Design and Analysis of Joint Data Detection and Frequency/Phase Estimation Algorithms

Chun-Hao Hsu *Student Member, IEEE,*

and Achilleas Anastasopoulos, *Member, IEEE*

Submitted: April 15, 2004; Revised December 5, 2004

Abstract

The problem of joint data detection and frequency/phase estimation is considered in this work. The traditional belief regarding *exact* generalized-likelihood-based joint detection and estimation is that its complexity is exponential in the sequence length N. This belief is justified due to the memory imposed on the transmitted sequence by the lack of knowledge of the auxiliary channel parameters. In this paper, we show that the exact solution can be performed with $O(N^4)$ worst-case complexity regardless of the operating signal-to-noise ratio. The concepts used in the proof of the polynomial complexity result are also utilized to evaluate tight performance bounds on the exact and a family of approximate algorithms.

I. Introduction

Coherent data communication requires perfect knowledge of the frequency and phase of the carrier signal. In practical communication systems, however, the mobility of the transmitter/receiver, in conjunction with the ambient and electronic noise at the receiver circuitry results in a time-varying phase that is unknown to the receiver. The presence of this frequency/phase jitter has multiple effects on the transmitted signal, the most severe of which is distortion of the transmitted signal and generation of inter-symbol interference (ISI) which becomes significant

This work was supported in part by the National Science Foundation under Grant CCF-0346977.

This paper was presented in part at the International Conference on Communications (ICC), Paris, France, June 2004.

The authors are with Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI 48109-2122, (e-mail: chhsu@umich.edu; anastas@umich.edu).

as the frequency/phase dynamics increase. Nevertheless, even when the channel dynamics are slow, the effect is a multiplicative phase distortion that can cause a significant performance loss, if not accounted for at the receiver.

There are three basic techniques (see [1] for a tutorial review) for detecting data in the presence of frequency/phase jitter: (i) a training sequence is transmitted periodically that aids at frequency and phase estimation, followed by coherent data detection; (ii) the received signal is passed through a memoryless nonlinearity (e.g., squaring binary phase shift keying (BPSK) signals,) that eliminates the data dependence, thus allowing channel estimation, followed by coherent data detection; and (iii) joint data detection and frequency/phase estimation. Clearly, the first two schemes do not exploit all the channel information available at the received sequence. As a result, the first two solutions are adequate for high signal-to-noise ratios (SNRs) and slow channel dynamics, but they perform very poorly (in terms of bandwidth efficiency or bit-error-rate (BER)) in more severe scenarios. Thus, when high-performance codes are employed over fast channels, some sort of joint detection and estimation needs to be performed.

Motivated by the desire to exploit the information from the whole sequence while keeping a low computational complexity, several joint detection/estimation algorithms have been proposed in the literature. These can be classified into the following two categories. The first category involves alternate maximization (or more formally, expectation maximization (EM)), where the tasks of channel-conditioned data detection and data-conditioned channel estimation are performed iteratively starting from an initial channel estimate, until convergence occurs [2]–[4]. The second category of algorithms is based on a suboptimal search over the set of all sequences by appropriately pruning the sequence tree according to a tree-pruning algorithm, such as the T-algorithm [5], the M-algorithm [6], or the per-survivor processing (PSP) algorithm [7]–[11]. It is noted that the above mentioned approaches are valid for any joint data detection and channel estimation problem and not only for the particular problem of joint data detection and frequency/phase estimation discussed in this paper. The impetus for the aforementioned research on approximate algorithms was the belief that the *exact* solution of the joint data detection and

parameter estimation problem has exponential complexity with respect to the sequence length.

Since future communication systems will operate close to their theoretical limits, joint data detection and parameter estimation will become indispensable in achieving the highest possible performance with the given resources, and thus, the following questions arise: (i) how accurate is the conventional wisdom that *exact* joint data detection and frequency/phase estimation requires exponential complexity with respect to the sequence length? (ii) what is the impact of a negative answer to the above question on the design of approximate algorithms? (iii) how can the complexity/performance tradeoff of these approximate algorithms be analyzed, and how can it be improved?

In this paper we continue the work initiated in [12] in trying to provide answers to the above questions. In particular, it is shown here that exact, generalized-likelihood-based, joint detection and frequency/phase jitter estimation of uncoded sequences is a polynomial-complexity problem in the sequence length. Furthermore, the proposed technique can be generalized to solve the problem of symbol-by-symbol soft-decision generation implied by the min-sum algorithm, which can be used when turbo-like coded sequences are utilized. In this paper, we concentrate in the uncoded sequence detection problem, since the extension to symbol-by-symbol soft-decision generation can be performed in a way similar to the one described in [12]. Furthermore, based on the proposed exact solution, we develop a class of approximate algorithms. Finally, a framework for analyzing the performance of both exact and approximate algorithms at arbitrary SNR is proposed. In the case of performance analysis for the exact algorithms, it is shown that analysis is possible exactly due to the novel polynomial-complexity structure. For the approximate algorithms, it is shown that their performance can get close to that of the exact algorithms with reasonable complexity.

The remaining of the paper is structured as follows. Section II develops the system and channel model under consideration. The polynomial-complexity exact algorithm for joint data detection and frequency/phase estimation is developed in Section III, after reviewing the corresponding results from [12]. Section IV presents the performance analysis for the exact, as well as several

approximate techniques. Numerical results are presented in Section V, while the conclusions are summarized in Section VI. For the sake of clarity, most of the proofs are relegated to an appendix at the end of the paper.

II. Channel Model

Consider the transmission of a length-N sequence of symbols $s_k \in \mathcal{A}$, where $\mathcal{A}$ is a set of complex-valued numbers with unit magnitude. The equivalent lowpass transmitted signal is of the form

$$s(t) = \sum_{k=1}^{N} s_k \sqrt{E_k} p(t - kT) \tag{1}$$

where $p(t)$ is a pulse shape function satisfying the no-ISI Nyquist criterion [13, Section 9.2.1], with unit energy, T is the symbol duration, and $E_k = E_s$ or E_p depending on whether s_k is an information symbol or pilot symbol, respectively. If this signal is transmitted over an additive white Gaussian noise (AWGN) channel and is further rotated by a phase process $\phi(t)$ unknown to both transmitter and receiver, then the received signal $z(t)$ can be modelled as[1]

$$z(t) = s(t)e^{j\phi(t)} + n(t) \tag{2}$$

where $n(t)$ is a zero mean complex white Gaussian process with one-sided power spectral density level N_0. For this observation model, optimal pre-processing depends on the phase process $\phi(t)$ and in the general case should include fractionally-spaced sampling. However, for the purpose of illustrating the basic ideas, we will assume symbol-spaced sampling, since all algorithms generalize easily to the fractionally-space model. We will further assume perfect epoch synchronization. Then, if the phase rotation process $\phi(t)$ is modelled by a constant unknown

[1]In this model, amplitude variations are ignored. This is done mainly in order to isolate the effects of phase rotation on the system performance. This assumption may be justified for channels where the phase rotation is due to imperfect local oscillators, and not due to user mobility. In the latter case, this model can be justified by the existence of an automatic gain control mechanism that compensates for the amplitude variations.

phase $\theta \in [0, 2\pi)$, i.e., if

$$\phi(t) = \theta \quad \forall t \in [0, NT), \tag{3}$$

then there will be no ISI and we have the following equivalent discrete-time model

$$\mathbf{z} = \mathbf{D}\mathbf{s}e^{j\theta} + \mathbf{n} \tag{4}$$

where $\mathbf{s} \triangleq [s_1, s_2, \ldots, s_N]^T$, $\mathbf{D}$ is a diagonal matrix with diagonal elements $(\sqrt{E_1}, \sqrt{E_2}, \ldots, \sqrt{E_N})$, $\mathbf{z} \triangleq [z_1, z_2, \ldots, z_N]^T$ is the vector of symbol-spaced observations, and $\mathbf{n} \triangleq [n_1, n_2, \ldots, n_N]^T$ is a vector of i.i.d. zero-mean circularly symmetric complex Gaussian random variables with variance $N_0/2$ per real an imaginary component. On the other hand, if the phase rotation process is modelled by

$$\phi(t) = 2\pi(f_d/T)t + \theta \quad \forall t \in [0, NT), \tag{5}$$

where $f_d \in [0, 1)$ is the normalized frequency jitter, and $\theta \in [0, 2\pi)$ is the phase shift, then as shown in [1] under the assumption of no ISI and perfect gain control, the discrete-time model becomes

$$z_k = s_k\sqrt{E_k}e^{j(2\pi f_d k+\theta)} + n_k \quad \forall k = 1, 2, \ldots, N. \tag{6}$$

III. Algorithms for Exact Generalized-likelihood Detection

For this section and the rest of the paper, we will restrict our attention to the special case where we use equally probable antipodal signaling, i.e., $\mathcal{A} = \{+1, -1\}$ for simplicity. Moreover, we will assume that the first one, and the first two symbols are pilot symbols for models (4), and (6), respectively. With a little abuse of notation, it is understood that all following detection algorithms are applied only to information symbols by setting beforehand the pilot symbols to $+1$. However, we note that all the algorithms and results presented in this paper can be generalized to arbitrary alphabet $\mathcal{A}$, unequal a-priori probabilities, arbitrary pilot symbol positioning, and energy assignments with a complexity increase by a factor $O(|\mathcal{A}|^2)$.

A. Background: The Constant Phase Model

In this subsection, we review the low complexity exact generalized-likelihood algorithm proposed in [12], [14], [15] for completeness and for the purpose of showing the basic idea from which we develop the exact algorithm for the more complicated model (6). The sequence estimate based on the generalized-likelihood ratio test (GLRT) for model (4) can be written in the following double maximization form

$$\hat{\mathbf{s}}_{GLRT} \triangleq \arg\max_{\tilde{\mathbf{s}}\in\mathcal{A}^N}\left\{\max_{\tilde{\theta}\in[0,2\pi)} p(\mathbf{z}|\tilde{\mathbf{s}},\tilde{\theta})\right\} = \arg\max_{\tilde{\mathbf{s}}\in\mathcal{A}^N}\left\{\max_{\tilde{\theta}\in[0,2\pi)} \Re\{\mathbf{z}^H\mathbf{D}\tilde{\mathbf{s}}e^{j\tilde{\theta}}\}\right\} \tag{7a}$$

$$= \arg\max_{\tilde{\mathbf{s}}\in\mathcal{A}^N} |\mathbf{z}^H\mathbf{D}\tilde{\mathbf{s}}|. \tag{7b}$$

It is well known that, under the assumption that the constant phase rotation θ is uniformly distributed in $[0, 2\pi)$, the GLRT solution coincides with the maximum likelihood sequence detection (MLSD) solution. Furthermore, this optimal solution can be obtained exactly with $O(N \log N)$ complexity as follows (for details on the correctness of the algorithm, refer to [12]).

1) Calculate the set of partitioning points (or thresholds) from the received signal $\mathbf{z}$ as follows

$$\Phi \triangleq \{\phi_i \in [0, 2\pi) : \phi_i \triangleq \angle z_i \pm \frac{\pi}{2}, \quad \forall i = 2, 3, \ldots, N\} \tag{8}$$

2) Sort Φ so that $\Phi = \{0 \le t_1 \le t_2 \le \ldots \le t_{2(N-1)} \le 2\pi\}$. We use $j(i)$ to denote the index of the observation symbol $z_{j(i)}$ associated with t_i.

3) Find the candidate sequence supported by the first partition, which contains phase 0,

$$\mathbf{s}_1 = \hat{\mathbf{s}}(0) \tag{9}$$

where

$$\hat{\mathbf{s}}(\tilde{\theta}) \triangleq \arg\max_{\tilde{\mathbf{s}}\in\mathcal{A}^N} \Re\{\mathbf{z}^H\mathbf{D}\tilde{\mathbf{s}}e^{j\tilde{\theta}}\} \tag{10}$$

4) Sequentially evaluate the candidate sequence $\mathbf{s}_{i+1}$ from $\mathbf{s}_i$ by flipping the $j(i)$th bit of $\mathbf{s}_i$.

5) As explicitly proved in [12], $\hat{\mathbf{s}}_{GLRT}$ must be one of the candidate sequences $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{2(N-1)}$. Hence we have

$$\hat{\mathbf{s}}_{GLRT} = \arg\max_{\tilde{\mathbf{s}}\in\{\mathbf{s}_1,\mathbf{s}_2,\ldots,\mathbf{s}_{2(N-1)}\}} \Re\{\mathbf{z}^H\mathbf{D}\tilde{\mathbf{s}}e^{j\hat{\theta}(\tilde{\mathbf{s}})}\} = \arg\max_{\tilde{\mathbf{s}}\in\{\mathbf{s}_1,\mathbf{s}_2,\ldots,\mathbf{s}_{2(N-1)}\}} |\mathbf{z}^H\mathbf{D}\tilde{\mathbf{s}}| \tag{11}$$

where

$$\hat{\theta}(\tilde{\mathbf{s}}) \triangleq \arg \max_{\tilde{\theta} \in [0,2\pi)} \Re\{\mathbf{z}^H \mathbf{D}\tilde{\mathbf{s}} e^{j\tilde{\theta}}\} = -\angle(\mathbf{z}^H \mathbf{D}\tilde{\mathbf{s}}) \tag{12}$$

The actual algorithm in [12], [14], [15] further simplifies the metric evaluation in (11), so that the overall complexity is determined by the sorting in step 2) which can be performed in $O(N \log N)$ time. Note that the basic idea behind the algorithm is to partition the parameter space $[0, 2\pi)$ in such a way that performing coherent sequence detection with one hypothesized parameter taken from each set of the partition provides a sufficient set of candidate sequences for the GLRT problem.

B. The Linear Phase Model

Consider the model in (6). The GLRT solution is given by

$$\hat{\mathbf{s}}_{GLRT} \triangleq \arg \max_{\tilde{\mathbf{s}}} \left\{ \max_{(\tilde{f}_d, \tilde{\theta}) \in \Lambda} p(\mathbf{z}|\tilde{\mathbf{s}}, \tilde{f}_d, \tilde{\theta}) \right\} \tag{13}$$

where $\Lambda \triangleq [0, 1) \times [0, 2\pi)$ is the 2-dimensional parameter space over which maximization of the unknown parameters f_d and θ is performed for each possible transmitted sequence. In the rest of the section, we will aim at finding this exact GLRT solution with polynomial complexity.

Defining

$$\hat{f}_d(\tilde{\mathbf{s}}) \triangleq \arg \max_{\tilde{f}_d} \left\{ \max_{\tilde{\theta}} p(\mathbf{z}|\tilde{\mathbf{s}}, \tilde{f}_d, \tilde{\theta}) \right\} = \arg \max_{\tilde{f}_d} \left| \sum_k z_k^* \tilde{s}_k \sqrt{E_k} e^{j2\pi \tilde{f}_d k} \right| \tag{14a}$$

and

$$\hat{\theta}(\tilde{\mathbf{s}}) \triangleq \arg \max_{\tilde{\theta}} \left\{ \max_{\tilde{f}_d} p(\mathbf{z}|\tilde{\mathbf{s}}, \tilde{f}_d, \tilde{\theta}) \right\} = -\angle \left(\sum_k z_k^* \tilde{s}_k \sqrt{E_k} e^{j2\pi \hat{f}_d(\tilde{\mathbf{s}}) k} \right) \tag{14b}$$

to be the sequence-conditioned phase/frequency estimates, we have

$$\hat{\mathbf{s}}_{GLRT} = \arg \max_{\tilde{\mathbf{s}} \in \mathcal{A}^N} p(\mathbf{z}|\tilde{\mathbf{s}}, \hat{f}_d(\tilde{\mathbf{s}}), \hat{\theta}(\tilde{\mathbf{s}})) \tag{15}$$

Looking at (15) one can make the following observation. There are two sources of complexity in finding the GLRT solution. The first one is combinatorial in nature and is related to the exponential growth of sequences (or metrics) that need to be maximized. The second source of complexity

is computational in nature and is related to the evaluation of each metric $p(\mathbf{z}|\tilde{\mathbf{s}}, \hat{f}_d(\tilde{\mathbf{s}}), \hat{\theta}(\tilde{\mathbf{s}}))$ for a given hypothesized sequence. This latter complexity is linear in N for the constant phase model as evidenced in (7b). It should be clear however, that there is no hope in finding a closed-form solution for the metric $p(\mathbf{z}|\tilde{\mathbf{s}}, \hat{f}_d(\tilde{\mathbf{s}}), \hat{\theta}(\tilde{\mathbf{s}}))$, since this is equivalent to finding a closed-form solution to the sequence-conditioned frequency estimate $\hat{f}_d(\tilde{\mathbf{s}})$ in (14a). But this, in turn, involves maximization of the magnitude of the discrete-time Fourier transform (DTFT) of the sequence $\{z_k \tilde{s}_k^* \sqrt{E_k}\}_{k=1}^N$ over the continuous interval $[0, 1)$ [1]. Therefore, frequency estimation can only be performed within a pre-specified accuracy with finite complexity in all practical algorithms. Since this complexity is unavoidable even when the data sequence is known, we are only interested in the first source of complexity manifesting itself through the combinatorial explosion of the data sequences. As can be seen in (15), even if we omit the frequency-estimation complexity, the exact GLRT solution still appears to demand $O(2^N)$ complexity; we will now propose an exact algorithm, which performs this task with $O(N^4)$ complexity regardless of the SNR.

Defining a parameter-conditioned sequence estimate

$$\hat{\mathbf{s}}(\tilde{f}_d, \tilde{\theta}) \triangleq \arg \max_{\tilde{\mathbf{s}} \in \mathcal{A}^N} p(\mathbf{z}|\tilde{\mathbf{s}}, \tilde{f}_d, \tilde{\theta}), \tag{16}$$

the GLRT problem can be restated as

$$\hat{\mathbf{s}}_{GLRT} \triangleq \max_{(\tilde{f}_d, \tilde{\theta}) \in \Lambda} p(\mathbf{z}|\hat{\mathbf{s}}(\tilde{f}_d, \tilde{\theta}), \tilde{f}_d, \tilde{\theta}). \tag{17}$$

Furthermore, collecting all candidate GLRT solutions in the sufficient set

$$\mathcal{T} \triangleq \{\hat{\mathbf{s}}(\tilde{f}_d, \tilde{\theta}) | (\tilde{f}_d, \tilde{\theta}) \in \Lambda\}, \tag{18}$$

the GLRT problem becomes

$$\hat{\mathbf{s}}_{GLRT} = \arg \max_{\tilde{\mathbf{s}} \in \mathcal{T}} p(\mathbf{z}|\tilde{\mathbf{s}}, \hat{f}_d(\tilde{\mathbf{s}}), \hat{\theta}(\tilde{\mathbf{s}})). \tag{19}$$

It is now clear how one should proceed to prove the polynomial-complexity result: if the size of $\mathcal{T}$ grows only polynomially with N, and $\mathcal{T}$ can be constructed by a polynomial-complexity algorithm, then using (19) one can solve the problem with polynomial combinatorial complexity.

As shown in [12] for a general GLRT problem, constructing the sufficient set $\mathcal{T}$ is equivalent to partitioning the parameter space Λ into subsets in such a way that all parameter pairs $(\tilde{f}_d, \tilde{\theta})$ in each subset result in the same sequence $\hat{\mathrm{s}}(\tilde{f}_d, \tilde{\theta})$. It was further shown that this partitioning can be accomplished by superimposing the boundaries defined by equations of the form

$$\begin{aligned} |z_k - \sqrt{E_k}(+1)e^{j(2\pi\tilde{f}_d k+\tilde{\theta})}|^2 &= |z_k - \sqrt{E_k}(-1)e^{j(2\pi\tilde{f}_d k+\tilde{\theta})}|^2 \\ \Leftrightarrow 2\pi\tilde{f}_d k + \tilde{\theta} &= \angle z_k \pm \frac{\pi}{2} + 2\pi m, \end{aligned} \tag{20}$$

for all $k = 1, 2, \ldots, N$, and all integers m. An example of the resulting partitioned parameter space is shown in Fig. 1.

By observing that there are $2k$ lines for each k, we conclude that there are $O(N^2)$ lines in total implied by (20). Using a well-known result from computational geometry [16], one can deduce that at most $|\mathcal{T}| = O(N^4)$ polygons can be generated by superimposing $O(N^2)$ lines and that there is an $O(N^4)$-complexity algorithm that can construct these polygons, and thus the sufficient set $\mathcal{T}$. We further improve this estimate by using the additional information that every line with $k = i$ intersects at most one line with $k = j, j \neq i$, and is parallel to all lines with $k = i$. Indeed, due to the above observation, it is straightforward to show that there are at most $|\mathcal{T}| = O(N^3)$ partitions resulted from these lines since each line intersects at most $N - 1$ other lines. Moreover, to compute all intersection points of each additional line at $k = i$ with all previously drawn lines, the following algorithm can be used.

1) Find the entrance intersection point on the polygon adjacent to the left or upper boundary of the parameter space Λ. It is sufficient to look only at the left and upper boundary, because all lines have negative slopes. The complexity of this step is $O(i^2)$ for all $2i$ lines.
2) Find the exit intersection point among all sides of this polygon. Note that the number of sides per polygon is constant on the average, resulting in constant complexity for this step.
3) Identify the next polygon (constant complexity with proper indexing). If there is no next polygon, i.e., the exit point hits the lower or right boundary, then quit. Otherwise, go to step 2.

Since there are $O(N^2)$ lines in total, and each line visits at most N polygons as shown above, we have shown that the size of the partition and its construction complexity is $O(N^3)$.

After constructing all $O(N^3)$ polygons, we take one parameter pair (f_d, θ) from each polygon and do coherent symbol by symbol detection hypothesizing (f_d, θ) as the CSI. The resulting $O(N^3)$ sequences from all polygons then constitute the the sufficient set $\mathcal{T}$. Finally, the frequency estimate and the noncoherent metric is evaluated for each of these sequences and is maximized to find the GLRT solution according to

$$\hat{\mathbf{s}}_{GLRT} = \arg \max_{\tilde{\mathbf{s}} \in \mathcal{T}} \left| \sum_{k=1}^{N} z_k^* \tilde{s}_k \sqrt{E_k} e^{j2\pi \hat{f}_d(\tilde{\mathbf{s}})k} \right|. \tag{21}$$

Since these two final steps require $O(N)$ complexity for each polygon and since we have $O(N^3)$ polygons, we have verified that the whole algorithm finds the exact GLRT solution with $O(N^4)$ complexity. We summarize the above discussion in the following proposition.

Proposition 1 *The proposed algorithm in this subsection finds the exact GLRT solution with worst case complexity $O(N^4)$ when omitting the complexity of the frequency estimator, where N is the sequence length.*

Proof: Follows from the above discussion and the framework developed in [12]. ■

IV. Performance Analysis of Exact and Approximate Algorithms

The purpose of this section is to demonstrate that the polynomial-complexity results developed above, aside from their conceptual value, are also useful in at least two additional respects. The first is that they enable accurate performance analysis of the corresponding exact GLRT algorithms, which would otherwise be impossible or result in loose performance bounds. The second is that they imply a family of approximate algorithms that can be easily implemented and analyzed. In the following, the exact and two approximate algorithms are analyzed for the constant phase model and one approximate algorithm is analyzed for the linear phase model.

A. Constant Phase Model: Exact GLRT Algorithm

In this subsection, we present a performance analysis for the polynomial-complexity exact GLRT algorithm for the model in (4). The resulting performance upper bound turns out to be extremely tight as can be seen in section V, and is now given by the following proposition.

Proposition 2 *Under the assumptions of the model in* (4)*, the sequence error probability for the exact GLRT solution (which is the same as the optimal MLSD solution), can be upper bounded by*

$$\begin{aligned} P_{MLSD} \leq & 2\sum_{w=1}^{N-2}\int_0^\infty R(r,E_t,\frac{E_tN_0}{2})Q_1(\frac{r|E_t-2wE_s|}{\sqrt{2wE_tE_s(E_t-wE_s)N_0}},\frac{r\sqrt{E_t}}{\sqrt{2wE_s(E_t-wE_s)N_0}})dr \\ & + \int_0^\infty R(r,E_t,\frac{E_tN_0}{2})Q_1(\frac{r|E_p-(N-1)E_s|}{\sqrt{2(N-1)E_tE_sE_pN_0}},\frac{r\sqrt{E_t}}{\sqrt{2(N-1)E_sE_pN_0}})dr \\ & + 1-(1-Q(\sqrt{\frac{2E_s}{N_0}}))^{N-1}, \end{aligned} \tag{22a}$$

where

$$E_t \triangleq E_p + (N-1)E_s \tag{22b}$$

denotes the total signalling energy,

$$R(r,s,\sigma^2) \triangleq \frac{r}{\sigma^2}e^{-\frac{r^2+s^2}{2\sigma^2}}I_0(\frac{rs}{\sigma^2}) \tag{22c}$$

is a Ricean density with $I_0(\cdot)$ being the 0th-order modified Bessel function of the first kind, and $Q_1(\cdot,\cdot)$, $Q(\cdot)$ are the the Marcum's Q function, and the Gaussian tail function, respectively.

Proof: See Appendix A. ■

The main idea of the proof is to use union bound over a much smaller sufficient set of sequences rather than all the 2^N sequences. This can be done only because the set of all possible MLSD solutions had been tremendously reduced to have a size linear in N as was shown in the previous section.

B. Constant Phase Model: Pilot-Only (PO) Algorithm

It is possibly the simplest approximate algorithm, which uses the channel information provided by the pilot symbol only, and performs symbol-by-symbol detection as follows

$$\hat{s}_i = \arg\max_{\tilde{s}\in\mathcal{A}} p(z_i|\tilde{s}, z_1) \quad i = 2, \ldots, N, \tag{23}$$

This pilot-only (PO) algorithm has complexity $O(N)$, and its performance is given in the following proposition.

Proposition 3 *The probability of bit error of the PO algorithm is exactly*

$$P_b(PO) = 1 - \int_0^{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} T(y+x, \frac{E_s}{N_0}) T(x, \frac{E_p}{N_0}) dy dx \tag{24}$$

where

$$T(x, S) \triangleq \frac{1}{2\pi} e^{-S} + \sqrt{\frac{S}{\pi}} \cos(x) e^{-S\sin^2(x)} \left(1 - Q(\sqrt{2S}\cos(x))\right) \tag{25}$$

is the probability density function (pdf) of the phase component of the complex Gaussian random variable with mean $\sqrt{S}$ and variance 1. Therefore, the corresponding probability of sequence error is

$$P_{PO} = 1 - \left(\int_0^{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} T(y+x, \frac{E_s}{N_0}) T(x, \frac{E_p}{N_0}) dy dx \right)^{N-1} \tag{26}$$

Proof: See Appendix B. ■

It is worth noting that even if we drop the magnitude information of z_1 and condition only on $\angle z_1$ in (23), we obtain the same performance in this uncoded sequence case. However, if we are also interested in the soft decisions of the receiver as required in the coded sequence case, then the magnitude information provided by the pilot symbol would help improve the performance as shown in [17].

C. Constant Phase Model: Uniform Sampling (US) Algorithm

Motivated by the exact algorithm described in Section IV-A, we now consider a simple approximate analogue of it and analyze its performance. The idea behind this algorithm is that instead of optimally partitioning the parameter space $[0, 2\pi)$–a process that has complexity $O(N \log N)$–one can partition the parameter space in a sensible, but somewhat arbitrary way, in order to reduce complexity. The simplest such partitioning is the uniform partitioning (or sampling), which results in the following algorithm.

1) Define L samples $\{a_i\}_{i=1}^{L}$ as

$$a_i \triangleq \frac{2\pi i}{L}, \quad \forall\ i = 1, 2, \ldots, L \tag{27}$$

2) Obtain L candidate sequences $\hat{\mathbf{s}}(a_i)$ for all i and select

$$\hat{\mathbf{s}}_{US} = \arg \max_{\hat{\mathbf{s}}(a_i): i \in \{1,2,\ldots,L\}} |\mathbf{z}^H \mathbf{D} \hat{\mathbf{s}}(a_i)| \tag{28}$$

This algorithm has complexity $O(LN)$ and a characterization of its performance is given by the following proposition.

Proposition 4 *Under the assumption that $L \geq 2$, the probability of sequence error for the US algorithm can be bounded by*

$$\begin{aligned} P_{MLSD} \leq P_{US} \leq & P_{MLSD} + (N-1)(N-2) \\ & \int_0^\infty R(r, E_t, \frac{E_t N_0}{2}) \left[\int_{-\frac{\pi}{L}}^{\frac{\pi}{L}} T(\theta - \frac{\pi}{2}, \frac{E_s}{E_t(E_t - E_s)N_0} r^2) d\theta \right]^2 dr \qquad (29a) \\ = & P_{MLSD} + O(\frac{N^2}{L^2} e^{-\frac{E_s}{N_0}}) \qquad (29b) \end{aligned}$$

Proof: See Appendix C. ■

The main idea behind the proof can again be attributed to the parameter space partitioning perspective. The event that we miss the MLSD solution is exactly the event that none of our sampling points is in the same partition with the MLSD solution. It is the probability of this event that is characterized and upper bounded in the proof.

Equation (29a) can be numerically evaluated to provide a measure of how far the performance of the US algorithm is from that of the MLSD algorithm. Also, the asymptotic expression in (29b) reveals some interesting insight. First, for high SNR, the performance of this approximate algorithm converges exponentially to that of the exact algorithm for any N and $L \geq 2$. Moreover, for any fixed SNR, if L increases slightly faster than N (e.g., as $N^{1+\epsilon}$, for an arbitrarily small positive ϵ), then the performance of this approximate algorithm approaches that of the MLSD algorithm for large N. As will be shown in Section V, this latter bound is indeed pessimistic: a small L is enough to make the performance of the US algorithm close to that of the MLSD algorithm for all SNR values of interest.

D. Linear phase Model: 2-Dimensional Uniform Sampling (US2D) Algorithm

Motivated by the good performance of the US algorithm for the constant phase case, we propose a similar approximate algorithm for the more complicated linear phase case as follows.

1) Define $Q_f Q_\theta$ samples of frequency-phase pairs as

$$(a_i, b_j) \triangleq (\frac{2i}{Q_f}, \frac{2\pi j}{Q_\theta}), \forall i = 1, \ldots, Q_f, j = 1, \ldots, Q_\theta \tag{30}$$

2) Obtain $Q_f Q_\theta$ candidate sequences $\hat{\mathbf{s}}(a_i, b_j)$ and the corresponding frequency estimates $f_{i,j} = \hat{f}_d(\hat{\mathbf{s}}(a_i, b_j))$ for all i, j , and select

$$\hat{\mathbf{s}}_{US2D} = \arg \max_{\substack{\hat{\mathbf{s}}(a_i,b_j) \\ i\in\{1,2,\ldots,Q_f\} \\ j\in\{1,2,\ldots,Q_\theta\}}} \left| \sum_{k=1}^{N} z_k^* s_k(a_i, b_j) \sqrt{E_k} e^{j2\pi f_{i,j} k} \right| \tag{31}$$

This algorithm has complexity $O(Q_f Q_\theta N)$ and its performance is given by the following proposition.

Proposition 5 *Under the assumption that $Q_f \geq 4N$ and $Q_\theta \geq 4$ and for any f_d and θ, the*

probability of sequence error for the US2D algorithm can be bounded by

$$P_{CSI} \leq P_{US2D} \leq P_{CSI} + P_{GLRT} + \{1 - 2\prod_{k=1}^{N}(1 - ql(k)) + \prod_{k=1}^{N}(1 - 2ql(k))\}$$

$$+ \left\{1 - 2(1 - qu)^N + (1 - 2qu)^N\right\} \tag{32a}$$

$$= P_{CSI} + P_{GLRT} + O(\frac{N^2}{Q_f} e^{-\frac{E_s}{N_0}}) + O(\frac{N}{Q_\theta} e^{-\frac{E_s}{N_0}}) \tag{32b}$$

where P_{CSI} and P_{GLRT} are the probability of error for the perfect channel state information (CSI) and the exact GLRT algorithms, respectively, and

$$ql(k) = \int_0^{\frac{2\pi k}{Q_f}} T(x - \frac{\pi}{2}, \frac{E_s}{N_0}) + T(x + \frac{\pi}{2}, \frac{E_s}{N_0}) dx \tag{33a}$$

$$qu = \int_0^{\frac{2\pi}{Q_\theta}} T(x - \frac{\pi}{2}, \frac{E_s}{N_0}) + T(x + \frac{\pi}{2}, \frac{E_s}{N_0}) dx. \tag{33b}$$

Proof: See Appendix D. ■

This result is very similar in spirit to Proposition 4. One difference is that because of the additional frequency jitter, we may need to increase Q_f quadratically with N in order to have a performance close to that of the exact GLRT algorithm. Another difference is that in this case, the upper and lower bounds do not agree even for large Q_f and Q_θ.

V. Numerical results

In this section numerical results are presented for the exact and approximate algorithms developed earlier for both the constant and the linear phase models.

A. The Constant Phase Model

Fig. 2 compares the performance bounds for the three algorithms (where the expression for the PO algorithm is exact) developed in Section IV for model (4) with optimally chosen pilot energies versus the information bit SNR, $E_b/N_0 \stackrel{\text{def}}{=} (E_t/N_0)/(N-1)/\log_2|\mathcal{A}|$. Also shown is the P_{CSI} performance curve, and a "naive" MLSD bound obtained by a union bound over all 2^N sequences. As can be seen in the figure, the upper bound of the US algorithm, which is

obtained by substituting the MLSD bound into (29a), performs almost identically to the upper bound of the exact MLSD algorithm, and is very close to the performance curve of the perfect CSI receiver. With a moderate choice of $L = 8$ in the approximate algorithm, it performs equally well for both $N = 4$ and $N = 32$ cases. As compared with the PO algorithm, we gain 2 dB when $N = 4$ and about 0.5 dB when $N = 32$ by using the US algorithm. Note also that the upper bound for the MLSD algorithm is actually very tight since the CSI performance is a lower bound.

B. The Linear Phase Model

In this subsection we provide various simulation results regarding the implementation of the US2D algorithm. For simplicity, we use $E_p = E_s$ instead of optimizing over E_p throughout our simulations.

One way to perform frequency estimation for a given sequence, is to zero-pad the sequence with D zeros and perform an $(N + D)$-point FFT to find the frequency component with the maximum magnitude. This entails $O((N+D)\log(N+D))$ complexity per sequence, which can be undesirable in the US2D algorithm. Since the US2D algorithm is aiming at low complexity, we adopt the simplest version of the Luise and Reggiannini (L&R) estimator [18] for this algorithm, which evaluates a frequency estimate as

$$\hat{f}_d(\mathbf{s}) = \angle \sum_{i=1}^{N-1} (z_{i+1} s_{i+1}^*)(z_i s_i^*)^*, \tag{34}$$

thus having linear complexity in N. Note that in general, the L&R estimator can use more autocorrelation coefficients than just the first one as in this case. As can be seen in Fig. 3, the considerably much faster L&R estimator suffers only about 1 dB performance loss (at BER=10^{-3}) compared to the higher accuracy FFT estimators. Therefore, in all the following simulations, we will employ this L&R frequency estimator for the US2D algorithm.

To examine the effect of Q_θ in the US2D algorithm, we have nullified the effect of Q_f by using a large Q_f such that no performance can be gained by further increasing this Q_f. Fig. 4

shows the performances of the US2D algorithm for a large enough Q_f, different values of Q_θ, and different M-PSK alphabets. As can be seen in the figure, very little improvement can be gained by selecting $Q_\theta > M$ regardless of N (Similar simulations for $N = 32$ were performed, resulting in the same conclusion). This suggests that Proposition 5, which shows that performance behaves as $O(\frac{N}{Q_\theta})$, provides a conservative estimate.

Due to the above observation, we fix $Q_\theta = M$ in the US2D algorithm and investigate how Q_f affects performance. Fig. 5 clearly shows that to achieve the best performance, Q_f should grow approximately quadratically with N as predicted in Proposition 5. This result for Q_f together with the previous one for Q_θ suggests that it is possible to employ an $O(N^3)$ complexity US2D algorithm to achieve a near-exact performance.

Fig. 6 shows that the US2D algorithm with a fast L&R frequency estimator can indeed achieve a performance close to that of the exact GLRT algorithm, which employs a very accurate FFT-based frequency estimator, by choosing a moderate Q_f and Q_θ. Notice that the number of frequency samples was increased quadratically with N. Also shown is the performance of an ad-hoc pilot-only algorithm, which first estimates f_d and θ using the first two pilots and then does symbol-by-symbol detection assuming the estimated CSI. It is obvious that for this linear phase model, it is much more desirable to extract the channel information from the whole received sequence by performing joint detection/estimation schemes rather than restricting our attention only to the channel information provided by pilots. Also shown in Fig. 6 are the upper bounds derived in Proposition 5. However, the parameters used for obtaining these bounds were $Q_\theta = 8$ and $Q_f = 8N$ which are different from those used to obtain the simulated performance. This was due to the fact that the bound is valid for $Q_\theta \geq 4$ and $Q_f \geq 4N$ as mentioned in Proposition 5.

VI. Conclusion and Discussion

Several low complexity joint detection/estimation algorithms for noncoherent channels with an unknown phase rotation are presented and analyzed in this paper. If the phase process is modelled by a constant phase shift, then we showed that the low complexity US algorithm can be applied, which yields a performance close to that of the ML decoder with perfect CSI. In the

case that the phase process is modelled by both a frequency jitter and a phase shift, we showed that the exact GLRT solution can be obtained with combinatorial complexity $O(N^4)$ or well approximated by the proposed US2D algorithm with complexity $O(Q_f Q_\theta N)$.

When powerful codes are utilized, symbol by symbol soft decisions will be desired by the decoder for iterative decoding. The proposed algorithms can also be adapted to provide this information. It is very similar to the one discussed in [12], and the additional manipulation will only increase the whole complexity of the algorithms by a factor of N.

We conclude by noting that the basic idea behind the exact algorithms is to think of decision regions in the parameter space rather than in the observation space as is the traditional approach. This is helpful since in the discussed problems, the dimension of the observation space grows with N, while that of the parameter space remains fixed. It is thus expected that similar results will hold every time the number of independent parameters that need to be estimated grows slower than N.

APPENDIX

A. *Proof of Proposition 2*

Let $\mathbf{s}_0, \hat{\mathbf{s}}_{CSI}$ be the transmitted sequence and the detected sequence for the hypothetical receiver that has perfect CSI. We have

$$P_{MLSD} = Pr(\hat{\mathbf{s}}_{MLSD} \neq \mathbf{s}_0) = Pr(\hat{\mathbf{s}}_{MLSD} \neq \mathbf{s}_0, \hat{\mathbf{s}}_{CSI} = \mathbf{s}_0) + Pr(\hat{\mathbf{s}}_{MLSD} \neq \mathbf{s}_0, \hat{\mathbf{s}}_{CSI} \neq \mathbf{s}_0)$$

$$\leq Pr(\{\hat{\mathbf{s}}_{MLSD} \neq \mathbf{s}_0\} \cap \{\mathbf{s}_0 = \mathbf{s}_j \text{ for some } 1 \leq j \leq 2(N-1)\}) + Pr(\hat{\mathbf{s}}_{CSI} \neq \mathbf{s}_0)$$

$$\leq Pr\left(\bigcup_{\substack{1 \leq i \leq 2(N-1), \\ i \neq j}} \{|\mathbf{z}^H \mathbf{D}\mathbf{s}_0| < |\mathbf{z}^H \mathbf{D}\mathbf{s}_i|\} \right) + 1 - (1 - Q(\sqrt{\frac{2E_s}{N_0}}))^{N-1} \tag{35}$$

where $\mathbf{s}_i, \forall i \neq 0$ are the candidate sequences as defined in the exact GLRT algorithm. To further evaluate the first term of (35), we define $X \triangleq \mathbf{z}^H \mathbf{D}\mathbf{s}_0$ and $Y_i \triangleq \mathbf{z}^H \mathbf{D}\mathbf{s}_i$ for $1 \leq i \leq 2(N-1)$. Since X and $Y_i's$ are jointly Gaussian, we have

$$p(Y_i|X) = \mathcal{CN}(Y_i, \frac{E_t - 2w_i E_s}{E_t} X, \frac{4w_i E_s (E_t - w_i E_s) N_0}{E_t}) \tag{36}$$

where $\mathcal{CN}(\cdot, m, \sigma^2)$ is the pdf of a complex Gaussian random variable with mean m and variance σ^2, and w_i is the number of places where $\mathbf{s}_0$ and $\mathbf{s}_i$ differ. Therefore, we have

$$Pr\left(\left|\mathbf{z}^H\mathbf{D}\mathbf{s}_0\right| < \left|\mathbf{z}^H\mathbf{D}\mathbf{s}_i\right|\right) = \int_{\mathcal{C}} Pr(|Y_i| > |x||X = x) f_X(x) dx \tag{37}$$

$$= \int_{\mathcal{C}} \int_{|x|}^{\infty} R(r, \left|\frac{E_t - 2w_iE_s}{E_t}\right| |x|, \frac{2w_iE_s(E_t - w_iE_s)N_0}{E_t}) f_X(x) dr dx$$

$$= \int_0^{\infty} R(r, E_t, \frac{E_tN_0}{2}) Q_1(\frac{r|E_t - 2w_iE_s|}{\sqrt{2w_iE_tE_s(E_t - w_iE_s)N_0}}, \frac{r\sqrt{E_t}}{\sqrt{2w_iE_s(E_t - w_iE_s)N_0}}) dr \tag{38}$$

where $\mathcal{C}$ is the set of all complex numbers and f_X is the pdf of X. Hence by taking the union bound in (35) and use the fact that candidate sequences corresponding to neighboring partitions of $[0, 2\pi)$ differ in only one symbol, which follows from the structure of the exact polynomial-complexity algorithm, we obtain (22a).

B. Proof of Proposition 3

Letting $z = \sqrt{E_s} + n$ where $n \sim \mathcal{CN}(0, N_0)$, $r = |z|$ and $t = \angle z - \angle z_1$, its bit error probability can be evaluated as follows:

$$\begin{aligned} P_b(PO) =& Pr\left(\int_0^{2\pi} e^{\frac{2\Re\{z\sqrt{E_s}e^{-j\theta}\}}{N_0}} e^{\frac{2|z_1|\sqrt{E_p}\cos(\angle z_1 - \theta)}{N_0}} d\theta < \int_0^{2\pi} e^{\frac{-2\Re\{z\sqrt{E_s}e^{-j\theta}\}}{N_0}} e^{\frac{2|z_1|\sqrt{E_p}\cos(\angle z_1 - \theta)}{N_0}} d\theta\right) \\ =& Pr\left(\int_0^{2\pi} \left[e^{\frac{2r\sqrt{E_s}\cos(\theta - \angle z)}{N_0}} - e^{\frac{-2r\sqrt{E_s}\cos(\theta - \angle z)}{N_0}}\right] e^{\frac{2|z_1|\sqrt{E_p}\cos(\angle z_1 - \theta)}{N_0}} d\theta < 0\right) \\ =& Pr\left(\int_0^{2\pi} \sinh(\frac{2r\sqrt{E_s}\cos(\theta - t)}{N_0}) e^{\frac{2|z_1|\sqrt{E_p}\cos(\angle z_1 - \theta)}{N_0}} d\theta < 0\right) \end{aligned} \tag{39}$$

Since $\sinh(\frac{2r\sqrt{E_s}\cos(\theta)}{N_0})$ and $e^{\frac{2|z_1|\sqrt{E_p}\cos(\angle z_1 - \theta)}{N_0}}$ are both symmetric bell shaped functions of θ in one period $[-\pi, \pi]$, their circular convolution would still be a bell shaped function. Observing further that $\int_0^{2\pi} \sinh(\frac{2r\sqrt{E_s}\cos(\theta - t)}{N_0}) e^{\frac{2|z_1|\sqrt{E_p}\cos(\angle z_1 - \theta)}{N_0}} d\theta$ attains maximum at $t = 0$, minimum at

$t=\pi$ and 0 at $t=\pm\frac{\pi}{2}$, we have

$$\begin{aligned}
P_b(PO) =& 1-p(t\in[-\frac{\pi}{2},\frac{\pi}{2}]) \\
=& 1-E_\theta\left[\int_0^{2\pi} p(\angle z\in[-\frac{\pi}{2}+\angle z_1,\frac{\pi}{2}+\angle z_1]|\theta)f(\angle z_1|\theta)d\angle z_1\right] \\
=& 1-E_\theta\left[\int_0^{2\pi}\int_{-\frac{\pi}{2}+x}^{\frac{\pi}{2}+x} T(y-\theta,\frac{E_s}{N_0})dyT(x-\theta,\frac{E_p}{N_0})dx\right] \\
=& 1-E_\theta\left[\int_0^{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} T(y+x,\frac{E_s}{N_0})T(x,\frac{E_p}{N_0})dydx\right] \\
=& 1-\int_0^{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} T(y+x,\frac{E_s}{N_0})T(x,\frac{E_p}{N_0})dydx
\end{aligned} \tag{40}$$

It then follows that $P_{PO}=1-(1-P_b(PO))^{N-1}$.

C. Proof of Proposition 4

The lower bound follows easily by the optimality of the MLSD metric. The upper bound of P_{US} is derived as follows:

$$\begin{aligned}
P_{US} =& Pr(\hat{\mathbf{s}}_{US}\neq\mathbf{s}_0)=Pr(\hat{\mathbf{s}}_{US}\neq\mathbf{s}_0,\hat{\mathbf{s}}_{MLSD}=\mathbf{s}_0)+Pr(\hat{\mathbf{s}}_{US}\neq\mathbf{s}_0,\hat{\mathbf{s}}_{MLSD}\neq\mathbf{s}_0) \\
\leq& Pr(\hat{\mathbf{s}}_{US}\neq\hat{\mathbf{s}}_{MLSD}=\mathbf{s}_0)+P_{MLSD}
\end{aligned} \tag{41}$$

where $\mathbf{s}_0$ is the transmitted sequence. The first term of this upper bound can be further bounded as

$$\begin{aligned}
& Pr(\hat{\mathbf{s}}_{US}\neq\hat{\mathbf{s}}_{MLSD}=\mathbf{s}_0)\leq Pr(\{\text{No sample point is in the same partition with }\hat{\theta}(\mathbf{s}_0)\}) \\
=& \sum_{i=1}^{L} Pr(\bigcup_{k,l:k\neq l}\{\phi_k,\phi_l\in(a_i,a_{i+1})\}\cap\{\hat{\theta}(\mathbf{s}_0)\in(\phi_k,\phi_l)\}) \\
\leq& \sum_{i=1}^{L}\sum_{k,l:k\neq l} E_\theta\left[Pr(\{\phi_k,\phi_l\in[a_i,a_{i+1}]\}\cap\{\hat{\theta}(\mathbf{s}_0)\in(\phi_k,\phi_l)\}|\theta)\right]
\end{aligned} \tag{42}$$

where the notation $\phi_l \in [a,b]$ is used to denote the event that one of the threshold pairs obtained from z_l as in (8) is inside $[a,b]$. Define

$$x_1 \triangleq E_p e^{j\theta} + \sqrt{E_p} n_1 \tag{43}$$

$$x_i \triangleq E_s e^{j\theta} + s_{0,i} n_i \quad \forall i = 2,3,\ldots,N \tag{44}$$

$$X \triangleq \sum_{i=1}^{N} x_i \tag{45}$$

We have

$$\phi_i = \angle z_i \pm \frac{\pi}{2} = s_{0,i}\angle z_i \pm \frac{\pi}{2} = \angle x_i \pm \frac{\pi}{2} \quad \forall i = 2,3,\ldots,N \tag{46}$$

$$\hat{\theta}(\mathbf{s}_0) = -\angle(\mathbf{z}^H \mathbf{s}_0) = \angle \sum_{i=1}^{N} \left(s_{0,i}^2 e^{j\theta} + s_{0,i} n_i \right) = \angle X \tag{47}$$

Hence

$$\begin{aligned} & p(\{\phi_k, \phi_l \in [a_i, a_{i+1}]\} \cap \{\hat{\theta}(\mathbf{s}_0) \text{ is between } \phi_k, \phi_l\} | \theta) \\ =& 2 \int_{X:\angle X \in [a_i, a_{i+1}]} \int_{x_k:\angle x_k \pm \frac{\pi}{2} \in [a_i, \angle X]} \int_{x_l:\angle x_l \pm \frac{\pi}{2} \in [\angle X, a_{i+1}]} f(x_k, x_l, X|\theta) dx_l dx_k dX \end{aligned} \tag{48}$$

which is under the assumption that $L \geq 2$ so that the ith pair of thresholds can not lie in the same sampling interval for all i. Since x_i's are i.i.d. complex Gaussian random variables and X is the sum of them, we have

$$f(X|\theta) = \frac{|X|}{\pi E_t N_0} e^{-\frac{||X| e^{j\angle X} - E_t e^{j\theta}|^2}{E_t N_0}} \tag{49}$$

$$f(x_l|X,\theta) = \mathcal{CN}\left(x_l, \frac{E_s}{E_t} X, E_s N_0 (1 - \frac{E_s}{E_t}) \right) \tag{50}$$

$$f(x_k|x_l, X, \theta) = \mathcal{CN}\left(x_k, \frac{E_s}{E_p + (N-2)E_s} (X - x_l), E_s N_0 (1 - \frac{E_s}{E_p + (N-2)E_s}) \right) \tag{51}$$

Letting

$$E_{N-i} \triangleq \frac{[\frac{E_s}{E_p + (N-i)E_s}]^2}{E_s (1 - \frac{E_s}{E_p + (N-i)E_s})} \quad \text{for } i \text{ integer} \tag{52}$$

to simplify notation, the above integral becomes

$$2\int_0^{\infty}\int_{a_i}^{a_{i+1}}\left[\int_{x_l:\angle x_l\pm\frac{\pi}{2}\in[\angle X,a_{i+1}]}\int_{\theta_k\pm\frac{\pi}{2}\in[a_i,\angle X]} T(\theta_k-\angle(X-x_l),\frac{E_{N-2}}{N_0}|X|^2)\right.$$
$$\left.\frac{|x_l|}{\pi E_s N_0(1-\frac{E_s}{E_t})}e^{-\frac{||x_l|e^{j\angle x_l}-\frac{E_s}{E_t}X|^2}{E_sN_0(1-\frac{E_s}{E_t})}}d\theta_k dx_l\right]\frac{|X|}{\pi E_t N_0}e^{-\frac{||X|e^{j\angle X}-E_te^{j\theta}|^2}{E_tN_0}}d\angle X d|X| \tag{53}$$

Now, observe that the integrand of (53) depends only on θ, which is uniformly distributed in $[0,2\pi)$. Since all sampling intervals have the same lengths, (53) should not depend on the choice the sampling interval i. Also, note that (53) does not depend on k nor l. Hence, after combining all the terms, we obtain

$$Pr(\hat{\mathbf{s}}_{US}\neq\hat{\mathbf{s}}_{MLSD}=\mathbf{s}_0)\leq L(N-1)(N-2)\int_0^{\infty}\int_0^{\frac{2\pi}{L}}\frac{1}{2\pi}\left[\int_{-\angle X}^{\angle X}T(\theta_1-\frac{\pi}{2},\frac{E_{N-2}}{N_0}|X|^2)d\theta_1\right]$$
$$\left[\int_{-\frac{2\pi}{L}+\angle X}^{\frac{2\pi}{L}-\angle X}T(\theta_2-\frac{\pi}{2},\frac{E_{N-1}}{N_0}|X|^2)d\theta_2\right]R(|X|,E_t,\frac{E_tN_0}{2})d\angle X d|X|,$$

which reduces to (29a).

D. Proof of Proposition 5

The lower bound is obvious. The upper bound can be derived as follows.

$$\begin{aligned}P_{US2D}=&Pr(\hat{\mathbf{s}}_{US2D}\neq\mathbf{s}_0)\\ =&Pr(\hat{\mathbf{s}}_{US2D}\neq\mathbf{s}_0,\hat{\mathbf{s}}_{CSI}=\mathbf{s}_0,\hat{\mathbf{s}}_{GLRT}=\mathbf{s}_0)+Pr(\hat{\mathbf{s}}_{US2D}\neq\mathbf{s}_0,\hat{\mathbf{s}}_{CSI}\neq\mathbf{s}_0,\hat{\mathbf{s}}_{GLRT}=\mathbf{s}_0)\\ &+Pr(\hat{\mathbf{s}}_{US2D}\neq\mathbf{s}_0,\hat{\mathbf{s}}_{GLRT}\neq\mathbf{s}_0)\\ \leq&Pr(\hat{\mathbf{s}}_{US2D}\neq\hat{\mathbf{s}}_{GLRT}=\hat{\mathbf{s}}_{CSI})+P_{CSI}+P_{GLRT}\end{aligned} \tag{54}$$

The first term of this upper bound can be further calculated as follows

$$\begin{aligned}&Pr(\hat{\mathbf{s}}_{US2D}\neq\hat{\mathbf{s}}_{GLRT}=\hat{\mathbf{s}}_{CSI})\\ \leq&Pr(\{\text{no sample pair is in the same partition with}(f_d,\theta)\})\\ =&Pr(\{\exists\text{ partitioning lines between }(f_d,\theta)\text{ and the nearest four sample pairs}\})\end{aligned} \tag{55}$$

To evaluate this expression, we define A, B, C and D to be the events that no partitioning line lies between $(f_d - \frac{1}{Q_f}, \theta)$ and (f_d, θ), $(f_d, \theta)$and $(f_d, \theta + \frac{2\pi}{Q_\theta})$, $(f_d, \theta)$and $(f_d + \frac{1}{Q_f}, \theta)$, and $(f_d, \theta - \frac{2\pi}{Q_\theta})$ and (f_d, θ), respectively. Also, recall that all partitioning lines have negative slopes as in (20). We have

$$\begin{aligned} &Pr(\{\exists \text{ partitioning lines between } (f_d, \theta) \text{ and the nearest four partitioning points}\}) \\ \leq &Pr((A^c \cup B^c) \cap (B^c \cup C^c) \cap (C^c \cup D^c) \cap (D^c \cup A^c)) \leq Pr(A^c \cap C^c) + Pr(B^c \cap D^c). \end{aligned} \quad (56)$$

We define further the event

$$\mathcal{E}_k(f_1, \theta_1, f_2, \theta_2) \triangleq \{\exists \text{ partitioning lines of the kth symbol that lie between } (f_1, \theta_1) \text{ and } (f_2, \theta_2)\} \quad (57)$$

and subsequently

$$ql(k) \triangleq Pr(\mathcal{E}_k(f_d - \frac{1}{Q_f}, \theta, f_d, \theta) | f_d, \theta) \quad (58a)$$

$$qu(k) \triangleq Pr(\mathcal{E}_k(f_d, \theta, f_d, \theta + \frac{2\pi}{Q_\theta}) | f_d, \theta) \quad (58b)$$

$$qr(k) \triangleq Pr(\mathcal{E}_k(f_d, \theta, f_d + \frac{1}{Q_f}, \theta) | f_d, \theta) \quad (58c)$$

$$qd(k) \triangleq Pr(\mathcal{E}_k(f_d, \theta - \frac{2\pi}{Q_\theta}, f_d, \theta) | f_d, \theta) \quad (58d)$$

Since conditioning on f_d and θ the N sets of partitioning lines are linearly independent, we have

$$\begin{aligned} &Pr(A^c \cap C^c) + Pr(B^c \cap D^c) \\ =&[1 - Pr(A) - Pr(C) + Pr(A \cap C)] + [1 - Pr(B) - Pr(D) + Pr(B \cap D)] \\ =&E_{f_d,\theta}\left[\left\{1 - \prod_{k=1}^{N}(1 - ql(k)) - \prod_{k=1}^{N}(1 - qr(k)) + \prod_{k=1}^{N}(1 - ql(k) - qr(k))\right\} + \right. \\ &\left. + \left\{1 - \prod_{k=1}^{N}(1 - qu(k)) - \prod_{k=1}^{N}(1 - qd(k)) + \prod_{k=1}^{N}(1 - qu(k) - qd(k))\right\}\right], \end{aligned} \quad (59)$$

where we have used the fact that $Q_\theta \geq 4$ and $Q_f \geq 4N$ in the last equality. Now, since $Q_f \geq 4N$, $ql(k)$ and $qr(k)$ can be evaluated as follows

$$\begin{aligned} ql(k) =& Pr(\frac{1}{2\pi k}(\angle z_k - \theta \pm \frac{\pi}{2}) \in [f_d - \frac{1}{Q_f}, f_d] | f_d, \theta) = \int_0^{\frac{2\pi k}{Q_f}} T(x - \frac{\pi}{2}, \frac{E_s}{N_0}) + T(x + \frac{\pi}{2}, \frac{E_s}{N_0}) dx \\ =& qr(k) = O(\frac{k}{Q_f} e^{-\frac{E_s}{N_0}}) \end{aligned} \tag{60}$$

Similarly, since $Q_\theta \geq 4$, we can evaluate $qu(k)$ and $qd(k)$ as follows

$$\begin{aligned} qu(k) =& Pr((\angle z_k - 2\pi k f_d \pm \frac{\pi}{2}) \in [\theta, \theta + \frac{2\pi}{Q_\theta}] | f_d, \theta) = \int_{-\frac{2\pi}{Q_\theta}}^{0} T(x - \frac{\pi}{2}, \frac{E_s}{N_0}) + T(x + \frac{\pi}{2}, \frac{E_s}{N_0}) dx \\ =& qd(k) = O(\frac{1}{Q_\theta} e^{-\frac{E_s}{N_0}}) \end{aligned} \tag{61}$$

Note that $ql(k), qr(k), qu(k)$ and $qd(k)$ actually do not depend on f_d and θ, and $qu(k)$ and $qd(k)$ do not depend on k. Consequently, we have

$$\begin{aligned} & Pr(\hat{\mathbf{s}}_{US2D} \neq \hat{\mathbf{s}}_{GLRT} = \hat{\mathbf{s}}_{CSI}) \\ \leq & \{1 - 2\prod_{k=1}^{N}(1 - ql(k)) + \prod_{k=1}^{N}(1 - 2ql(k))\} + \left\{1 - 2(1 - qu)^N + (1 - 2qu)^N\right\}, \end{aligned} \tag{62}$$

which reduces to the expression in (32a).

REFERENCES

[1] M. Morelli and U. Mengali, "Feedforward frequency estimation for PSK: a tutorial review," *European Trans. Telecommun.*, vol. 9, no. 2, pp. 103–116, March/April 1998.

[2] C. N. Georghiades and D. L. Snyder, "The expectation-maximization algorithm for symbol unsynchronized sequence detection," *IEEE Trans. Communications*, vol. 39, no. 1, pp. 54–61, Jan. 1991.

[3] M. Ghosh and C. L. Weber, "Maximum-likelihood blind equalization," in *Proc. SPIE*, July 1991, pp. 181–195.

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

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

[6] J. B. Anderson and S. Mohan, "Sequential coding algorithms: A survey cost analysis," *IEEE Trans. Communications*, vol. 32, pp. 169–176, Feb. 1984.

[7] J. Lodge and M. Moher, "Maximum likelihood estimation of CPM signals transmitted over Rayleigh flat fading channels," *IEEE Trans. Communications*, vol. 38, pp. 787–794, June 1990.

[8] R. A. Iltis, "A Bayesian maximum-likelihood sequence estimation algorithm for a priori unknown channels and symbol timing," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 579–588, Apr. 1992.

[9] A. N. D'Andrea, U. Mengali, and G. M. Vitetta, "Approximate ML decoding of coded PSK with no explicit carrier phase reference," *IEEE Trans. Communications*, vol. 42, pp. 1033–1040, Feb./Mar./April 1994.

[10] X. Yu and S. Pasupathy, "Innovations-based MLSE for Rayleigh fading channels," *IEEE Trans. Communications*, vol. 43, no. 2/3/4, pp. 1534–1544, Feb./Mar./Apr. 1995.

[11] R. Raheli, A. Polydoros, and C. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain environments," *IEEE Trans. Communications*, vol. 43, no. 2/3/4, pp. 354–364, Feb/Mar/Apr. 1995.

[12] I. Motedayen and A. Anastasopoulos, "Polynomial-complexity noncoherent symbol-by-symbol detection with application to adaptive iterative decoding of turbo-like codes," *IEEE Trans. Communications*, vol. 51, no. 2, pp. 197–207, Feb. 2003.

[13] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.

[14] K. M. Mackenthun, Jr., "A fast algorithm for multiple-symbol differential detection of MPSK," *IEEE Trans. Communications*, vol. 42, pp. 1471–1474, Feb./Mar./Apr. 1994.

[15] W. Sweldens, "Fast block noncoherent decoding," *IEEE Commun. Lett.*, vol. 5, no. 4, pp. 132–134, Apr. 2001.

[16] H. Edelsbrunner, J. O'Rourke, and R. Seidel, "Constructing arrangements of lines and hyperplanes with applications," *Society for Industrial and Applied Mathematics Journal on Computing*, vol. 15, pp. 341–363, 1986.

[17] R. Nuriyev and A. Anastasopoulos, "Pilot-symbol-assisted coded transmission over the block-noncoherent AWGN channel," *IEEE Trans. Communications*, vol. 51, no. 6, pp. 953–963, June 2003.

[18] M. Luise and R. Reggiannini, "Carrier frequency recovery in all-digital modems for burst-mode transmissions," *IEEE Trans. Communications*, vol. 43, no. 2/3/4, pp. 1169–1178, Mar. 1995.

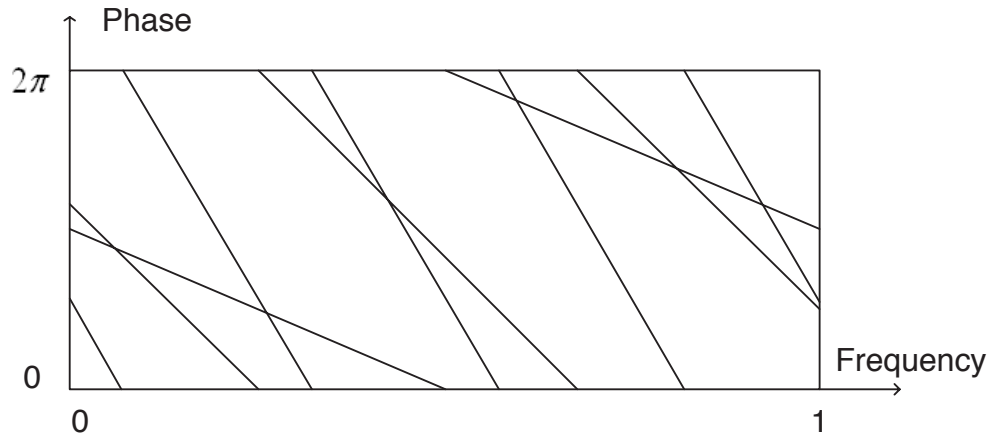

Fig. 1. An example of the partitioned parameter space for $N = 4$

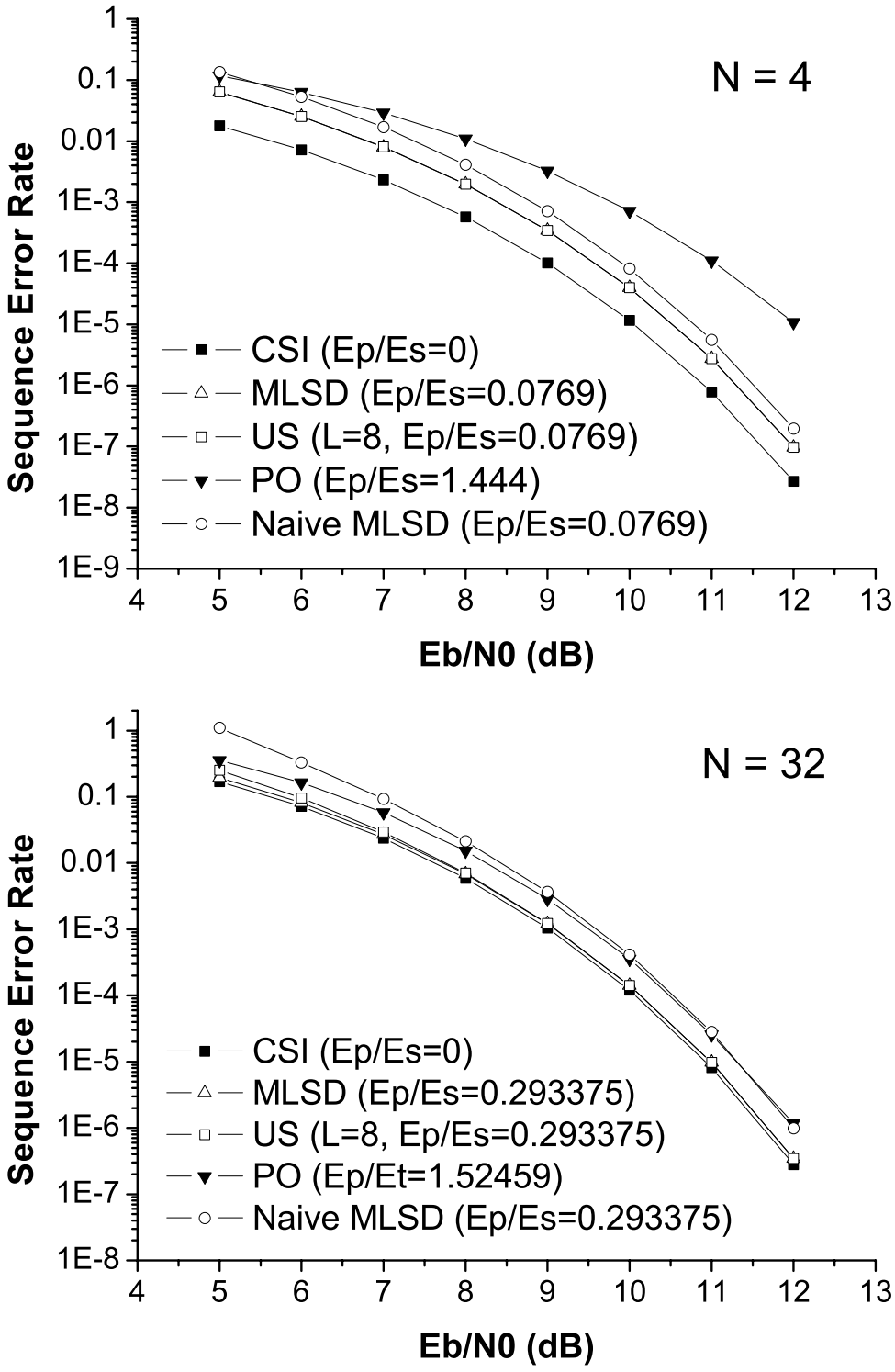

Fig. 2. Analytical results of exact and approximate algorithms for the constant phase model. Above is for $N = 4$ and below is for $N = 32$.

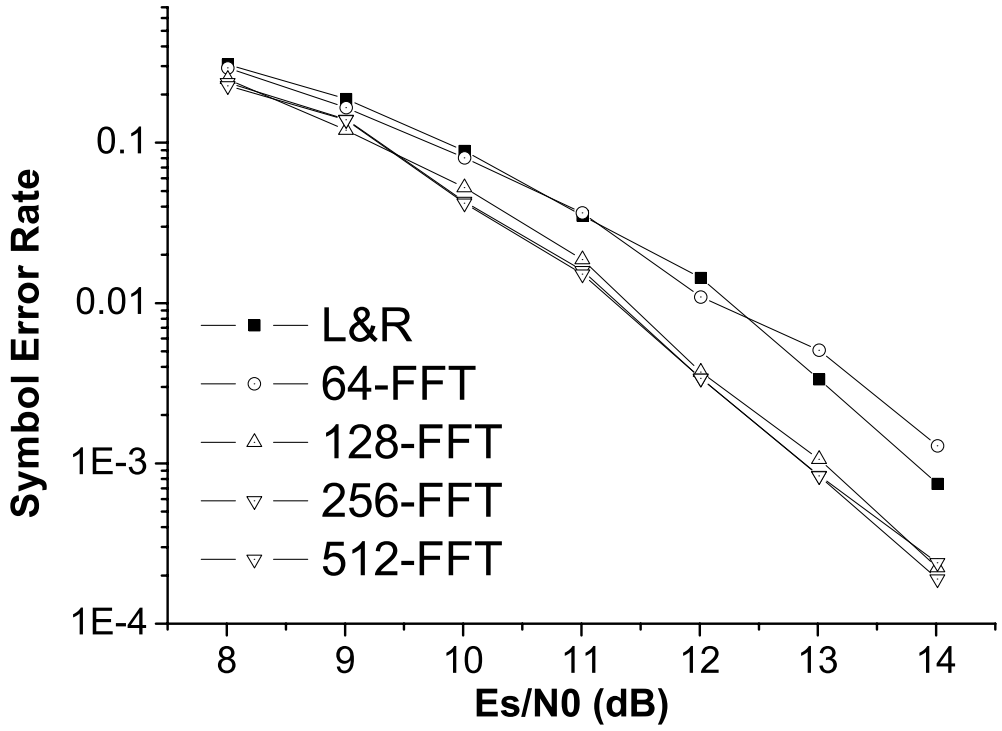

Fig. 3. Comparison of simulation results between different frequency estimators for QPSK modulated sequences with $N = 16$, $f_d = 0.2$, for the US2D algorithm with large Q_f and $Q_\theta = 4$.

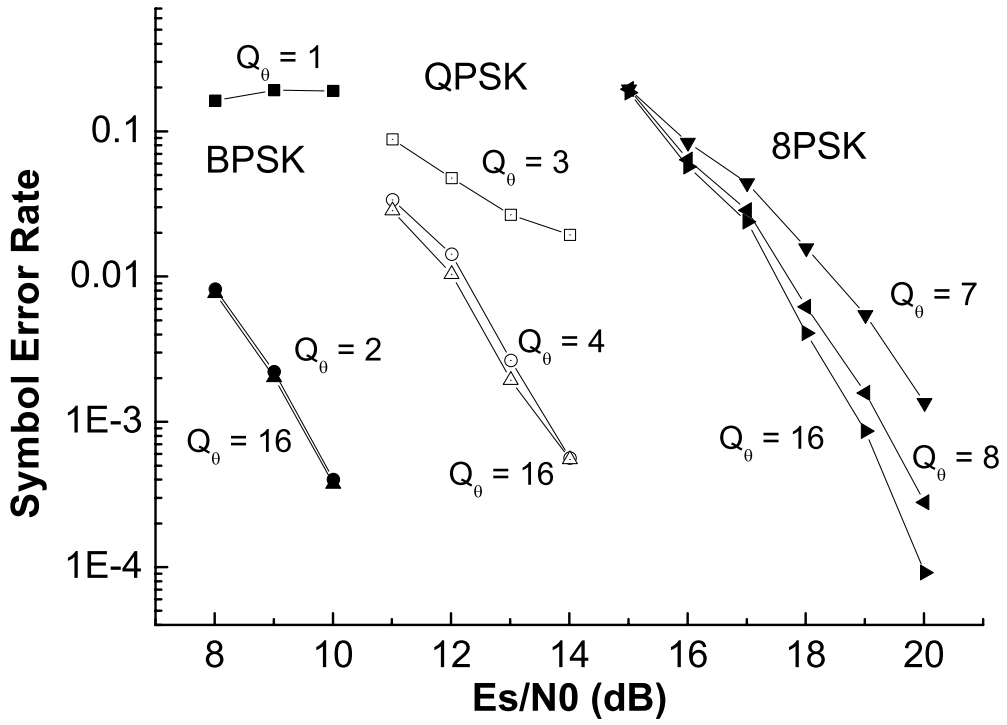

Fig. 4. Relation between M and Q_θ for the simulated US2D algorithm with $N = 16$, $f_d = 0.2$, and a large Q_f.

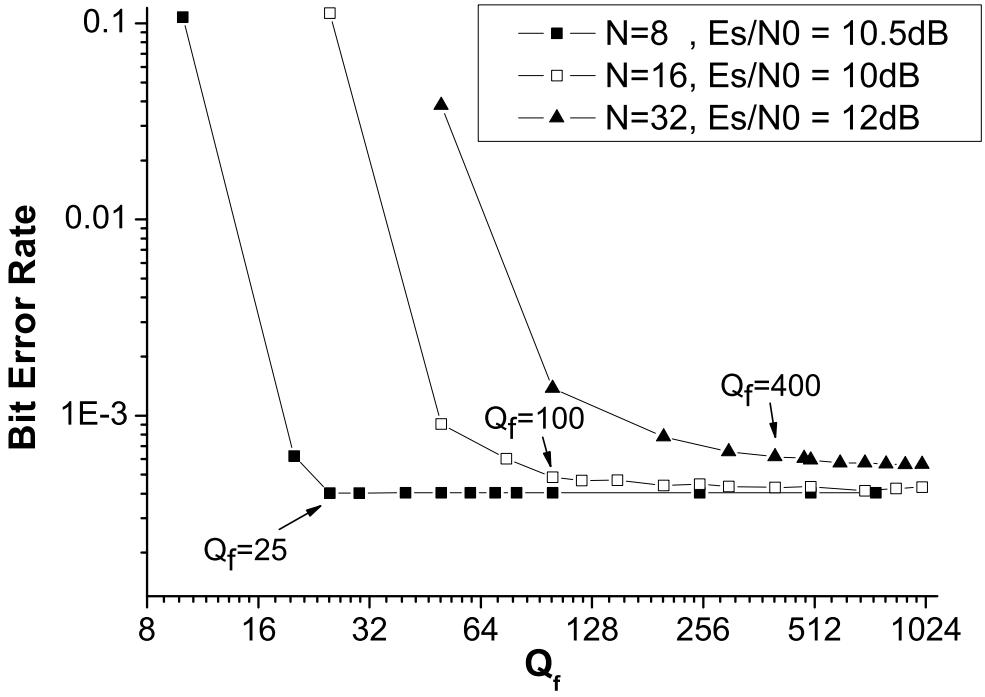

Fig. 5. Effect of Q_f on the simulated US2D algorithm with BPSK signals and f_d, θ uniformly random. In all simulations, $Q_\theta = 2$. The different SNR values are chosen to make all performance curves saturate in approximately the same bit error rate.

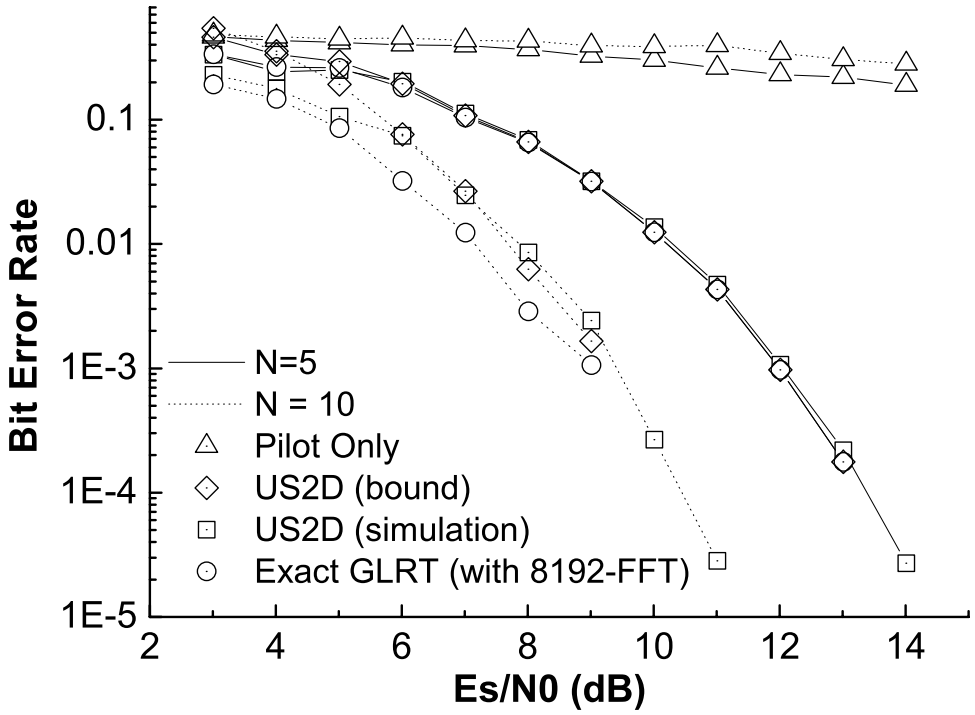

Fig. 6. Simulation results of the exact GLRT, US2D, and an ad hoc pilot-only algorithm for BPSK modulated sequences with f_d, θ uniformly random. $Q_\theta = 2$ and $Q_f = 10$, 40 for $N = 5$, 10, respectively for the simulated US2D algorithm.