat{\phi}_M^{(J-1)}]^t$, where $M - 2$ tentative decisions obtained from the previous iteration are employed to assist the testing. Note that the previous tentative decisions at the first iteration are set to zeros by default.

Obviously, the iterative rule differs from the block rule primarily in the frequency of computing the metric $|V(\phi)|^2$. Unlike the block rule, the metric is not computed for all phasor blocks by the iterative rule, but only for the phasor blocks that may be located near (in Hamming distance) to the correct one. Because the metric $|V(\phi)|^2$ is computed only for all L possible phasors at each of $I(M - 1)$ tentative decisions, the required complexity grows linearly with M. Table II compares the implementation complexity between the block and iterative rules in terms of the numbers of real operations for processing $\mathbf{z}$ in one block interval of KT seconds. The metric form (26) for $V(\phi)$ is considered for both rules. In counting the required operations, all the matrices $\phi^t \mathbf{D}^{-1/2} \mathbf{E}^t$ in the block rule and $\mathbf{D}^{-1/2} \mathbf{E}^t$ in the iterative rule are assumed prestored and not counted in complexity evaluation. As shown, the iterative rule is much simpler to implement than the block rule when M is very large.

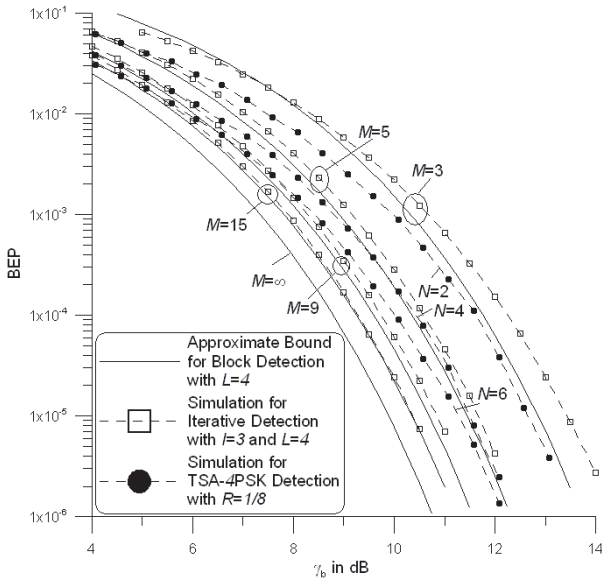[7]The equality in (40) is obtained due to the linearity of $V(\cdot)$.

Fig. 7.   BEP comparison among block detection scheme with $L = 4$, iterative detection scheme with $I = 3$ and $L = 4$, and TSA-4PSK detection scheme with $R = 1/8$.
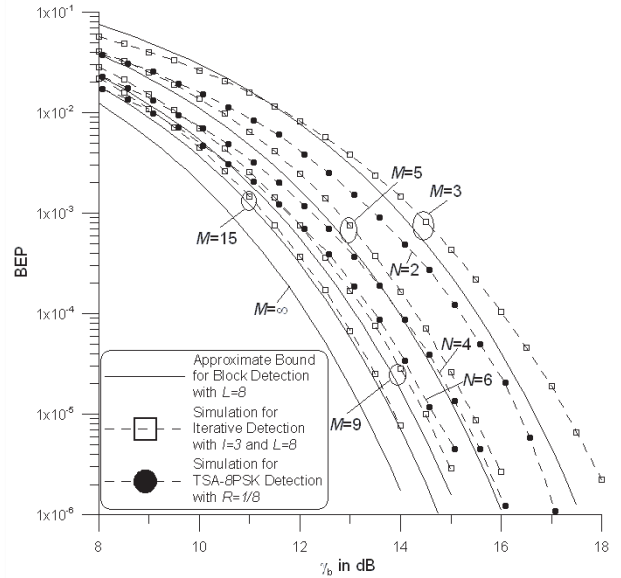

Fig. 8.   BEP comparison among block detection scheme with $L = 8$, iterative detection scheme with $I = 3$ and $L = 8$, and TSA-8PSK detection scheme with $R = 1/8$.

Digital simulation has been conducted to evaluate the BEP performance of the iterative detection scheme for demodulating the Gray-labeled data phasor block-modulated signals composed by the transform in *Example One* of Table I. It is found by the author that the iterative scheme can provide better performance by increasing iterations. Particularly, significant performance improvement is achieved by making final decision at the second iteration than at the first iteration and the improvement become infinitesimal by making final decision after more than three iterations. Figs. 7 and 8 illustrate the simulated BEP results of the iterative detection scheme with $I = 3$ for $L = 4$ and 8, respectively, and compare the results with the approximate bound results of the block detection scheme. It is shown that the iterative detection scheme provides the performance results only slightly degrading from those of the block detection scheme with the same block length. Moreover, the ultimate performance results of the block detection scheme with an infinite block length can be closely approached by the iterative detection scheme when $M \geq 15$.

C) Comparison with Conventional TSA Scheme: Figs. 7 and 8 also compare the BEP characteristics among the proposed schemes and the conventional TSA-LPSK scheme [13]. In the TSA-LPSK scheme, the symbol stream is transmitted in blocks, each block consisting of a sequence of N training symbols followed by a sequence of $N(1/R - 1)$ LPSK data symbols. Both DC offset and channel response are assumed to be constant over each block and estimated in the prefix training sequence by the least square approach in [13, eq. 22] with the training sequence optimized to exhibit a zero average [13, Subsection III.E]. With both DC offset and channel response compensated by the least square estimates [13, eq. 11], the LPSK detection is conducted in a standard coherent mode on the suffix data sequence. It is found from simulation that the TSA scheme can provide better performance by increasing N and the performance improvement becomes small when N is

set larger than six. Note also that the transmission rate (in bits/symbol) is given by $(1 - R) \log_2 L$ for the TSA scheme and $(1 - 2/(M + 1)) \log_2 L$ for the proposed schemes. With these in mind, we purposely illustrate in Figs. 7 and 8 the BEP comparison among the proposed schemes with $M = 15$ and the TSA scheme with $R = 1/8$ and N varied, all with the same transmission throughput efficiency. As shown, the proposed schemes with $M = 15$ remarkably outperform the TSA scheme with $R = 1/8$. Because the block length (in symbols) is given by $M + 1$ for the proposed schemes and N/R for the TSA scheme, the proposed schemes with $M = 15$ use a block length much smaller than the TSA scheme with $N \geq 4$. This implies that the proposed schemes can be used for broader applications where both DC offset and channel response are only required to remain constant over a short processing window.

V. CONCLUSION

In the paper, a generalized linear block transform is employed to counteract the redundant DC offset which may exist in the baseband processing of certain wireless communication receivers. Operating on the transform output, both ML techniques of implicit channel estimation and noncoherent block detection are investigated and analyzed for the AWGN channel. A new digital phasor block modulation format is also designed in conjunction with the noncoherent detection scheme to achieve performance prevalence. When the block length is long enough, the block scheme for noncoherently detecting the Gray-labeled L-ary phasor block-modulated signal in the presence of DC offset is analytically shown to perform comparably with the ideal coherent Gray-labeled LPSK signal detection in the absence of DC offset. To reduce the realization complexity, an iterative detection scheme is also modified from the block detection scheme and shown by simulation to perform comparably with its counterpart

block scheme. Both block and iterative detection schemes are shown to outperform the conventional training-sequence-aided PSK detection scheme under the same transmission throughput efficiency, while enabling broader applications where both DC offset and channel response are only required to remain constant over a short processing window.

REFERENCES

[1] D. Y. Kim and K. Feher, "Power suppression at the Nyquist frequency for pilot-aided PAM and QAM systems," *IEEE Trans. Commun.*, vol. 37, pp. 984-986, Sep. 1989.

[2] A. Bateman, "Feedforward transport tone-in-band: Its implementations and applications," *IEEE Trans. Veh. Technol.*, vol. 39, pp. 235-243, Aug. 1990.

[3] I. Korn, "Coherent detection of M-ary phase-shift keying in the satellite mobile channel with tone calibration," *IEEE Trans. Commun.*, vol. 37, pp. 997-1003, Oct. 1989.

[4] J. K. Cavers, "Performance of tone calibration with frequency offset and imperfect pilot filter," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 426-434, May 1991.

[5] M. K. Simon *et al.*, *Spread Spectrum Communications*. Rockville, MD: Computer Science, 1985, 3 vols.

[6] A. A. Hassan, "Error rate for optimal follower tone-jamming," *IEEE Trans. Commun.*, vol. 44, pp. 546-548, May 1996.

[7] M. Zeng and Q. Wang, "On the probability of differential phase perturbed by tone interference and Gaussian noise," *IEEE Trans. Commun.*, vol. 47, pp. 508-510, Apr. 1999.

[8] E. Larson, *RF and Microwave Circuit Design for Wireless Communications*. Norwood, MA: Artech House, 1996.

[9] S. Mirabbasi and K. Martin, "Classical and modern receiver architectures," *IEEE Commun. Mag.*, vol. 38, pp. 132-139, Nov. 2000.

[10] A. Bateman and D. M. Haines, "Direct conversion transceiver design for compact low-cost portable mobile radio terminals," in *Proc. 39th IEEE Veh. Technol. Conf.*, 1989, pp. 57-62.

[11] J. K. Cavers and M. W. Liao, "Adaptive compensation for imbalance and offset losses in direct conversion transceivers," *IEEE Trans. Veh. Technol.*, vol. 42, pp. 581-588, Nov. 1993.

[12] B. Lindquist, M. Isberg, and P. W. Dent, "A new approach to eliminate the DC offset in a TDMA direct conversion receiver," in *Proc. 43rd IEEE Veh. Technol. Conf.*, May 1993, pp. 754-757.

[13] B. Lindoff and P. Malm, "BER performance analysis of a direct conversion receiver," *IEEE Trans. Commun.*, vol. 50, pp. 856-865, May 2002.

[14] I.-H. Sohn, E.-R. Jeong, and Y.H. Lee, "Data-aided approach to I/Q mismatch and DC offset compensation in communication receivers," *IEEE Commun. Lett.*, vol. 6, pp. 547-549, Dec. 2002.

[15] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York: John Wiley and Sons, 1968.

[16] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis, 2nd ed*. New York: John Wiley and Sons, 1984.

[17] I. Korn, *Digital Communications*. New York: Van Nostrand Reinhold, 1985.

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

[19] H. Meyer, M. Moeneclaey, and S. Fechtel, *Digital Communication Receivers, Synchronization, Channel Estimation, and Signal Processing*. New York: Wiley, 1998.

[20] Y. Jiang, F.-W. Sun, and J. S. Baras, "On the performance limits of data-aided synchronization," *IEEE Trans. Inform. Theory*, vol. 49, pp. 191-203, Jan. 2003.

[21] S. J. Press, *Applied Multivariate Analysis*. New York: Holt, Rinehart and Winston, 1972.

[22] M.K. Simon, S.M. Hinedi, and W.C. Lindsey, *Digital Communication Techniques: Signal Design and Detection*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

Char-Dir Chung (S'87-M'89-SM'03) received the B.S. degree in electrical engineering from the National Taiwan University, Taipei, in 1983, and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles, in 1986 and 1989, respectively.

From 1989 to 1992 he was with the LinCom Corporation, Los Angeles, where he worked on analytical and simulation modeling of scientific and military satellite communication systems. From 1992 to 2005 he was with the National Central University, Chung-Li, Taiwan, ROC, first as an Associate Professor, then as a Professor since 1996. At the National Central University, he founded the Advanced Communications Laboratory in 1998, the Graduate Institute of Communications Engineering in 2000 and the Department of Communications Engineering in 2003, and was the founding heads of these organizations. Since 2005 he has been on the faculty at the National Taiwan University, Taipei, Taiwan, ROC, where he is now a Professor of the Graduate Institute of Communications Engineering and the Department of Electrical Engineering. His current research interests are in digital modulation theory, wireless communications, and spread spectrum communications. He has published more than 40 journal and conference papers and holds two patents in these areas.

Dr. Chung received the Group Achievement Award from the National Aeronautics and Space Administration, USA, in 1991, the Young Scientists Award from the International Union of Radio Science in 1993, the annual Research Award from the National Science Council, ROC, in 1992 and from 1994 to 2001, the Kentucky Colonel grade from the Commonwealth of Kentucky, USA, in 2003, and the FORMOSAT-2 Satellite Project Award from the National Space Center, ROC, in 2005. He served as the Chairman of IEEE Information Theory Society, Taipei Chapter, from 1997 to 1999. From 2000 to 2004 he was an Editor for the Journal of the Chinese Institute of Electrical Engineering, and since 2003 he has been an Editor for the Electrical Engineering Magazine of the same organization. Since 2004, he has been appointed the Chairman of the Wireless System Group of the National Science and Technology Program for Telecommunications, ROC. In past ten years, he has also actively served as a technical referee of numerous industrial product and technology development projects for several ROC government agencies.