Optimal and Approximate Methods for Detection of Uncoded Data with Carrier Phase Noise

Rajet Krishnan, Hani Mehrpouyan, Thomas Eriksson, and Tommy Svensson
Division of Communication Systems and Information Theory
Department of Signals and Systems, Chalmers University of Technology
E-mail: {rajet, hanim, thomase, tommy.svensson}@chalmers.se

Abstract— **Previous results in the literature have shown that derivation of the optimum *maximum-likelihood* (ML) receiver for *symbol-by-symbol* (SBS) detection of an uncoded data sequence in the presence of *random phase noise* is an intractable problem, since it involves the computation of the conditional *probability distribution function* (PDF) of the phase noise process. In this paper, we seek to minimize *symbol error probability* (SEP), which is achieved by SBS detection of the sequence based on all received signals. We show that the ML detector for this problem can be formulated as a weighted sum of central moments of the conditional PDF of phase noise. Given that the central moments of the conditional PDF of phase noise can be estimated, this new optimal structure is tractable with respect to the previously known optimal ML receiver. Furthermore, based on the new receiver structure, we propose a simple approximate method for SBS detection and investigate its scope and applicability. Simulation results demonstrate that SEP performance close to optimality can be obtained through the proposed method for scenarios of low phase noise variance and low *signal-to-noise ratio* (SNR).**

I. Introduction

Local oscillator instabilities that result in *random, time-varying* phase difference between the transmitter and receiver, have been one of the major impediments towards realizing a reliable coherent communication system [1]. This impairment, referred to as phase noise can result in significant performance loss if not compensated appropriately.

The problem of receiver design of uncoded data detection in the presence of a *random phase noise process* has been studied for decades, e.g., refer to [1], [2] and references therein. One of the earlier and important approaches to this problem was reported in [3], which proposed simultaneous *maximum-likelihood* (ML) estimation of the data sequence and phase noise. However, it was not proved if the approach ensures optimality in terms of achieving minimum *symbol error probability* (SEP). In [4], it was shown that the simultaneous approach proposed in [3] is optimal in the high *signal-to-noise ratio* (SNR) regime.

A receiver structure, optimum with respect to the minimum SEP criterion, was first derived in [5]. Specifically, it was illustrated that the optimum *symbol-by-symbol* (SBS) receiver has a separable estimator-detector structure, i.e., all the received signals are used to first compute/estimate the posteriori or conditional *probability density function* (PDF) of phase noise. The information in this posteriori density function is then used to detect a data symbol. However, it was observed that this optimum detector can only be realized if the conditional PDF of phase noise has a closed-form expression. In general, the problem of computing the conditional PDF of phase noise given all the received signals has been demonstrated to be an infinite dimensional problem [5], and the optimum receiver structure is found to only be analytically tractable under restrictive assumptions on the phase noise distribution and the receiver structure.

The analytical intractability of the optimum receiver structure in [5], combined with limited scope of data-aided phase estimation schemes, spurred interest in designing joint data sequence detection and phase estimation algorithms, instead of SBS detection. Some related examples are the per-survivor processing algorithm in [6], tree-pruning algorithm in [7], and expectation maximization algorithm in [8]. Generalized-likelihood based joint data sequence detection and phase noise estimation algorithm of polynomial complexity was proposed in [9]. For a *constant phase offset model*, the algorithm was observed to achieve performance close to that of the optimal ML receiver. Iterative methods based on factor-graphs for data detection and phase noise estimation were proposed in [10]. In [11], an adaptive maximum likelihood sequence detection scheme based on Viterbi algorithm was proposed for uncoded data sequence detection in the presence of a random phase noise process. However, it is well known that sequence detection schemes in [6]–[11] do not achieve optimal SEP performance in the presence of a random phase noise process. Application of Monte Carlo sampling methods to phase noise estimation and uncoded data detection was investigated in [12], which incurs high computational complexity.

In this paper, motivated by the optimal receiver structure derived in [5], we re-visit the problem of optimal detection of an uncoded data sequence in the presence of phase noise. We seek to minimize SEP, which is achieved by SBS detection of the sequence. Detection of each symbol is based on the entire received signal, corresponding to the entire data sequence, thereby accounting for correlated phase distortion (memory) in the received signals. The contributions and organization of this paper can be summarized as follows:

- In Section II, the system model for optimal detection of an uncoded data sequence in the presence of phase noise is presented. This is similar to that in [5].
- In Section III, without making any assumptions on the PDF of the phase noise process, we derive an alternative form of the ML receiver that is analytically tractable with respect to the original receiver in [5]. Specifically, the ML detector is formulated as a weighted sum of the central moments of the conditional PDF of the phase noise as opposed to the convoluted conditional PDF computation in [5].
- Furthermore, in Section III, we present an analytical method to approximate the alternative ML rule by truncating it to a finite number of terms while still ensuring that its SEP performance is close to optimal. Then, we truncate the new optimal ML rule to two terms for SBS detection and investigate its scope and applicability.
- In Section IV, we present our simulation results, which demonstrate that performance close to that of the optimal ML detector can be achieved for scenarios of low phase

noise variance and low SNR through the proposed truncation approach.

Notations: Expectation operator is denoted as $\mathbb{E}[\cdot]$. $[\cdot]^T$ denotes transpose and $[\cdot]^H$ denotes Hermitian of a vector, and $\mathbf{I}$ denotes the identity matrix. Re$(\cdot)$, Im$(\cdot)$, and arg$(\cdot)$ are the real, imaginary part, and angle of a complex number respectively.

II. System Model

Consider a system with the following received signal model in the kth time slot

$$r_k = m_k e^{j\theta_k} + n_k, \tag{1}$$

where, r_k is the received signal, m_k is the transmitted symbol, θ_k is the unknown phase noise, and n_k is complex Gaussian noise in the kth time slot. $\mathbf{r} \triangleq [r_0, \ldots, r_{L-1}]^T$ represents the vector of all L received symbols in L time slots. We assume transmission of uncoded data that are denoted in the vector form as $\mathbf{m} \triangleq [m_0, \ldots, m_{L-1}]^T$. Since the data is uncoded, we assume that all elements in $\mathbf{m}$ are independent of each other, and are transmitted with equal probability. Also, it can assume the value of any point $\{S_i, \forall i \in \{1, \ldots, C\}\}$ in the signal constellation, where C is the total number of signal points in the constellation. Let $\boldsymbol{\theta} \triangleq [\theta_0, \ldots, \theta_{L-1}]^T$ denote the vector of unknown phase noise random variables, where no assumptions are made on its PDF. It is assumed that $\mathbf{m}$ and $\boldsymbol{\theta}$ are independent of each other. The AWGN channel is $\mathbf{n} \triangleq [n_0, \ldots, n_{L-1}]^T$, i.e., it is a vector of independent identically distributed (i.i.d.) complex Gaussian random variables with $\mathbb{E}[\mathbf{n}] = [0, \ldots, 0]^T$, and $\mathbb{E}[\mathbf{n}\mathbf{n}^H] = N_0\mathbf{I}$, i.e., $n_k \sim \mathcal{CN}(0, N_0)$.

We investigate the problem of optimum symbol detection based on all received signals, $\mathbf{r}$, such that the SEP is minimized. It is known that optimum SBS detection of the kth symbol that minimizes SEP is obtained by ML detection [13]. Thus, the optimum receiver for the kth symbol is given by

$$\max_{i\in\{1,\ldots,C\}} L_i(k) = \max_{i\in\{1,\ldots,C\}} p(\mathbf{r}|m_k = S_i), \tag{2}$$

In the case of optimum ML detection, in [5], it has been shown that $L_i(k)$ reduces to the following

$$L_i(k) = \int_{-\pi}^{\pi} p(r_k|m_k = S_i, \theta(k)) p(\theta_k|\bar{\mathbf{r}}_k) d\theta_k, \tag{3}$$

where $\bar{\mathbf{r}}_k \triangleq [r_0, \ldots, r_{k-1}, r_{k+1}, \ldots, r_{L-1}]^T$, refers to all signals received outside the kth interval. The optimum ML detector first involves the estimation of the conditional PDF of phase noise in an interval using all signals received outside it. This conditional PDF is then used to perform data detection using (3). As shown in [5], the conditional PDF $p(\mathbf{r}|m_k = S_i)$ can be determined only in special cases.

The detector in (3) reduces to the conventional receiver approach when the carrier phase is first recovered by a phase estimator, followed by coherent detection of the symbols, i.e., the recovered phase $\hat{\theta}_k$ is treated as the true value of θ_k. To illustrate the aforementioned amenability, consider the conditional PDF of θ_k to be a distribution with zero variance or equivalently a delta function, i.e., $p(\theta_k|\bar{\mathbf{r}}_k) = \delta(\theta - \hat{\theta}_k)$. The ML data decision rule is then derived from (2) and (3) as follows

$$\max_{i\in\{1,\ldots,C\}} L_i(k) = \max_{i\in\{1,\ldots,C\}} \frac{e^{-\frac{|r_k - S_i e^{j\hat{\theta}_k}|^2}{2N_0}}}{(2\pi N_0)^{1/2}}. \tag{4}$$

Thus, if the recovered $\hat{\theta}_k$is treated as the true value of θ_k, the ML decision rule in (3) becomes equivalent to the minimum distance based coherent detection rule [14].

III. Alternative form for ML Decision Rule

In this section, we seek to derive alternative forms of the optimum receiver for uncoded data in the presence of phase noise. Particularly of interest are ML detector structures that are tractable in their exact or approximate form. Adopting the system model discussed above, consider the problem of data detection in the kth time slot. Assume θ_k to be drawn from an *arbitrary probability distribution*. Then, by performing Taylor series expansion of $f(\theta_k) = p(r_k|m_k = S_i, \theta_k)$ about $\theta_k = \hat{\theta}_k$ in (3), it can be rewritten as

$$\begin{aligned}\max_{i\in\{1,\ldots,C\}} L_i(k) &= \max_{i\in\{1,\ldots,C\}} \frac{1}{(2\pi N_0)^{1/2}} \int_{-\pi}^{\pi} \left[\frac{f(\hat{\theta}_k)}{0!} + \frac{f^{\{1\}}(\hat{\theta}_k)}{1!} \right. \\ &\quad \left. \times(\theta_k - \hat{\theta}_k) + \frac{f^{\{2\}}(\hat{\theta}_k)}{2!}(\theta_k - \hat{\theta}_k)^2 + \ldots \right] \\ &\quad \times p(\theta_k|\bar{\mathbf{r}}_k) d\theta_k, \\ &= \max_{i\in\{1,\ldots,C\}} \frac{1}{(2\pi N_0)^{1/2}} \left[\frac{f(\hat{\theta}_k) M_0}{0!} \right. \\ &\quad \left. + \frac{f^{\{1\}}(\hat{\theta}_k) M_1}{1!} \ldots + \frac{f^{\{n\}}(\hat{\theta}_k) M_n}{n!} + \ldots \right].\end{aligned} \tag{5}$$

That is, the decision rule in equation (3) is equivalent to the maximization of the weighted sum of $M_j, j \in \mathbb{Z}^+$ over $S_i \in \{1, \ldots, C\}$. Here, M_j is the jth central moment of the conditional PDF, $p(\theta_k|\bar{\mathbf{r}}_k)$, and $f^{\{n\}}(\hat{\theta}_k)$ is the nth derivative of $f(\theta_k)$ given by

$$f(\theta_k) = \frac{e^{-\frac{|r_k - S_i e^{j\theta_k}|^2}{2N_0}}}{(2\pi N_0)^{1/2}}, \tag{6}$$

and evaluated at $\theta_k = \hat{\theta}_k$. For the Taylor series expansion in (5) to converge to $f(\theta_k)$ for all values of θ_k, it is required that $f(\theta_k)$ be an entire function in θ_k, $\forall \theta_k \in \mathbb{R}$. The proof for this is given in Appendix A.

In deriving (5) no restrictive assumptions are made on the distribution of θ_k. In addition, we do not assume any form of decision feedback or data-aided mode of operation at the receiver. Thus, the problem of determining the optimum ML detector is reduced to estimating the central moments of the conditional PDF of θ_k, as opposed to estimating the PDF itself in [5]. The central moments of a distribution can be estimated for a given data set [17]. In its exact form, the new receiver structure incurs high computational complexity on the receiver, thereby constraining practical utility. As we shall see in the sequel, the parametric form of the ML detection rule in (5) allows simple approximation by using a finite number of terms and obtains performance close to that of the original ML rule.

A. Truncation of the Sum-of-Central-Moments ML Rule

The alternative ML decision rule in (5) shows that the optimal ML decision for achieving minimum SEP requires a knowledge of all central moments of the conditional PDF, which is equivalent to having complete knowledge of the distribution. In this section, we present two techniques to truncate the rule in (5) to a finite number of terms.

1) Determine the number of terms to be retained, n: We first seek to determine the number of terms that are to be retained in a truncated version of (5) such that the error of this approximate rule, with respect to the original ML rule is very small. This ensures that the SEP performance of the approximate ML rule is close to that of the optimal ML rule. Consider an approximate

ML rule obtained by retaining n terms in the Taylor series as follows

$$\max_{i\in\{1,\ldots,C\}} L_i(k) = \max_{i\in\{1,\ldots,C\}} \frac{1}{(2\pi N_0)^{1/2}} \left[\frac{f(\hat{\theta}_k)M_0}{0!} + \frac{f^{\{1\}}(\hat{\theta}_k)M_1}{1!} + \ldots + \frac{f^{\{n\}}(\hat{\theta}_k)M_n}{n!} \right]. \quad (7)$$

The upper bound on the error for this approximation with respect to the original ML rule is given as

$$\epsilon_{n+1} \leq \frac{1}{(2\pi N_0)^{1/2}} \frac{\left(\frac{\left|\mathrm{Im}\{r_k^* S_i e^{j\theta_1}\}\left(\theta_m - \hat{\theta}_k\right)\right|}{N_0} \right)^{n+1}}{(n+1)!}, \quad (8)$$

where the proof for (8) is presented in Appendix B. Equation (8) can be used to determine the number of terms to be retained in the approximate ML rule, such that the error due to truncation of (5) to a finite number of terms diminishes to a sufficiently small value. This in turn ensures that the approximate rule is close to the original ML rule. For the error to diminish to a sufficiently small value, it is straightforward to see that we need n, such that

$$\left(\frac{\left|\mathrm{Im}\{r_k^* S_i e^{j\theta_1}\}\left(\theta_m - \hat{\theta}_k\right)\right|}{N_0} \right)^{n+1} < (n+1)!. \quad (9)$$

Since an upper bound on the approximation error is used in (9), numerically solving the inequality gives an upper bound on the number of terms that are to be retained in the approximate decision rule.

2) Fix the number of terms to be retained, n: Another approach to approximating the ML decision rule by truncation is to fix the number of terms in the Taylor series expansion, n, and investigate the various scenarios where the approximate decision rule achieves SEP performance close to that of the ML decision rule. Consider the case where the Taylor series is truncated to $n = 2$; i.e., only the first three terms in the optimal decision rule in (5) are considered. This case is particularly interesting as it corresponds to scenarios where the conditional distribution of θ_k is unknown except for its mean and variance. We first define an approximate SBS detection rule for uncoded data over AWGN channel as

$$\begin{aligned}\max_{i\in\{1,\ldots,C\}} A_i(k) &= \max_{i\in\{1,\ldots,C\}} \left[\frac{f(\hat{\theta}_k)M_0}{0!} + \frac{f^{\{1\}}(\hat{\theta}_k)M_1}{1!} + \frac{f^{\{2\}}(\hat{\theta}_k)M_2}{2!} \right], \\ &= \max_{i\in\{1,\ldots,C\}} \left[\frac{f(\hat{\theta}_k)M_0}{0!} + \frac{f^{\{2\}}(\hat{\theta}_k)\sigma_p^2}{2!} \right]. \quad (10)\end{aligned}$$

The second-order approximate ML rule in (10) consists of two terms; the first term is the zero-th order term from the Taylor series and is identical to the minimum distance based coherent symbol detection rule. The second term is the variance of the conditional PDF of θ_k weighted by the second derivative of $f(\theta_k) = p(r_k|m_k = S_i, \theta_k)$, which intuitively gives a measure of sharpness or curvature of $p(r_k|m_k = S_i, \theta_k)$ about $r_k = S_i e^{j\hat{\theta}_k}$. The most likely symbols result in high magnitude of sharpness of $p(r_k|m_k = S_i, \theta_k)$ about $r_k = S_i e^{j\hat{\theta}_k}$. Thus, the objective function in the optimization problem characterizing the decision rule in (10) is intuitively appealing, in that it can be viewed as a weighted combination of the distance based measure (as from coherent detection) and the curvature of $p(r_k|m_k = S_i, \theta_k)$ weighted by the variance of the conditional distribution of θ_k.

The upper bound on the error of this approximation is given as

$$\epsilon_3 \leq \left| \frac{\left(\theta_m - \hat{\theta}_k\right)^3}{6(2\pi N_0)^{1/2}} \right| \left[\left| - \left(\frac{\mathrm{Im}\{r_k^* S_i e^{j\theta_1}\}}{N_0} \right)^3 \right| + \left| \frac{\mathrm{Im}\{r_k^* S_i e^{j\theta_1}\}}{N_0} \right| + \frac{3\left|\mathrm{Im}\{r_k^* S_i e^{j\theta_1}\}\right| \left|\mathrm{Re}\{r_k^* S_i e^{j\theta_2}\}\right|}{N_0^2} \right] \quad (11)$$

The proof for the approximation error bound is presented in Appendix B. From (11), we develop insight into the scenarios where the approximate decision rule would be close to the ML decision rule in SEP performance.

- The error in approximation is inversely proportional to AWGN channel noise variance. Hence the error decreases with increase in the variance of the AWGN channel or equivalently with decreasing SNR for a given constellation and phase noise variance.
- The error in approximation is directly proportional to the magnitude of phase noise relative to the mean $\hat{\theta}_k$ of the conditional PDF of θ_k.
- The error in approximation is directly proportional to the magnitude of the symbol point in the constellation. This implies that the error in the second-order approximated ML rule increases with increase in the size of the constellation for a given AWGN channel noise variance and phase noise variance.

IV. Simulations and Discussion

By simulations, we first seek to investigate the number of terms, n, required to diminish the error in the approximate decision rule with respect to the optimal ML decision rule. It is difficult to simulate the approximate decision rule when the required number of terms, n, is large. Hence, we simulate the performance of the decision rule given by the difference between the original rule in (3), and the upper bound on the error in (8) resulting from truncation of the ML rule to an arbitrary n terms. The phase noise random variables are considered to be Gaussian i.i.d. with variance σ_{p}^2. This model is valid for a phase noise process in the case of a locked *phase-locked loop* (PLL) with small loop bandwidth [18]. We consider 16-QAM modulation scheme ($C = 16$), a fixed SNR of 11 dB, and different conditional PDF variance values, $\sigma_{\mathrm{p}}^2 = 10^{-1}, 10^{-2}, 10^{-3}$. Fig. 1 illustrates the dependence of SEP performance of the approximate rule on n, for different values of σ_{p}^2. For a given constellation and SNR, we observe that the number of terms required increases as the phase noise variance increases. Therefore, when σ_{p}^2 is large, higher number of central moments of the conditional PDF of θ_k, and higher order derivatives of $f(\theta_k)$ are required in the approximate decision rule. Note that n in Fig. 1 is an upper bound on the number of terms to be retained in the truncated ML rule. In the ensuing discussion, we observe that the approximate ML rule with just two terms approaches optimal SEP performance for cases of low phase noise variance and low SNR. This also includes the case of $\sigma_{\mathrm{p}}^2 \leq 10^{-2}$ considered above for 16-QAM.

We now discuss simulation results demonstrating SEP performance versus SNR per bit using (10) as the decision rule for detecting uncoded data. Two modulation schemes with relatively different constellation order are considered for study: (i) A lower order $16-$QAM constellation, and (ii) a higher order $1024-$QAM

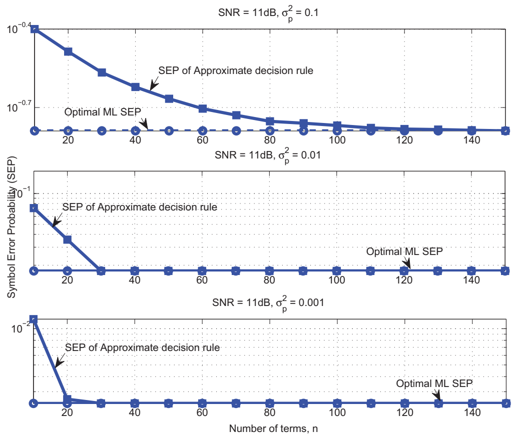

Fig. 1: Convergence of the approximate decision rule to the optimal ML with increase in n for different σ_{p}^2 values.

constellation. This choice is motivated by the analytical observation that the error in approximation would depend on the size of the constellation. The phase noise random variables are again considered to be Gaussian i.i.d. with variance σ_p^2.

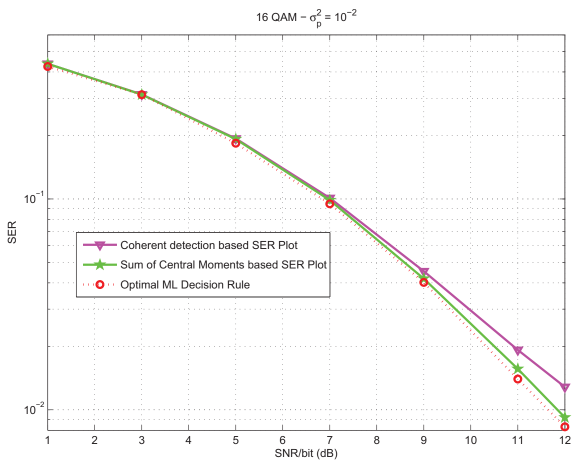

Fig. 2: Comparison of SEP performance between the Optimal ML decision rule, approximate rule and coherent detection for 16 QAM, $\sigma_{\mathrm{p}} = 10^{-2}$.

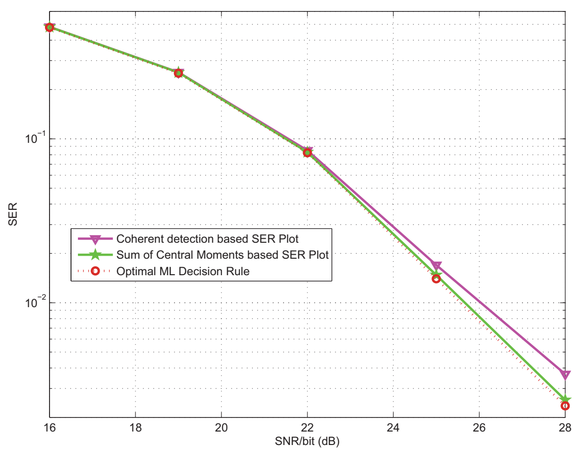

Fig. 3: Comparison of SEP performance between the Optimal ML decision rule, approximate rule and coherent detection for 1024 QAM, $\sigma_{\mathrm{p}} = 10^{-4}$

1) Dependence on AWGN variance: Fig. 2 presents results for $C = 16$, $\sigma_p^2 = 10^{-2}$ and SNR per bit up to 12 dB. It can be easily observed that the approximate rule in (10) outperforms the coherent detector (4) till around 12 dB SNR per bit. At 12 dB, a gain of 1 dB in SNR per bit is observed for the approximate method when compared to the case of coherent detection. Also, in this SNR regime, the performance of the approximate decision rule is observed to be close to that of the optimal ML. Similar observations can be made for the case $C = 1024$, $\sigma_{\mathrm{p}}^2 = 10^{-4}$ in Fig. 3, where a gain of around 1 dB is observed at around 28 dB SNR per bit. We observe that as AWGN variance is decreased, the maximum variance of the conditional PDF of phase noise for which the approximation gives performance close to optimal SEP, decreases.
2) Dependence on phase noise variance: We observe that SEP performance of the rule in (10) is close to that of the optimal ML rule when $\sigma_{\mathrm{p}}^2 < 10^{-2}$ for $C = 16$, and SNR ≤ 12 dB. This is also observed when $\sigma_{\mathrm{p}}^2 \leq 10^{-4}$ for $C = 1024$, and SNR ≤ 28 dB. In both the cases, a gain of around 1 dB can be observed when compared to the performance of coherent detection. As σ_{p}^2 is increased, the maximum SNR for which the SEP performance of the approximate rule is close to the optimum, decreases.
3) Dependence on size of constellation: For a given SNR and phase noise variance, the error in the second-order approximated ML rule increases with increase in the distance of a symbol point from the origin of the constellation. With increase in the size of the constellation, the maximum variance for which the approximation renders SEP close to that of the optimal ML rule, decreases.

In general, we observe that the second order approximation of the ML rule operates close to the original ML SEP in scenarios of low phase noise variance and high AWGN noise. When the decision rule in (10) deviates significantly from the original ML rule, higher number of terms needs to be included in order to realize SEP performance that is close to optimum. Note that no form of decision feedback or data aided schemes have been employed at the receiver while implementing the approximate ML rule. The additional computation involved relates to evaluating the function $f(\theta_k)$, its derivatives at $\hat{\theta}_k$, and the mean and variance of the conditional PDF of θ_k.

V. Conclusions

We show that the ML data detector for symbol by symbol detection in the presence of phase noise can be formulated as a weighted sum of central moments of the conditional PDF of phase noise. We present an analytical method to determine the number of terms to be retained in the approximate ML decision rule that still ensures SEP performance close to optimum. Furthermore, we approximate the optimal structure by truncating the ML rule to two terms and observe that this approximation renders SEP performance close to optimum for low phase noise variance and low SNR.

APPENDIX A

Proof that $f(\theta_k)$ is an entire function in θ_k

Lemma 1: If $f(\theta_k) = \frac{1}{(2\pi N_0)^{1/2}} e^{-\frac{|r_k - S_i e^{j\theta_k}|^2}{2N_0}}$, then the nth derivative of $f(\theta_k)$ evaluated at any arbitrary point $\theta_k = \hat{\theta}_k$ is of the form

$$f^{\{n\}}(\hat{\theta}_k) = w_n \left(\frac{\operatorname{Im}\{r_k^* S_i e^{j\hat{\theta}_k}\}}{N_0} \right)^n, \tag{A-1}$$

where $\left(\frac{\operatorname{Im}\{r_k S_i e^{j\hat{\theta}_k}\}}{N_0} \right)^n$ is the highest exponential power in $f^{\{n\}}\left(\hat{\theta}_k\right)$, and w_n is a function of $r_k, S_i, e^{j\hat{\theta}_k}$, and N_0.

Proof: For $n = 1$, it is trivial to see that $f^{\{1\}}(\hat{\theta}_k) = w_1 \frac{\mathrm{Im}\{r_k S_i e^{j\hat{\theta}_k}\}}{N_0}$. Hence, we first prove that (A-1) holds for $n = 2$ as follows

$$\begin{aligned} f^{\{2\}}(\hat{\theta}_k) &= \frac{e^{-\frac{|r_k - S_i e^{j\hat{\theta}_k}|^2}{2N_0}}}{(2\pi N_0)^{1/2}} \frac{\left(r_k^* S_i e^{j\hat{\theta}_k} - r_k S_i^* e^{-j\hat{\theta}_k}\right)^2}{4N_0^2} \\ &\times \left[-1 - \frac{2N_0\left(r_k^* S_i e^{j\hat{\theta}_k} + r_k S_i^* e^{-j\hat{\theta}_k}\right)}{\left(r_k^* S_i e^{j\hat{\theta}_k} - r_k S_i^* e^{-j\hat{\theta}_k}\right)^2}\right], \\ &= w_2 \left(\frac{\mathrm{Im}\{r_k^* S_i e^{j\hat{\theta}_k}\}}{N_0}\right)^2. \end{aligned} \quad \text{(A-2)}$$

Assume that (A-1) holds true for $n \in \mathbb{N}$, i.e.,

$$\begin{aligned} f^{\{n\}}(\hat{\theta}_k) &= w_n \left(\frac{\mathrm{Im}\{r_k S_i e^{j\hat{\theta}_k}\}}{N_0}\right)^n, \\ &= w_n' e^{\frac{-|r_k - S_i e^{j\hat{\theta}_k}|^2}{2N_0}} \left(\frac{\mathrm{Im}\{r_k^* S_i e^{j\hat{\theta}_k}\}}{N_0}\right)^n. \end{aligned} \quad \text{(A-3)}$$

Now the $(n+1)$th derivative is evaluated as in (A-4). Hence result in (A-1) also holds for $f^{\{n+1\}}(\hat{\theta}_k)$. Since both the basis and the inductive steps have been proven, (A-1) holds true $\forall n \in \mathbb{N}$. ■

Lemma 2: $f(\theta_k)$ is an entire function in θ_k, i.e., the Taylor series expansion of $f(\theta_k)$ is equal to the function for all values of $\theta_k \in \mathbb{R}$.

Proof: The function represented by $f(\theta_k)$ is a real function in $\theta_k \in \mathbb{R}$, given that both its domain and range are real valued. Hence it is analytic $\forall \theta_k \in \mathbb{R}$ if, and only if, it is infinitely differentiable and can be represented by a convergent power series evaluated about any arbitrary point $\hat{\theta}_k \in \mathbb{R}$ [15]. The power series representation of $f(\theta_k)$ about $\hat{\theta}_k \in \mathbb{R}$ is given as

$$f(\theta_k) = \sum_{n=0}^{\infty} \frac{f^{\{n\}}(\hat{\theta}_k)}{n!} \left(\theta_k - \hat{\theta}_k\right)^n, \quad \text{(A-5)}$$

where $f^{\{n\}}, n \in \mathbb{Z}^+$ is the nth derivative of $f(\theta_k)$. $f(\theta_k)$ is a function that is a composition of the exponential function in θ_k. Given that the exponential function is infinitely differentiable; any function that is a composition of an exponential function is also infinitely differentiable [15].

The convergence of the power series (A-5) $\forall \theta_k \in \mathbb{R}$ can be proved by the ratio test. Let a_n denote the n-th term in the power series given in (A-5), where

$$\begin{aligned} a_n &= \frac{f^{\{n\}}(\hat{\theta}_k)}{n!} \left(\theta_k - \hat{\theta}_k\right)^n \\ &= \frac{w_n \left(\frac{\mathrm{Im}\{r_k S_i e^{j\hat{\theta}_k}\}}{N_0}\right)^n}{n!} \left(\theta_k - \hat{\theta}_k\right)^n. \end{aligned}$$

Here, $f^{\{n\}}(\theta_k)$ at $\theta_k = \hat{\theta}_k$ can be obtained from Lemma 1. Hence by ratio test, we have

$$\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = \lim_{n\to\infty} \left|\frac{w_{n+1} \frac{\mathrm{Im}\{r_k S_i e^{j\hat{\theta}_k}\}}{N_0} \left(\theta_k - \hat{\theta}_k\right)}{w_n n}\right| = 0.$$

The radius of convergence r_{conv} of the series is evaluated as

$$r_{\text{conv}} = \lim_{n\to\infty} \left|\frac{n w_n}{\frac{\mathrm{Im}\{r_k S_i e^{j\hat{\theta}_k}\}}{N_0} w_{n+1}}\right|. \quad \text{(A-6)}$$

When $n \to \infty$, the radius of convergence $r_{\text{conv}} \to \infty$ in (A-6). Now we have that $f(\theta_k)$ is infinitely differentiable and can be represented by a convergent power series for $\forall \theta_k \in \mathbb{R}$. It remains to be shown that the Taylor series converges to the original function $f(\theta_k), \forall \theta_k \in \mathbb{R}$. To prove this, consider a finite Taylor series expansion of $f(\theta_k)$

$$f(\theta_k) = \sum_{n=0}^{n} \frac{f^{\{n\}}(\hat{\theta}_k)}{n!} \left(\theta_k - \hat{\theta}_k\right)^n. \quad \text{(A-7)}$$

The error in this truncated Taylor series with respect to the original function is given by the remainder term in Taylor's theorem as

$$R^{\{n+1\}}(\theta_k) = \frac{f^{\{n+1\}}(\theta_c)}{(n+1)!} \left(\theta_k - \hat{\theta}_k\right)^{n+1}. \quad \text{(A-8)}$$

Here $\theta_c \in \left(\theta_k, \hat{\theta}_k\right)$ if $\theta_k < \hat{\theta}_k$, or $\theta_c \in \left(\hat{\theta}_k, \theta_k\right)$ if $\theta_k > \hat{\theta}_k$. In the limit $n \to \infty$, we have

$$\begin{aligned} \lim_{n\to\infty} R^{\{n+1\}}(\theta_k) &= \lim_{n\to\infty} \frac{f^{\{n+1\}}(\theta_c)}{(n+1)!} \left(\theta_k - \hat{\theta}_k\right)^{n+1}, \\ &= \lim_{n\to\infty} \frac{w_{n+1} \left(\left(\theta_k - \hat{\theta}_k\right) \frac{\mathrm{Im}\{r_k S_i e^{j\hat{\theta}_k}\}}{N_0}\right)^{n+1}}{(n+1)!}, \\ &= 0. \end{aligned} \quad \text{(A-9)}$$

The limit $n \to \infty$ in (A-9) evaluates to zero since in the limit, $n!$ grows faster than $\left(\frac{\mathrm{Im}\{r_k S_i e^{j\hat{\theta}_k}\}}{N_0}\right)^{n+1} \times \left(\theta_k - \hat{\theta}_k\right)^{n+1}$, which is an exponential function in n [16]. We thus prove that $f(\theta_k)$ is an entire function in θ_k. ■

APPENDIX B

PROOF FOR UPPER BOUND ON ϵ_{n+1}

Consider an approximate ML rule obtained by retaining n terms in the Taylor series as follows

$$\max_{i\in\{1,\ldots,C\}} L_i(k) = \max_{i\in\{1,\ldots,C\}} \frac{1}{(2\pi N_0)^{1/2}} \left[\frac{f(\hat{\theta}_k) M_0}{0!} + \frac{f^{\{1\}}(\hat{\theta}_k) M_1}{1!} + \ldots + \frac{f^{\{n\}}(\hat{\theta}_k) M_n}{n!}\right].$$

$$\begin{aligned} f^{\{n+1\}}(\theta_k) &= \frac{dw_n'}{d\theta_k} e^{\frac{-|r_k - S_i e^{j\theta_k}|^2}{2N_0}} \left(\frac{\mathrm{Im}\{r_k^* S_i e^{j\theta_k}\}}{N_0}\right)^n + \frac{\left(r_k^* S_i e^{j\theta_k} - r_k S_i^* e^{-j\theta_k}\right)}{2N_0} \left(\frac{\mathrm{Im}\{r_k^* S_i e^{j\theta_k}\}}{N_0}\right)^n w_n \\ &+ \frac{d\left(\frac{\mathrm{Im}\{r_k^* S_i e^{j\theta_k}\}}{N_0}\right)^{n+1}}{d\theta_k} e^{\frac{-|r_k - S_i e^{j\theta_k}|^2}{2N_0}} w_n' = w_{n+1} \left(\frac{\mathrm{Im}\{r_k^* S_i e^{j\theta_k}\}}{N_0}\right)^{n+1}. \end{aligned} \quad \text{(A-4)}$$

Then, the absolute value of the error of this approximation with respect to the original ML rule is given

$$\epsilon_{n+1} = \int_{-\pi}^{\pi} \underbrace{\left| \frac{f^{\{n+1\}}(\theta_c)}{(n+1)!} \left(\theta_k - \hat{\theta}_k\right)^{n+1} \right|}_{\triangleq |\tau_{n+1}|} p(\theta_k|\bar{\mathbf{r}}_k) d\theta_k, \quad \text{(B-1)}$$

where τ_{n+1} is the error arising from the truncation of the Taylor series [15]. The variable θ_c in τ_{n+1} depends on both $\hat{\theta}_k$ and θ_k, and cannot be explicitly determined. Note that we consider the absolute value of τ_{n+1}, rather than its actual value, since it can be positive or negative due to each θ_k value. Using (B-1), the upper bound on ϵ_{n+1} can be determined as follows

$$\begin{aligned} \epsilon_{n+1} \quad &\approx \quad \int_{-\pi}^{\pi} \underbrace{e^{-\frac{|r_k - S_i e^{j\hat{\theta}_k}|^2}{2N_0}}}_{\triangleq \phi} \frac{\left|\text{Im}\{r_k^* S_i e^{j\theta_c}\} \left(\theta_k - \hat{\theta}_k\right)\right|^{n+1}}{{N_0}^{N+1} (2\pi N_0)^{1/2} (n+1)!} \\ & \quad \times p(\theta_k|\bar{\mathbf{r}}_k) d\theta_k, \quad \text{(B-2a)} \\ &\leq \quad \frac{1}{(2\pi N_0)^{1/2}} \frac{\left(\frac{\left|\text{Im}\{r_k^* S_i e^{j\theta_1}\}(\theta_m - \hat{\theta}_k)\right|}{N_0}\right)^{n+1}}{(n+1)!}, \quad \text{(B-2b)} \end{aligned}$$

where (B-2a) is obtained by first using the approximation

$$f^{\{n+1\}}(\theta_c) \approx \frac{e^{-\frac{|r_k - S_i e^{j\hat{\theta}_k}|^2}{2N_0}}}{(2\pi N_0)^{1/2}} \left(\frac{\text{Im}\{r_k^* S_i e^{j\theta_c}\}}{N_0}\right)^{n+1}, \quad \text{(B-3)}$$

for large values of $\frac{\text{Im}\{r_k^* S_i e^{j\theta_c}\}}{N_0}$, which is verified in Lemma 1 in Appendix A. We have $\text{Im}\{r_k^* S_i e^{j\theta_c}\} = |r_k^* S_i| \sin(\theta_c + \arg\{r_k^* S_i\})$ that is maximum when $\theta_c + \arg\{r_k^* S_i\} = \pi/2$. Thus we set θ_c as $\theta_1 = \pi/2 - \arg\{r_k^* S_i\}$ as in (B-2b). Though θ_k is drawn from $p(\theta_k|\bar{\mathbf{r}}_k)$ and can take any values between $[-\infty, \infty]$, it can be upper bounded to $\theta_m = k\sigma_p$ using the Chebyshev inequality [13], where $k \in \mathbb{R}$ and σ_p^2 is the variance of $p(\theta_k|\bar{\mathbf{r}}_k)$ such that

$$\Pr\left(|\theta_k - \hat{\theta}_k| \geq k\sigma_p = \theta_m\right) \leq \frac{1}{k^2}. \quad \text{(B-4)}$$

In (B-2a), the term $\phi \in [0, 1]$ and is upper bounded to one to finally obtain (B-2b).

A. Derivation of Upper Bound for ϵ_3

The bound on the error for the approximate ML rule in (10) is evaluated as

$$\begin{aligned} \epsilon_3 \quad &= \quad \int_{-\pi}^{\pi} \underbrace{\left| \frac{f^{\{3\}}(\theta_c)}{(3)!} \left(\theta_k - \hat{\theta}_k\right)^3 \right|}_{\triangleq |\tau_3|} p(\theta_k|\bar{\mathbf{r}}_k) d\theta_k, \quad \text{(B-5)} \\ &\leq \quad \left| f^{\{\text{UB}\}}(\theta_c, \theta_k) \right| \int_{-\pi}^{\pi} p(\theta_k|\bar{\mathbf{r}}_k) d\theta_k = \left| f^{\{\text{UB}\}}(\theta_c, \theta_k) \right| \end{aligned}$$

Here $f^{\{\text{UB}\}}(\theta_c, \theta_k)$ refers to the upper bound of τ_3, which is given by

$$\begin{aligned} |\tau_3| &\leq \left| \frac{\left(\theta_k - \hat{\theta}_k\right)^3}{6(2\pi N_0)^{1/2}} \right| \left[\left| - \left(\frac{\text{Im}\{r_k^* S_i e^{j\theta_c}\}}{N_0}\right)^3 \right| \right. \quad \text{(B-6a)} \\ & \left. + \left| \frac{\text{Im}\{r_k^* S_i e^{j\theta_c}\}}{N_0} \right| + \frac{3\left|\text{Im}\{r_k^* S_i e^{j\theta_c}\}\right| \left|\text{Re}\{r_k^* S_i e^{j\theta_c}\}\right|}{N_0^2} \right] \\ &\leq \left| \frac{\left(\theta_m - \hat{\theta}_k\right)^3}{6(2\pi N_0)^{1/2}} \right| \left[\left| - \left(\frac{\text{Im}\{r_k^* S_i e^{j\theta_1}\}}{N_0}\right)^3 \right| \right. \quad \text{(B-6b)} \\ & \left. + \left| \frac{\text{Im}\{r_k^* S_i e^{j\theta_1}\}}{N_0} \right| + \frac{3\left|\text{Im}\{r_k^* S_i e^{j\theta_1}\}\right| \left|\text{Re}\{r_k^* S_i e^{j\theta_2}\}\right|}{N_0^2} \right] \end{aligned}$$

First, Triangle and Cauchy-Schwartz inequality [15] are applied to $|\tau_3|$ in order to obtain (B-6a). Equation (B-6a) is monotonically increasing for all values of θ_k. Hence we let θ_k to be upper-bounded by θ_m as in (B-4). Then, $\text{Im}\{r_k^* S_i e^{j\theta_c}\}$ is upper bounded by setting $\theta_c = \theta_1$, where $\theta_1 = \pi/2 - \arg\{r_k^* S_i\}$ as discussed before. Similarly, we have $\text{Re}\{r_k^* S_i e^{j\theta_c}\} = |r_k^* S_i| \cos(\theta_c + \arg\{r_k^* S_i\})$ and this becomes maximum when $\theta_c + \arg\{r_k^* S_i\} = 0$. Hence, we upper bound $\text{Re}\{r_k^* S_i e^{j\theta_c}\}$ by setting $\theta_c = \theta_2$, where $\theta_2 = -\arg\{r_k^* S_i\}$,

References

[1] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers, Synchronization, Channel Estimation, and Signal Processing.* Wiley, 1998.

[2] U. Mengali, and A. N. Andra, *Synchronization Techniques for Digital Receivers.* New York: Plenum Press, 1997.

[3] H. Kobayashi, "Simultaneous Adaptive Estimation and Decision Algorithm for Carrier Modulated Data Transmission Systems," *IEEE Trans. Commun. Tech.*, vol. 19, no. 3, pp. 268-280, Jun. 1971.

[4] D. Falconer, and J. Salz, "Optimal reception of digital data over the Gaussian channel with unknown delay and phase jitter," *IEEE Trans. Info. Theory*, vol. 23, no. 1, pp. 117- 126, Jan. 1977.

[5] P. Y. Kam, S. S. Ng, and T. S. Ng , "Optimum symbol-by-symbol detection of uncoded digital data over the Gaussian channel with unknown carrier phase," *IEEE Trans. Commun.*, vol. 42, no. 8, pp. 2543-2552, Aug. 1994.

[6] R. Raheli, A. Polydoros, and Ching-Kae Tzou, "Per-Survivor Processing: a general approach to MLSE in uncertain environments," *IEEE Trans. Commun.*, vol. 43, no. 234, pp. 354-364, Feb. Mar. Apr. 1995.

[7] S. J. Simmons, "Breadth-first trellis decoding with adaptive effort," *IEEE Trans. Commun.*, vol. 38, no. 1, pp. 3-12, Jan 1990.

[8] C. N. Georghiades and J. C. Han, "Sequence estimation in the presence of random parameters via the EM algorithm," *IEEE Trans. Commun.*, vol. 45, no. 3, pp. 300-308, Mar. 1997.

[9] C. H. Hsu, A. Anastasopoulos, "Design and analysis of joint data detection and frequency/phase estimation algorithms," *IEEE Journal on Selected Areas in Commun.*, vol. 23, no. 9, pp. 1707-1717, Sept. 2005.

[10] G. Colavolpe, A. Barbieri, and G. Caire, "Algorithms for iterative decoding in the presence of strong phase noise," *IEEE Journal on Selected Areas in Commun.*, vol. 23, no. 9, pp. 1748-1757, Sept. 2005

[11] Y. Li, P. Y. Kam and C. C. Chui, "Sequence detection for MPSK/MQAM with adaptive phase tracking", *Science Direct - Phys. Commun.*, vol. 2, no. 3, pp. 217-227, 2009.

[12] F. Simoens, D. Duyck, H. irpan, E. Panayirci, and M. Moeneclaey, "Monte Carlo Solutions for Blind Phase Noise Estimation," *EURASIP Journal on Wireless Commun. and Networking - Special Issue on Synchronization*, vol. 2009, pp. 1-11, Jan. 2009.

[13] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I.* New York: Wiley, 1968.

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

[15] W. Kaplan, *Advanced Calculus*. MA: Addison-Wesley, 1991.

[16] S. S. Yan, *Number Theory for Computing*, Germany: Springer, 2000.

[17] P. M. Heffernan, "Unbiased Estimation of Central Moments By Using U-Statistics," *Journal of the Royal Statistical Society, Series B*, vol. 59, no. 4, pp. 861-863, 1996.

[18] A. J. Viterbi, "Phase-locked loop dynamics in the presence of noise by Fokker-Planck techniques," *Proceedings of the IEEE*, vol. 51, no. 12, pp. 1737-1753, Dec. 1963.