

A Maximum Likelihood Digital Receiver Using Coordinate Ascent and the Discrete Wavelet Transform

Ilan Sharfer and Alfred O. Hero III

ABSTRACT

In this paper a Maximum Likelihood (ML) method is presented for joint estimation of amplitude, phase, time delay, and data symbols in a single user direct sequence spread spectrum communication system. Since maximization of the likelihood function is analytically intractable a novel coordinate ascent algorithm is used to obtain sequential updates of the data symbols and all unknown nuisance parameters. The novelty of the algorithm is due to the use of a multi-resolution expansion of the received signal and the use of polynomial rooting in the complex plane in place of a line search over the signal delay parameter. The multi-resolution structure of the algorithm is exploited to reduce sensitivity to impulsive noise via wavelet thresholding. Computer simulations of the single user system show that the algorithm has fast convergence, and comparison to theoretical lower bounds establishes that the algorithm achieves nearly optimal error performance.

Key Words: Wireless communications, sequence estimation, bit synchronization, carrier phase recovery, impulsive noise mitigation.

Submitted to *IEEE Trans. on Signal Processing* (1996)

Corresponding author: Professor Alfred O. Hero III
Room 4229, Dept. of EECS
University of Michigan
1301 Beal Avenue
Ann Arbor, MI 48109-2122.

Tel: (313) 763-0564
FAX: (313) 763-8041
e-mail: hero eeecs.umich.edu.

⁰This research was supported in part by Army Research Office grant DAAH04-96-1-0337

LIST OF FIGURES

Fig 1: Tiling of the time frequency plane.

Fig 2: Wavelet index sets.

Fig 3: Block diagram of the ML receiver.

Fig 4: Block diagram of the wavelet thresholding receiver.

Fig 5: Bit error probabilities and PSK error bound vs. SNR for ML and DLL receivers.

Fig 6: Synchronization RMS error and CR bound vs. SNR for ML and DLL receivers.

Fig 7: Phase tracking RMS error and CR bound vs. SNR for ML and DAL.

Fig 8: Bit error probabilities of conventional and wavelet thresholding receivers in impulsive noise.

1 Introduction

The new technologies of multi-user wireless communication systems, mobile radio, and personal communication networks require advanced signal processing methods for improved efficiency and reliability. The growing computing power and shrinking cost of Digital Signal Processing (DSP) technology makes sophisticated signal processing algorithms both practical and affordable. Wireless communication systems must frequently operate in harsh environmental conditions that adversely affect the transmitted signal through such phenomena as time varying channel response, multipath fading, multi-access and co-channel interference, non-Gaussian noise and loss of synchronization. A parametric model of the received signal typically includes unknown amplitude, phase, and time delay, which should be accurately estimated to ensure near optimal decoding of the data symbols. Traditionally, these nuisance parameters are estimated by non-optimal combination of several sub-systems, each specialized to estimating a particular parameter. For example, carrier phase and time delay estimation are frequently done with a Phase Locked Loop (PLL) and a Delay Locked Loop (DLL), respectively [24]. The PLL and DLL techniques are equivalent to the optimal Maximum Likelihood (ML) or Maximum A posteriori Probability (MAP) estimators, under the assumptions that the data symbols are known, and that either the symbol timing is known (PLL) or the carrier phase is known (DLL). However, the overall estimator is sub optimal and it may unnecessarily reduce the system's operational threshold. Therefore an optimal receiver which jointly estimates the nuisance parameters as well as the data symbols is sought in order to achieve better performance.

A considerable amount of research has gone into improving the performance of the basic PLL and DLL synchronization techniques, e.g. by using decision feedback in a Data-Aided Loop (DAL) [23, 2], a decision directed receiver [22], or by deriving optimal non data-aided estimation structures [27]. More recently digital implementations of Pseudo-Noise (PN) code tracking algorithms using the extended Kalman filter have been proposed [21]. The problem of maximum likelihood sequence estimation for the Inter Symbol Interference (ISI) channel has been treated in the classical work [15] using a whitened matched filter followed by a Viterbi decoding algorithm. Most of these estimation techniques critically depend on the Gaussian noise assumption, so they become ineffective when the interference contains a dominant impulsive noise component. The joint parameter estimation and data demodulation problem becomes even more difficult in a multi-user communication system such as the Code Division Multiple Access (CDMA) system, which provided the motivation for the work reported in this paper.

This paper presents a novel joint ML estimator of the unknown parameters in the context of a single user receiver. First an ML estimation algorithm for the additive white Gaussian noise (AWGN) channel is developed, which is subsequently robustified for the non-Gaussian impulsive noise channel. The ML estimator yields an optimal receiver which generates estimates of the nuisance parameters that are asymptotically uniform minimum variance unbiased (UMVUE) in the sense of achieving the CR bound in the limit of large number of observations. The single user ML algorithm has been adapted to multi-user channels but due to space limitations this is treated elsewhere [36, 37]. Also, while such extensions have not yet been fully explored, it is probably straightforward to apply the approach of this paper to the case of fading and ISI channels.

It is well known that direct maximization of the joint likelihood function is analytically intractable due to the unknown delay parameter. Several suboptimal schemes for approximating the joint maximum likelihood sequence and synchronization parameters have been recently proposed [5, 34, 33]. Another approach is iterative ML estimation using steepest descent, Newton-Raphson, or the Expectation-Maximization (EM) algorithm which have been used in similar problems involving superposition of signals [3, 4, 12, 11, 16]. These algorithms are limited to block processing: all the unknown parameters, and in particular all the data symbols within a block, must be updated simultaneously. Furthermore, many of these algorithms are known to suffer from divergence or very slow convergence. This results in a large processing delay which can only be compensated by increasing processor speed and receiver power consumption. For the case of known signal delay (synchronous case) the Viterbi Algorithm performs maximum likelihood sequential decoding in a single pass over the symbols sequence [15]. However the single pass Viterbi algorithm cannot be implemented as an ML decoder for the asynchronous case. The iterative ML algorithm proposed in this paper uses coordinate ascent maximization to jointly estimate amplitude, phase, time delay and data symbols in a sequential manner. Furthermore, by performing the parameter updates on the coefficients of a multi-resolution decomposition of the received signal a single pass algorithm is obtained.

Coordinate ascent is an iterative maximization technique which cycles over groups (or coordinates) of parameters and guarantees that the likelihood is increased at each iteration. It is related to a class of EM algorithms known as Space-Alternating Generalized EM (SAGE) algorithms introduced in [13] and applied to the synchronous multi-user synchronous CDMA detection problem in [30]. Standard EM algorithms have been applied to synchronous sequence estimation in the

presence of fading [17] and unknown carrier phase [32]. The algorithm in this paper handles both unknown carrier phase and unknown bit synchronization and is equivalent to a SAGE algorithm for a special choice of hidden data sets (complete data) which are defined as one-to-one transformations of the observations (incomplete data). As a result, it can be shown that this coordinate ascent algorithm is a SAGE algorithm which has the fastest possible asymptotic convergence for this particular parameter cycling strategy. Since for signal superposition problems, such as the communications application in this paper, the asymptotic convergence speed of SAGE is significantly faster than standard EM [13], this coordinate ascent algorithm does not suffer from the slow convergence inherent to EM algorithms.

As in [28, 19] a well adapted orthogonal signal representation of the measurements is used as a way to concentrate parameter and symbol information into a low number of coefficients. These representations cover the conventional cardinal series basis, the time limited sinusoidal basis functions [20], the Walsh basis, the exponential basis [19], the Karhunen-Loeve basis, and the Slepian basis [38]. The orthogonal representation is exploited by re-expressing the likelihood function as a polynomial in a complex variable which is then used to solve for the unknown delay by fast root finding methods. Next this is specialized to the Daubechies wavelet basis [7, 41] to obtain a multi-resolution data recursive implementation of the coordinate ascent ML algorithm. Wavelets have the advantage of time and scale localization which make them well suited for recursive parameter update algorithms for which each parameter affects the signal at different times and scales. In the present application, the amplitude and phase parameters are coarse scale parameters in the sense that changes in these parameters affects the signal by a global scale factor over time. On the other hand the data symbols are finer scale parameters since they only affect the signal locally in time. The multi-resolution property is important for the implementation of the coordinate ascent algorithm as it separately encodes coarse scale parameters, e.g. signal amplitude and phase, and finer scale parameters, e.g. data symbols and timing, into different sets of wavelet coefficients.

Wavelet based signal processing techniques have recently seen very rapid growth in related areas. They have demonstrated their usefulness in seismology, image processing, and data compression as well as in many other applications. Wavelet bases enjoy a very strong optimality property for general inverse problems in that their use can achieve accurate and parsimonious representation of the input signal and simultaneously diagonalize the channel [10]. In addition, wavelet and other orthogonal modulation schemes have been suggested as a means of attaining reliable communication

over multiple access, fading, and ISI channels [43, 45, 44]. While such modulation schemes are not required for implementation of the ML receiver presented here, the receiver is naturally adapted to them.

Wavelet domain signal processing can also be very effective in suppression of impulsive noise which exists in wireless and optical communications as well as wireline communications such as digital subscriber lines and cable TV plants. This type of temporally localized noise contaminates a significant portion of the usable transmission spectrum. The optimum receiver for impulsive noise is hard to derive, while narrow band noise rejection techniques are not effective. Although the influence of impulsive noise can be reduced by the use of additional coding [29], this comes at a cost of reduced data rate and/or increased delay. The temporal localization of the wavelet basis can be used to detect and reject impulsive events using a simple thresholding algorithm [1, 10]. The reasoning is that only the largest wavelet coefficients are likely to be significantly corrupted by the noise impulses. It will be shown how this thresholding technique can be easily incorporated within the general framework of iterative ML sequence estimation.

The paper is organized as follows. In Section 2 the single user observation model is defined. In Section 3 the form of the likelihood function is given and simplifications of the line search step are discussed which result in a coordinate ascent algorithm with increasing complexity as a function of the block size. In Section 4 a fixed complexity algorithm is obtained by using wavelet representations. In Section 5 the fixed complexity algorithm is robustified against impulsive noise interference via wavelet thresholding. In Section 6 simulation results are presented and Section 7 concludes the paper.

2 Observation Model

Consider the following single user complex baseband model for the received signal:

$$Y(t) = s(t) + u(t) \quad 0 \leq t \leq T, \quad (1)$$

where $s(t)$ is an attenuated and delayed version of the data modulated signal transmitted by the user,

$$s(t) = a \sum_{n=0}^{N-1} b_n p(t - nT_b - d). \quad (2)$$

Note that $s(t)$ is a superposition of delayed and scaled versions of the signaling waveform $p(t)$ which is assumed to be a known PN code. For simplicity it is assumed that the period of the PN code is

equal to the data symbol time interval T_b . Note that it is not necessary to assume that the support of $p(t)$ is the same as the symbol interval – this property would be critical for extension of the algorithm to ISI channels. The noise $u(t)$ will be assumed to be a complex white Gaussian process with power spectral density $N_0/2$.

The unknown parameters in the model (2) are the complex gain $a \in \{R \exp(i\phi) : R > 0, 0 \leq \phi < \pi\}$, the data symbols $b_n \in \{\pm 1\}$ for $n = 0, \dots, N - 1$, and the time delay d . It is assumed that d is contained in the interval $[-T_d/2, T_d/2]$ for some positive $T_d \ll T$. The n index set notation $I_N = \{0, \dots, N - 1\}$ will be used. For simplicity the data symbols b_n in (2) are assumed to be in BPSK modulation format, but with minor modifications the cases of M-ary PSK, PAM and QAM can be treated. The restriction that a lies in the upper half of the complex plane is required in order to remove the 180 degree phase ambiguity and to enable coherent demodulation of the transmitted symbols. There are several techniques for resolving phase ambiguity at the expense of a small decrease in spectral efficiency [24, chap. 2]. Alternatively, DPSK modulation can be used to encode the source into phase changes rather than absolute phase of the transmitted signal.

3 Joint Maximum Likelihood Estimation

3.1 Direct Maximization of Likelihood

Let $\{\psi_k\}_{k=1}^{\infty}$ be a real orthonormal basis for the space of square integrable function \mathcal{H} . Relative to this basis the continuous time observation Y can be converted to the equivalent discrete time observation

$$Y_k = \langle Y, \psi_k \rangle \quad k = 1, 2, \dots$$

where $\langle f, g \rangle$ denotes the inner product $\int f(t)g^*(t)dt$. The original observation can be recovered from the discrete time observation through the reconstruction formula

$$Y(t) = \sum_k Y_k \psi_k(t). \quad (3)$$

Therefore the coefficients $\{Y_k\}$ are sufficient statistics for the original continuous time observations Y . Given that the observation noise $u(t)$ is a white complex Gaussian process with spectral level $N_0/2$, it follows from (1) and (2) that $\{Y_k\}$ is a set of independent jointly Gaussian complex random variables with mean

$$E\{Y_k\} = a \sum_{n=0}^{N-1} b_n w_k(n; d), \quad k = 1, 2, \dots$$

and variance $N_0/2$, where $w_k(n; d)$ are the projection coefficients of the time function $\{p(t - nT_b - d)\}_{t \in [0, T]}$,

$$w_k(n; d) = \int_{-\infty}^{\infty} p(t - nT_b - d)\psi_k(t)dt. \quad (4)$$

In the sequel, the unknown parameters are collectively referred to as the $N+2$ element parameter vector $\theta = [a, d, b_0, \dots, b_{N-1}]^T$. Thus, up to known additive constants, the log-likelihood function can be expressed as

$$\begin{aligned} \log p(Y; \theta) &= \frac{2}{N_0} [a^* \sum_k Y_k \sum_{n=0}^{N-1} b_n^* w_k^*(n; d) \\ &\quad + a \sum_k Y_k^* \sum_{n=0}^{N-1} b_n w_k(n; d) \\ &\quad - |a|^2 \sum_k | \sum_{n=0}^{N-1} b_n w_k(n; d) |^2]. \end{aligned} \quad (5)$$

A necessary and sufficient condition for $\hat{\theta} = [\hat{a}, \hat{d}, \hat{b}_0, \dots, \hat{b}_{N-1}]^T$ to jointly maximize the log likelihood (5) is that the following three equations be satisfied.

$$\hat{d} = \arg \max_{d \in [-T_d/2, T_d/2]} \log p(Y; [\hat{a}, d, \hat{b}_0, \dots, \hat{b}_{N-1}]) \quad (6)$$

$$\hat{b}_n = \arg \max_{b_n \in \{-1, 1\}} \log p(Y; [\hat{a}, \hat{d}, \hat{b}_0, \dots, b_n, \dots, \hat{b}_{N-1}]), \quad n = 0, \dots, N-1 \quad (7)$$

and

$$\hat{a} = \frac{\sum_k Y_k \sum_{n=0}^{N-1} \hat{b}_n^* w_k^*(n; \hat{d})}{\sum_k | \sum_{n=0}^{N-1} \hat{b}_n w_k(n; \hat{d}) |^2} \quad (8)$$

Note that only one of these equations (8) is explicit. The solution of (7) requires searching over 2^N possible symbol values $\{b_n \in \{-1, 1\}\}_{n=0}^{N-1}$ and the solution of (6) requires performing a line search over $d \in [-T_d/2, T_d/2]$. The high complexity of these latter solutions makes direct maximization impractical and provides the primary motivation for this work. The first step is to simplify the line search.

3.2 Line Search via Polynomial Rooting

First it will be necessary to transform (5) into a polynomial equation in the complex variable $z = e^{j\omega_0 d}$ by representing the projection coefficients $w_k(n; d)$ in Fourier series in the delay variable

d . The benefits of this transformation will become clear: a computationally intensive line search for the maximum over d of (5) is eliminated, and the numerical solution can be found by standard polynomial rooting algorithms which are fast and reliable. Suppose that one is interested in the inner product $w_k(n; d) = \langle p(\cdot - nT_b - d, \psi_k) \rangle$ on some interval $d \in [-T_0/2, T_0/2]$ which contains the *a priori* time delay uncertainty interval $[-T_d/2, T_d/2]$. Define the Fourier coefficients

$$c_{k,m}(n) = \sqrt{T_0} \int_{-\infty}^{\infty} v_{T_0}(t) w_k(n; t) e^{-jm\omega_0 t} dt, \quad (9)$$

where

$$v_{T_0}(t) = \begin{cases} 1/T_0 & |t| \leq T_0/2 \\ 0 & |t| > T_0/2, \end{cases}$$

and $\omega_0 = 2\pi/T_0$. By the convolution property of the Fourier transform

$$c_{k,m}(n) = \frac{\sqrt{T_0}}{2\pi} W_k(n; \omega) * V_{T_0}(\omega) |_{\omega=m\omega_0}, \quad (10)$$

where $V_{T_0}(\omega) = \sin(\omega T_0/2)/(\omega T_0/2)$ is the Fourier transform of $v_{T_0}(t)$ and $W_k(n; \omega) = \int_{-\infty}^{\infty} w_k(n; \tau) e^{-j\omega\tau} d\tau$ is the Fourier transform of $w_k(n; d)$ with respect to d .

The Fourier series expansion of $w_k(n; d)$ over $d \in [-T_0/2, T_0/2]$ can be written as a function of the complex variable z

$$w_k(n; d) = \sum_{m \in \mathbb{Z}} c_{k,m}(n) z^m, \quad (11)$$

with $z = e^{j\omega_0 d}$. Since $w_k(n; d)$ is a real function of d the Fourier coefficients $c_{k,m}(n)$ are conjugate symmetric in m

$$c_{k,m}(n) = c_{k,-m}^*(n) \quad m \in \mathbb{Z}.$$

and $c_{k,0}(n)$ is real.

With the series representation (11) the log-likelihood (5) can be approximated to arbitrary accuracy by a polynomial in the complex variable z . To see this, use conjugate symmetry of the Fourier coefficients to rewrite (5) as the double-sided infinite power series

$$\log p(Y; \theta) = \sum_{m \in \mathbb{Z}} q_m z^m = F(z) + F^*(z^{*-1}), \quad (12)$$

where $F(z) = \sum_{m=1}^{\infty} q_m z^m + q_0/2$ and q_m is the coefficient of $e^{jm\omega_0 d}$ in the log likelihood expansion, that is,

$$q_m = \frac{2}{N_0} [-|a|^2 \sum_{n=0}^{N-1} \sum_k \sum_{p \in \mathbb{Z}} c_{k,p}(n) c_{k,m-p}(n) + a \sum_{n=0}^{N-1} b_n \sum_k Y_k^* c_{k,m}(n) + a^* \sum_{n=0}^{N-1} b_n^* \sum_k Y_k c_{k,m}^*(n)].$$

Note that $F(z)$ depends on the parameter values a and b_0, \dots, b_{N-1} .

Now let $\epsilon > 0$ be arbitrary and define D_f as the smallest positive integer such that

$$\frac{1}{N} \sum_{m > D_f} E[|q_m|^2] < \epsilon/2.$$

Define the order D_f polynomial $\tilde{F}(z) = \sum_{m=1}^{D_f} q_m z^m + q_0/2$. Then

$$Q(z) = \tilde{F}(z) + \tilde{F}^*(z^{*-1})$$

is a polynomial approximation to the log likelihood in the mean square sense

$$\frac{1}{N} E[|\log p(Y; \theta) - Q(z)|^2] < \epsilon.$$

Note that the degree D_f of $\tilde{F}(z)$ is finite since the observations have finite power. D_f will principally depend on the bandwidth of the signaling waveform $p(t)$, which for PN sequences is proportional to the number of chips per symbol interval, i.e. the processing gain, since the Fourier coefficients $c_{k,m}(n)$ decay rapidly as $m\omega_0$ increases beyond the bandwidth of $p(t)$.

In the sequel the notation $Q = Q_{a,\underline{b}}$ will be used to clarify the fact that Q is a function of the amplitude a and the symbols $\underline{b} = [b_0, \dots, b_{N-1}]$. Note that since $Q_{a,\underline{b}}$ is a smooth function the maximum $d = d_{max}$ of $Q_{a,\underline{b}}(e^{j\omega_0 d})$ over $d \in [-T_0/2, T_0/2]$ will occur within $O(\epsilon)$ of the ML estimate \hat{d} and there is a stationary point at this maximum, i.e. $\frac{\partial}{\partial d} Q_{a,\underline{b}}(e^{j\omega_0 d})|_{d=d_{max}} = 0$. The idea is to add a penalty function which will force $Q_{a,\underline{b}}(z)$ to have a stationary point near $z = e^{j\omega_0 d_{max}}$ and such that the stationary points can be found by rooting a polynomial. To this aim define the penalized polynomial

$$\phi_{a,\underline{b}}(z) = Q_{a,\underline{b}}(z) + \log \Pi(z) \tag{13}$$

where $\Pi(z)$ is a rational function of the form

$$\Pi(z) = \frac{G(z)G^*(z^{*-1})}{H(z)H^*(z^{*-1})}, \tag{14}$$

where $G(z)$ and $H(z)$ are polynomials of degree D_g and D_h , respectively, over \mathbb{C} ,

$$G(z) = \sum_{n=0}^{D_g} g_n z^n; H(z) = \sum_{n=0}^{D_h} h_n z^n. \quad (15)$$

Now assume that $\Pi(z)$ is a function which satisfies: 1) $\Pi(z)$ is positive and reasonably flat on the semi-circle $\{z = e^{i\omega} : |\omega| \leq \omega_0 T_d/2\}$; 2) magnitude decays quickly to zero as z deviates from semi-circle. Then any stationary points of $\phi_{a,\underline{b}}(z)$ which are on the unit circle will be close to the stationary points of $Q_{a,\underline{b}}(e^{j\omega_0 d})$ over $d \in [-T_d/2, T_d/2]$. It follows that the line search step (6) of the joint ML estimator can be approximated by the following steps:

1. Differentiate the function $\phi(z) = \phi_{a,\underline{b}}(z)$ (13) with respect to z . This gives a rational function

$$\phi'(z) = \frac{Q'(z)\Pi(z) + \Pi'(z)}{\Pi(z)} = A(z)/B(z) \quad (16)$$

where $A(z)$ is a polynomial of degree $D = 2(D_f + D_g + D_h)$.

2. Find the roots of the numerator polynomial $A(z)$ which lie on the unit circle.
3. Evaluate the objective $\phi(z)$ at each of these roots and select the root which maximizes the penalized objective (13).
4. The phase of the maximizing root specifies the estimate of d .

Since the numerator polynomial of (16) has fixed degree D there is a fixed computational complexity per iteration of the delay estimation step. The penalty function $\Pi(z)$ can be implemented by applying standard filter design methods [31] to the transfer function $G(z)/H(z)$ in (14). The degrees D_g and D_h of the penalty polynomials depend on the desired precision of the time delay estimate which determines the passband ripple and transition bandwidth of $\Pi(z)$. Typically they are much smaller than D_f . The choice of penalty function is analogous to the choice of penalty in the penalized maximum likelihood algorithm.

3.3 Iterative Penalized Maximum Likelihood (PML)

The penalized ML estimate of a $\theta = [a, d, b_0, \dots, b_{N-1}]^T$ is defined as

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \{\log p(Y; \theta) - P(\theta)\}. \quad (17)$$

Here $P(\theta)$ is a user specified penalty function. Penalties have frequently been introduced to regularize the estimator [40], to promote faster convergence [14], or to take advantage of prior information [39]. It will be convenient to express the penalty function as the function $P(z) = -\log \Pi(z)$ of the complex variable $z = e^{j\omega_0 d}$ defined in (13). In this way it can be seen that incorporation of the aforementioned polynomial rooting estimation procedure for estimation of d is equivalent to implementing a PML solution to the joint estimation of θ .

In cases where direct maximization of the objective (17) is not possible iterative optimization is often useful. The iterative algorithm developed in this paper is a coordinate ascent type algorithm that is related to Gauss-Seidel iterations [18] for minimizing quadratic objectives. Fix a coordinate θ_i to be updated and define the parameter vector θ_{-i} as θ with the coordinate θ_i left out: $\theta_{-i} = [\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_{N+2}]^T$. Define j_i a mapping from the natural numbers $1, 2, \dots$ to the parameter index set $\{1, \dots, N+2\}$ which satisfies the property that the sequence j_i , $i = 1, 2, \dots$ cycles through all parameter indices an infinite number of times. For $i = 1, 2, \dots$, the coordinate ascent algorithm produces a sequence $\{\theta^{(i)}\}_{i=1}^{\infty}$ by the following iteration

$$\begin{aligned}\theta_{j_i}^{(i)} &= \arg \max_{\theta_{j_i}} \left\{ \log p(Y; \theta_{j_i}, \theta_{-j_i}^{(i-1)}) - P(\theta_{j_i}, \theta_{-j_i}^{(i-1)}) \right\}, \\ \theta_{-j_i}^{(i)} &= \theta_{-j_i}^{(i-1)}.\end{aligned}$$

Note that in the i -th iteration only the parameter θ_{j_i} of θ is updated while the other parameters θ_{-j_i} are held fixed.

For the present application the penalized log likelihood function is of the form $\log p(Y; \theta) + \log \Pi(e^{j\omega_0 d})$, where $\log p(Y; \theta)$ is given by (5). The convergence rate of coordinate ascent depends on the order of parameter updates and the number of consecutive updates of a given parameter. For positive integers k, N denote by $[k]_N$ the integer k modulo N . Given initial conditions $a^{(0)}, d^{(0)}, b_0^{(0)}, \dots, b_{N-1}^{(0)}$, the following is an algorithm which updates in the order: $d^{(i)} \rightarrow d^{(i+1)}$, $b_{[i-1]_{N+2}}^{(i)} \rightarrow b_{[i-1]_{N+2}}^{(i+1)}$, $a^{(i)} \rightarrow a^{(i+1)}$.

Increasing Complexity Single User Algorithm

For $i = 0, 1, \dots$, update i -th cycle:

1. Find $d^{(i+1)}$ by rooting the polynomial $\phi_{a^{(i)}, \underline{b}^{(i)}}(z)$ as explained in previous section.

2. Update the bit $b_{[i-1]_{N+2}}$ by hard decision

$$b_{[i-1]_{N+2}}^{(i+1)} = \text{sign}(\text{Re} \left\{ a^{(i)} \sum_k w_k([i-1]_{N+2}; d^{(i)}) [Y_k^* - a^{(i)*} I_k^{(i)*}([i-1]_{N+2})] \right\}). \quad (18)$$

where

$$I_k^{(i)}(n) = \sum_{n'=0, n' \neq n}^{N-1} b_{n'}^{(i)} w_k(n'; d^{(i)}). \quad (19)$$

3. Update the amplitude a as follows. Define temporary variable

$$\hat{a} = \frac{\sum_k Y_k \sum_{n=0}^{N-1} b_n^{(i)*} w_k^*(n; d^{(i)})}{\sum_k |\sum_{n=0}^{N-1} b_n^{(i)} w_k(n; d^{(i)})|^2}. \quad (20)$$

If $0 \leq \arg(\hat{a}) < \pi$ set $a^{(i+1)} = \hat{a}$ and $b_n^{(i+1)} = b_n^{(i)}$ for all $n \in I_N$, otherwise set $a^{(i+1)} = -\hat{a}$, and $b_n^{(i+1)} = -b_n^{(i)}$, for all $n \in I_N$.

Note that since at each iteration the objective function is maximized over the associated coordinate, the algorithm ensures a monotone increase in likelihood as it progresses. Use of the temporary variable \hat{a} is necessary to ensure that a has non-negative phase. A deficiency of the algorithm is the problem of growing memory and computation: it must cycle over an increasingly large set of symbols as the number N of these increases.

4 Complexity Reduction Via Wavelet Basis

The growing memory and growing complexity problem will be solved by prescribing a single pass acyclic version of the PML algorithm which updates only those symbols falling within a sliding time window of fixed length. In this scheme the time localized parameters, i.e. symbols b_n , are only updated a finite number of times while the global parameters, i.e. a and d , are updated an infinite number of times. Recall that in Section 2 a general orthogonal representation for the received waveform $\{Y(t)\}_{t \in [0, T]}$ was used, whereby projecting it on an orthogonal basis $\{\psi_k(t)\}_{k > 0}$ an equivalent set of measurements $Y_k = \langle Y(\cdot), \psi_k(\cdot) \rangle$, $k = 1, 2, \dots$, was obtained. To achieve decoupling between the localized and non-localized parameter updates it will be convenient to specialize the basis $\{\psi_k\}$ to one that has the multi-resolution property [25]. This will produce coefficients Y_k which contain only information specific to a particular time and scale component of $\{Y(t)\}_{t \in [0, T]}$. In this way a kind of parsimony of the data representation is achieved: information

needed for a particular local or non-local parameter update is concentrated in only a few coefficients Y_k .

Figure 1 shows time scale tiling diagrams of three orthonormal bases: the Dirac basis, the Fourier basis and the wavelet basis. While the Dirac and Fourier bases are localized only in the time and frequency dimensions, respectively, the wavelet basis has the multi-resolution property: each basis is localized in both dimensions, with a timing resolution which gets finer at smaller scales. The time localized Dirac basis is not parsimonious for updating parameters which are localized in frequency. On the other hand, the Fourier basis is not parsimonious for updating time localized parameters such as the symbol sequence. The same deficiencies would hold for other bases which do not have the multi-resolution property, e.g., the cardinal series basis, the Slepian (prolate spheroid) basis, and the Walsh basis.

Note that there exist nonorthonormal expansions which also have the time-frequency resolution property. For example Gabor frames [7] offer more regular tiling of the time frequency plane. However, non-orthogonality causes leakage across scales and complicates the maximization of the likelihood function. Among the many wavelet bases which can be used, e.g., the Daubechies wavelets, Battle-Lemarie wavelets, wavelet packets [8, 42], local cosine bases and bi-orthogonal wavelets [7], in this paper the Daubechies wavelet basis [6] is adopted. An advantage of this basis is that the coefficients can be computed in real time using the discrete time wavelet transform (DTWT) algorithm [35].

The discrete wavelet basis is defined by time scaling and translation of the single function, called the basic wavelet, $\psi(t)$ [8] producing the double indexed set of basis elements

$$\psi_{jk}(t) = 2^{-j/2}\psi(2^{-j}t - k) \quad (j, k) \in \mathbb{Z}^2.$$

Use of the wavelet basis produces a double indexed set of equivalent measurements $Y_{jk} = \langle Y(\cdot), \psi_{jk}(\cdot) \rangle$, $(j, k) \in \mathbb{Z}^2$, signaling waveform projections $w_{jk}(n; d) = \langle p(\cdot - nT_b - d), \psi_{jk}(\cdot) \rangle$, and a triple indexed set of Fourier coefficients $c_{jk,m}$ of the projections, as defined in the previous section.

As only certain subsets of the wavelet coefficients will be used for each parameter update, it is convenient to define an increasing sequence of subsets of wavelet indices $\{W^{(i)}\}_{i=0}^{\infty}$ such that

$$\begin{aligned} W^{(i)} &\subset W^{(i+1)} \quad i \geq 0 \\ \bigcup_i W^{(i)} &= \mathbb{Z}^2. \end{aligned}$$

The following conditions will be necessary

- i.* The signaling waveform $p(t)$ and the basic wavelet $\psi(t)$ are supported on $[-T_p/2, T_p/2]$ and $[-T_\psi/2, T_\psi/2]$ respectively for some T_p and T_ψ , where $T_p \geq T_b$.
- ii.* A finite number of scales is used in the wavelet decomposition, so the scale index j satisfies $q' \leq j \leq q$, for some $q', q \in \mathbb{Z}$.
- iii.* The wavelet index sets $W^{(i)}$ used in the algorithm are chosen sequentially in such a way that

$$\min\{2^j k : (j, k) \in W^{(i+1)} - W^{(i)}\} \geq \max\{2^j k : (j, k) \in W^{(i)}\}$$

for all $i \in \mathbb{N}$.

The purpose of these assumptions, which are not restrictive, will now be explained. The first assumption is needed to eliminate all but a small neighborhood of overlapping symbols $\{b_n\}$ appearing in the likelihood function. The second assumption is required so that the Fourier coefficients $c_{jk,m}(n)$ can be stored in a table of finite size and that the algorithm have a finite delay. The smallest scale q' can be determined from the bandwidth of the signaling waveform $p(t)$, and the largest scale q can be determined by the maximum tolerable processing delay, e.g. T . The third condition means that the wavelet coefficients are processed over all scales simultaneously and sequentially in time according to increasing temporal localization variable $2^j k$. A graphical illustration of the actual sequence of wavelet indices used in the algorithm is given in Fig. 2.

It is useful to remark that if $T_b = r 2^q$ for some rational number $r = r_1/r_0$ then it is possible to store all the Fourier coefficients $c_{jk,m}(n)$ in a finite table, because they can all be mapped to those with symbol indices $0 \dots, r_0 - 1$:

$$c_{jk,m}(n) = c_{j, k - n_0 r_1 2^{q-j}, m}(n_1), \quad 0 \leq n_1 < r_0,$$

where $n_1 = n \pmod{r_0}$.

Next define the following intermediate variables which will be updated at each iteration of the algorithm

$$\begin{aligned} \mu_m^{(i)}(n) &= \sum_{jk \in W^{(i)}} Y_{jk} c_{jk,m}^*(n) \quad m \in \mathbb{Z}, n \in I_N, \\ \nu_m^{(i)}(n_1, n_2) &= \sum_{jk \in W^{(i)}} d_{jk,m}(n_1, n_2) \quad m \in \mathbb{Z}, n_1, n_2 \in I_N, \end{aligned}$$

where

$$d_{jk,m}(n_1, n_2) = \sum_{p \in \mathbb{Z}} c_{jk,p}(n_1) c_{jk,p-m}^*(n_2). \quad (21)$$

Similar to the Fourier coefficients $c_{jk,m}(n)$ the constants $d_{jk,m}(n_1, n_2)$ can be stored in a finite table. It is easy to verify that $\nu_m^{(i)}(n_1, n_2)$ satisfies the following symmetry properties

$$\begin{aligned} \nu_m^{(i)}(n_1, n_2) &= \nu_m^{(i)}(n_2, n_1) \\ \nu_{-m}^{(i)}(n_1, n_2) &= \nu_m^{(i)*}(n_1, n_2). \end{aligned}$$

Notice that $\mu_m^{(i)}(n)$ is the resultant matched filtering of $\{Y_{jk}\}$ to the Fourier coefficients of the correlation $w_{jk}(n; d)$, while $\nu_m^{(i)}(n_1, n_2)$ is a deterministic quantity.

To reduce the computational complexity of updating the intermediate variables μ and ν it is useful to exploit the finite support property of the basic wavelet. To this aim define the following quantity

$$B = \left\lceil \left(\frac{T_p + T_0}{2} + 2^{q-1} T_\psi \right) / T_b \right\rceil, \quad (22)$$

where $\lceil x \rceil$ indicates the smallest integer greater or equal to x , and the reader should recall that T_p , T_ψ and T_0 are the signaling pulse width, the width of the basic wavelet, and the Fourier analysis window length defining the coefficients $c_{jk,m}$, respectively. Also, for a wavelet index set $W^{(i)}$, define the indices $n_-^{(i)}$ and $n_+^{(i)}$

$$\begin{aligned} n_-^{(i)} &= \min\{n : n \geq \frac{2^j k}{T_b}, \text{ for all } (j, k) \in W^{(i)}, n \in \mathbb{Z}\} \\ n_+^{(i)} &= \max\{n : n \leq \frac{2^j k}{T_b}, \text{ for some } (j, k) \in W^{(i)}, m \in \mathbb{Z}\}. \end{aligned} \quad (23)$$

Proposition 1 of the appendix shows that B is a timewidth parameter for ν_m in the sense that $\nu_m^{(i)}(n_1, n_2) = 0$ for all m for $|n_1 - n_2| \geq 2B$. Proposition 2 of the appendix shows that $\nu_m^{(i)}(n_1, n_2) = 0$ whenever n_1 or n_2 exceed $n_-^{(i)} + B$ and $\mu_m^{(i)}(n) = 0$ whenever n exceeds $n_-^{(i)} + B$ for all $m \in \mathbb{Z}$. The combination of these two propositions specifies a small region of wavelet indices for which the summands of $\mu_m^{(i)}$ and $\nu_m^{(i)}$ are non-zero. Proposition 3 of the appendix shows that if $n_1 \leq n_+^{(i)} - B$ or $n_2 \leq n_+^{(i)} - B$ then $\nu_m^{(p)}(n_1, n_2) = \nu_m^{(i)}(n_1, n_2)$ for all $p \geq i$ and all $m \in \mathbb{Z}$ and similarly for $\mu_m^{(i)}$. Thus this specifies a set of indices for which these intermediate variables need not be updated.

Once the following update index sets are defined a concrete specification of the algorithm can be given.

$$\begin{aligned} T^{(i)} &= \{n : n_+^{(i-1)} - B < n < n_-^{(i)} + B\} \\ U^{(i)} &= \{n : n_+^{(i-1)} - B < n \leq n_+^{(i)} - B\} \end{aligned} \quad (24)$$

Notice that the set $T^{(i)}$ is not empty because $n_-^{(i)} \geq n_+^{(i-1)}$ and $B \geq 1$, while the set $U^{(i)}$ may be empty. Corresponding to the above sets, define the following two dimensional index sets:

$$\begin{aligned} R^{(i)} &= \{(n_1, n_2) : n_1 \in T^{(i)}, n_1 - 2B < n_2 \leq n_1\} \cup \\ &\quad \{(n_1, n_2) : n_2 \in T^{(i)}, n_2 - 2B < n_1 \leq n_2\} \\ V^{(i)} &= \{(n_1, n_2) : n_1 \in U^{(i)}, n_1 - 2B < n_2 \leq n_1\} \cup \\ &\quad \{(n_1, n_2) : n_2 \in U^{(i)}, n_2 - 2B < n_1 \leq n_2\}. \end{aligned} \quad (25)$$

With the above, a fourth condition can be specified on the sliding symbol-update window

iv. During the i -th iteration the algorithm only updates the symbols $\{b_n : n \in T^{(i)}\}$.

Note that for symbols not in $T^{(i)}$ the intermediate variables do not change as a result of the most recent multiresolution samples having indices in $W^{(i)} - W^{(i-1)}$. Hence the only change in the objective of these symbols is a result of updating other parameters, which should have a secondary effect. Consequently, the algorithm no longer will have the cyclic update structure that it had before in Section 3. The set $T^{(i)}$ can therefore be referred to as the "current symbols" index set, while the set $U^{(i)}$ corresponds to "past symbols", that is, symbols whose estimates will not be further updated. The two dimensional sets $R^{(i)}$ and $V^{(i)}$ that define pairs of symbol indices can be interpreted in a similar manner.

The fixed memory wavelet version of the coordinate ascent algorithm presented in Section 3 is given below. To simplify the notation, the Fourier index will be omitted, and the corresponding variables will be typed in boldface, e.g., $\mathbf{c}_{jk}(n)$ instead of $c_{jk,m}(n)$.

Fixed Complexity Single User Algorithm

For $i = 0, 1, \dots$,

1. Choose the wavelet index set $W^{(i+1)}$.

2. Update the local variables $\boldsymbol{\mu}^{(i)}$ and $\boldsymbol{\nu}^{(i)}$ as follows:

$$\begin{aligned}\boldsymbol{\mu}^{(i+1)}(n) &= \boldsymbol{\mu}^{(i)}(n) + \sum_{jk \in D^{(i+1)}} y_{jk} \mathbf{c}_{jk}^*(n), \quad n \in T^{(i+1)}; \\ \boldsymbol{\nu}^{(i+1)}(n_1, n_2) &= \boldsymbol{\nu}^{(i)}(n_1, n_2) + \sum_{jk \in D^{(i+1)}} \mathbf{d}_{jk}(n_1, n_2), \\ n_1, n_2 &\in T^{(i+1)}, |n_1 - n_2| < 2B,\end{aligned}\tag{26}$$

where $D^{(i+1)} = W^{(i+1)} - W^{(i)}$.

3. Update the state variables $\boldsymbol{\alpha}^{(i)}$ and $\boldsymbol{\beta}^{(i)}$ as follows:

$$\begin{aligned}\boldsymbol{\alpha}^{(i+1)} &= \boldsymbol{\alpha}^{(i)} + \sum_{(n_1, n_2) \in V^{(i)}} b_{n_1}^{(i)} b_{n_2}^{(i)*} \boldsymbol{\nu}^{(i)}(n_1, n_2), \\ \boldsymbol{\beta}^{(i+1)} &= \boldsymbol{\beta}^{(i)} + \sum_{n \in U^{(i)}} b_n^{(i)*} \boldsymbol{\mu}^{(i)}(n).\end{aligned}\tag{27}$$

4. Update the delay estimate d :

- Compute the following temporary variables:

$$\begin{aligned}\boldsymbol{\gamma}^{(i+1)} &= |a^{(i)}|^2 [\boldsymbol{\alpha}^{(i+1)} + \sum_{(n_1, n_2) \in R^{(i+1)}} b_{n_1}^{(i)} b_{n_2}^{(i)*} \boldsymbol{\nu}^{(i+1)}(n_1, n_2)], \\ \boldsymbol{\delta}^{(i+1)} &= a^{(i)*} [\boldsymbol{\beta}^{(i+1)} + \sum_{n \in T^{(i+1)}} b_n^{(i)*} \boldsymbol{\mu}^{(i+1)}(n)].\end{aligned}\tag{28}$$

Set up the penalized objective for $z = e^{i\omega_0 d}$:

$$\begin{aligned}\phi^{(i+1)}(z) &= \frac{2}{N_0} \sum_{|m| \leq D_f} (\delta_{-m}^{(i+1)} + \delta_m^{(i+1)*} - \gamma_m^{(i+1)}) z^m \\ &\quad + \log \Pi(z).\end{aligned}\tag{29}$$

- Maximize the penalized objective using polynomial rooting:

$$z^{(i+1)} = \underset{z=e^{i\omega}}{\operatorname{argmax}} \{ \phi^{(i+1)}(z) \}.$$

5. Update symbol estimates b_n , $n \in T^{(i+1)}$:

- Update the local variable:

$$\boldsymbol{\epsilon}^{(i+1)}(n) = \boldsymbol{\epsilon}^{(i)}(n) + \sum_{\substack{n_2 \in U^{(i)} \\ |n-n_2| < 2B}} b_{n_2}^{(i)*} \boldsymbol{\nu}^{(i)}(n, n_2).\tag{30}$$

Next include the most recent data samples:

$$\begin{aligned} \varepsilon^{(i+1)}(n) = & a^{(i)} \sum_{|m| \leq D_f} [\mu_m^{(i+1)*}(n) - a^{(i)*}(\epsilon_m^{(i+1)}(n) \\ & + \sum_{\substack{n_2 \in T^{(i+1)} \\ n_2 \neq n}} b_{n_2}^{(i)*} \nu_m^{(i+1)}(n, n_2))] z^{(i)m}. \end{aligned} \quad (31)$$

The objective for $\{b_n\}$ is (need not compute):

$$Q^{(i+1)}(b_n) = \frac{2}{N_0} [b_n \varepsilon^{(i+1)}(n) + b_n^* \varepsilon^{(i+1)*}(n)]. \quad (32)$$

- Maximize (hard decision):

$$b_n^{(i+1)} = \text{sign}[\text{Re}(\varepsilon^{(i+1)}(n))]. \quad (33)$$

6. Update the amplitude estimate a :

- Compute the following temporary variables:

$$\begin{aligned} \zeta^{(i+1)} &= \sum_{|m| \leq D_f} z^{(i)m} (\beta_{-m}^{(i+1)} + \sum_{n \in T^{(i+1)}} b_n^{(i)*} \mu_{-m}^{(i+1)}(n)), \\ \eta^{(i+1)} &= \sum_{|m| \leq D_f} (\alpha_m^{(i+1)} + \sum_{(n_1, n_2) \in R^{(i+1)}} b_{n_1}^{(i)} b_{n_2}^{(i)*} \nu_m^{(i+1)}(n_1, n_2)) z^{(i)m}. \end{aligned} \quad (34)$$

The objective for a is (need not compute):

$$Q^{(i+1)}(a) = a^* \zeta^{(i+1)} + a \zeta^{(i+1)*} - |a|^2 \eta^{(i+1)}. \quad (35)$$

- Maximize the objective. Let,

$$\hat{a} = \zeta^{(i+1)} / \eta^{(i+1)}. \quad (36)$$

If $0 \leq \arg(\hat{a}) < \pi$ set $a^{(i+1)} = \hat{a}$ and $b_n^{(i+1)} = b_n^{(i)}$ for all $n \in I_N$, otherwise set $a^{(i+1)} = -\hat{a}$, $b_n^{(i+1)} = -b_n^{(i)}$ for all $n \in I_N$, and multiply $\beta^{(i+1)}$, $\{\epsilon^{(i+1)}(n) : n \in T^{(i+1)}\}$ by -1 .

Notice that inversion of the signs of all the symbols, as implied by the last step, can be most easily done by keeping track of the inversions and performing it just once after the final iteration. In this way multiple writes to memory are avoided.

An implementation of the PML receiver is shown by the block diagram on Figure 3. The incoming complex baseband signal $y(t)$ is oversampled and digitized. The oversampling should be

sufficiently high so that the discrete wavelet transform in continuous time is well approximated by a discrete time wavelet transform. The sampled sequence $\{y_n\}$ is wavelet transformed using the DTWT algorithm [35], which can be efficiently implemented with an octave-band filter bank structure. The fully digital algorithm described above performs ML estimation of the parameters in a time recursive manner, and passes the estimated symbol sequence \hat{b}_n on to further decoding stages.

5 Impulsive Noise Robustification

In this section a robustification of the algorithm is given for impulsive noise channels. The wavelet shrinkage method of [10] uses a soft wavelet shrinkage algorithm to optimally reconstruct a signal from samples contaminated by additive white Gaussian noise. In this method, small wavelet coefficients are set to zero, since they are likely to contain little signal energy, and, to compensate, larger wavelet coefficients are scaled down since they are likely to contain greater signal energy. This method, similarly to [1], eliminates noisy wavelet coefficients by comparison to a predetermined high threshold. Wavelet coefficients larger than the threshold are rejected because they have most likely been corrupted by the impulsive events. In the method of [1] the rejected wavelet coefficients are reconstructed via an FFT based interpolation algorithm. Assuming that the noise is dominated by the impulsive component while the Gaussian noise is very weak, there is practically no loss of performance if the noisy coefficients are rejected along with the signal component. This is because the residual signal power is still very large compared to the Gaussian noise power. While such an extension is not pursued in this paper, one could also simultaneously implement soft thresholding to reduce the effect of high power Gaussian noise.

A general block diagram of the wavelet based impulsive noise receiver is shown in Figure 4. This receiver is very similar to the optimum receiver for the AWGN channel described in Section 4. The additional blocks in the diagram perform the threshold estimation algorithm proposed in [1] and the wavelet thresholding operation. In the threshold estimation algorithm the coefficient sequence in each wavelet band is divided into segments, whose length is determined by the length of the time interval where the signal is assumed locally stationary. In each signal segment the standard deviation of the signal is estimated by the square root of the sample variance. The threshold at scale index j , Th_j , is set to a fixed multiple of the standard deviation estimate in that band. At the finer scales the impulsive events are isolated, i.e. no two impulses are captured by a single

wavelet coefficient. In the coarser scales the impulses are no longer isolated, so the threshold should be decreased in order to effectively detect the corrupted coefficients. This can be done by multiplying the threshold with a slowly decreasing function of scale. The wavelet thresholding operator generates a new wavelet coefficient sequence, $\{Y'_{jk}\}$, as follows:

$$Y'_{jk} = \begin{cases} Y_{jk} & |Y_{jk}| < Th_j \\ 0 & |Y_{jk}| \geq Th_j \end{cases} \quad (37)$$

The rest of the algorithm remains as before.

6 Simulations

The algorithm of Section 4 has been evaluated by means of a simulation program written in MATLAB. Several system parameters can be varied in order to examine their effect on the receiver's performance, e.g. choice of wavelet family, number of scales, step size, extent of ISI, choice of PN codes, and number of data symbols. The performance criteria of interest were the symbol error probability, and root mean squared (RMS) error of the time delay and phase estimates.

In this paper only a small subset of all possible variations of system parameters will be presented. We have restricted our attention to an uncoded single-user system at the low SNR range of -1dB to 8dB. The bit error probability results were obtained from simulation runs on contiguous data blocks of 2048 or 4096 bits each. The gain and time delay estimates converged within the first 90 bits of each data block. Afterwards, no further updates of these parameters were necessary. The phase and synchronization error performance results were obtained from 100 Monte-Carlo simulations on 32 bit long data blocks. The signal parameters were chosen randomly and independently of each other. The data bits $\{b_n\}$ were selected as either +1 or -1, and the time delay d was uniformly distributed in $[-T_b/2, T_b/2]$.

A 7 chip PN code was used for the spreading sequence. The transmitted signal was passed through a band limited AWGN channel, $|f| \leq 1/2T_c$, where T_c is the chip time. The Daubechies wavelets [6] of length 4, which are the shortest continuous compactly supported wavelets, were used in the wavelet decomposition of the received signal. The decomposition was done on 6 scales in order to capture a large percentage of the signal energy, close to 90% on the average. The number of new wavelet coefficients added per cycle, i.e. the size of $W^{(i+1)} - W^{(i)}$, was 11, and the number of wavelet samples per bit interval was approximately 63. The time delay penalty $\Pi(z)$ was designed using a sixth order Chebychev equiripple filter. The simulations focus on aided

acquisition or tracking performance as opposed to general acquisition performance to reduce the additional complications of local maxima in the likelihood function. Hence the time delay estimate was initialized close to the true parameter by adjusting the delay penalty function so that the maximum with respect to d is sought within 70% of T_c from the true time delay. Similar procedure was adopted to ensure proper phase initialization during the first 4 symbols of each transmission. The issue of global phase and synchronization acquisition should be addressed in a future study.

Figure 5 shows the bit error probability for the AWGN channel. The results were obtained from approximately 8000 simulated bits at the highest SNR value, to 2000 bits at the lowest SNR. We observe that the simulation results closely match the theoretical lower bound on probability of error for a BPSK decoder [24]. The performance of a phase coherent DLL with comparable response time is also shown. The parameter $\delta = 2/W_L T$ is the normalized response time of the DLL, where W_L is the two-sided loop bandwidth, and T is the signaling period, so a value of $\delta = 31$ was used. The loop damping of the equivalent PLL was taken as $\zeta = 0.707$, and zero detuning was assumed. The DLL error probability was found by numerical integration of the conditional PSK bit error probability with respect to the solution of a stochastic PDE of the steady state synchronization error [24]. Notice that the DLL performance has degraded by more than 1dB with respect to the ideal PSK error bound, while the performance of the coordinate ascent algorithm is essentially optimal. Figure 6 shows the normalized synchronization performance of the coordinate ascent algorithm ($T_b = 1$). The coordinate ascent algorithm achieves an RMS synchronization error that is much smaller than the chip time of $T_c \approx 0.14$, which explains its nearly optimal symbol estimation performance. The synchronization error of the DLL is seen to be much larger. The Cramer-Rao (CR) bound on time delay estimation is also shown for reference. Figure 7 compares the RMS phase error of the coordinate ascent algorithm with that of a Data Aided Loop (DAL). The CR bound on phase error is shown for reference [36]. The DAL performance was calculated with the same value of δ as above, for the case of a suppressed carrier with NRZ data ($m = 0$, see [23]). The two systems have a similar performance, but it should be noted that the DAL relies on a perfectly synchronized reference.

Finally, impulse noise channel results are shown on Figure 8. The impulses arrived according to a Poisson process with an average inter arrival time of approximately $1.5T_b$. The impulse amplitudes were generated according to an i.i.d. complex Gaussian process, and the width of each impulsive event was equal to the chip time T_c , thus effectively covering the signal spectrum. Also,

a weak background AWGN process whose power was 18dB below signal power was present. The thresholds were computed by the following simple iterative algorithm. First the standard deviation σ of each scale is estimated, then coefficients whose magnitude is larger than 3σ are removed, and the process is repeated using the remaining coefficients until convergence. The thresholds were scaled by dividing by \sqrt{j} where the scale index j ranged between 1 to 6. The simulations compared bit error probability estimates of the conventional receiver with those of the wavelet thresholding receiver over a range of Signal to Impulse ratio (SIR). The figure shows that a marked improvement in bit error probability has been achieved by using the wavelet thresholding technique coupled with the coordinate ascent algorithm.

7 Conclusions

This paper introduces a new grouped coordinate ascent method for joint timing and phase synchronization and optimal ML detection of transmitted symbols in a single user receiver. Wavelet thresholding can also be incorporated to adapt the receiver to impulsive noise interference channels. Fourier series and polynomial rooting were used to simplify the delay parameter line search, and a multiresolution wavelet representation of the received signal was used to efficiently match parameter updates to data coefficients. The flexibility in choosing several system parameters such as the wavelet basis, the step size of the algorithm, and the penalty functions, makes the algorithm suitable for a variety of applications depending on technological feasibility and cost considerations. It should be pointed out that the problem of selecting an optimal wavelet basis has not been considered in this work but that several optimal selection algorithms have been proposed elsewhere such as best basis [42, chap. 8] or the matching pursuit [26] which could be implemented for this purpose. Finally, space limitations prevented presentation of asynchronous multi-user detection results. Similarly to the benefits of using the SAGE algorithm for synchronous multi-user detection reported in [30], the grouped coordinate ascent algorithm can be used to accelerate convergence, and to simplify implementation, of iterative ML multi-user detection by decoupling the updates of the parameters of each user. The interested reader can refer to [36, 37] for details of this extension.

References

- [1] P.L. Ainsleigh and C.K. Chui, "A B-wavelet-based noise-reduction algorithm", IEEE Trans. SP, 44(5):1279-1284, May 1996.

- [2] J.W.M. Bergmans and H.W. Wong-Lam, "A class of data-aided timing-recovery schemes", IEEE Trans. COM, 43(2/3/4):1819-27, Feb./Mar./Apr. 1995.
- [3] C. Bouman, K. Sauer, "A local update strategy for iterative reconstruction from projections", IEEE Trans. ASSP, 41(2):534-548, Feb. 1993.
- [4] C. Bouman, K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization", IEEE Trans. on Medical Imaging, to appear Aug. 1996.
- [5] K.-H. Chang, C.N. Georghiades, "Joint maximum-likelihood timing and data estimation for MSK signals from matched-filter samples," Proceedings of IEEE ICC, v. 3, 1994.
- [6] I. Daubechies, "Orthonormal bases of compactly supported wavelets", Comm. Pure Appl. Math., 41 (1988), no. 7, 909-996.
- [7] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis", IEEE Trans. IT, 36(5):961-1005, Sep. 1990.
- [8] I. Daubechies, *Ten lectures on wavelets*, Philadelphia, PA: SIAM, 1992.
- [9] A.P. Dempster, N.M. Laird, D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", J. Roy Stat. Soc. series B, 39(1):1-38, 1977.
- [10] D. L. Donoho, "Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data", in *Different perspectives on wavelets* (I. Daubechies Ed.), Amer. Math. Soc., Providence, RI, 1993.
- [11] M. Feder and J. A. Catipovic, "Algorithms for joint channel estimation and data recovery—application to equalization in underwater communications", IEEE J. Oceanic Eng., 16:42-55, Jan. 1991.
- [12] M. Feder, E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm", IEEE Trans. ASSP, 36(4):477-489, Apr. 1988.
- [13] J.A. Fessler, A.O. Hero, "Space-alternating generalized expectation-maximization algorithm", IEEE Trans. ASSP, 42(10):2664-2677, Oct. 1994.
- [14] J.A. Fessler, A.O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms", IEEE Trans. IP, 4(10):1417-1429, Oct. 1995.

- [15] G.D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference", IEEE Trans. IT, 18(3):363–378, May 1972.
- [16] C.N. Georghiades, D.L. Snyder, "The expectation-maximization algorithm for symbol unsynchronized sequence detection", IEEE Trans. COM, 39(1):54-61, Jan. 1991.
- [17] C.N. Georghiades, J.C. Han, " Sequence estimation in the presence of random parameters via the EM algorithm," IEEE Trans. COM, v 45, n 3, p 300-308, Mar 1997.
- [18] , G.H. Golub, C.F. Van Loan *Matrix computations*, Baltimore, MD: Johns Hopkins University Press, 1983.
- [19] J. Huber, Weilin Liu, "Data-aided synchronization of coherent CPM-receivers," IEEE Trans Commun, v 40, n 1, Jan. 1992, p 178-189.
- [20] J. Huber, Weilin Liu, "Data-aided synchronization of M greater than equivalent to 2-ary coherent CPM receivers," Proceedings of IEEE GLOBECOM, v 2, p 709-713, 1990.
- [21] R.A. Iltis, "An EKF-based joint estimator for interference, multipath, and code delay in a DS spread-spectrum receiver", IEEE Trans. COM, 42(2/3/4):1288-99, Feb./Mar./Apr. 1994.
- [22] H. Kobayashi, "Simultaneous adaptive estimation and decision algorithm for carrier modulated data transmission systems", IEEE Trans. COM, 19:268-280, Jun. 1971.
- [23] W.C. Lindsey, M.K. Simon, "Data-aided carrier tracking loops", IEEE Trans. COM, 19(2):157-168, Apr. 1971.
- [24] W.C. Lindsey, M.K. Simon, *Telecommunication systems engineering*, Englewood Cliffs, NJ: Prentice-Hall, 1973.
- [25] S.G. Mallat, "Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$ ", Trans. Amer. Math. Soc., 315(1):69–87, 1989.
- [26] S.G. Mallat, Z. Zhang, "Matching pursuits with time-frequency dictionaries", IEEE Trans. SP, 41(12):3397-3415, Dec. 1993.
- [27] M.H. Meyers, L.E. Franks, "Joint carrier phase and symbol timing recovery for PAM systems", IEEE Trans. COM, 28(8):1121-1129, Aug. 1980.

- [28] H. Meyr, M. Oerder, A. Polydoros, "On sampling rate, analog prefiltering, and sufficient statistics for digital receivers", IEEE Trans. COM, 42:3208-14, Dec. 1994.
- [29] J.W. Modestino, D. Sargrad, R.E. Bollen, "Use of coding to combat impulse noise on digital subscriber loops", IEEE Trans. COM, 36(5):529-537, May 1988.
- [30] L.B. Nelson, H.V. Poor, "Iterative multiuser receivers for CDMA channels: an EM-based approach," IEEE Trans. COM, v 44, n 12, p 1700-1710, Dec 1996.
- [31] A.V. Oppenheim, R.W. Schaffer, *Discrete-time signal processing*, Englewood Cliffs, NJ: Prentice Hall, 1989.
- [32] J. Qin, "Demodulation of binary PSK signals without explicit carrier synchronization," Proc 1993 IEEE Int Conf Commun, p 498-501, 1993.
- [33] S.-C. Rezeanu, R.E. Ziemer, M.A. Wickert, "Joint maximum-likelihood parameter estimation for burst DS spread-spectrum transmission," IEEE Trans. Communications; v COM-45, n 2, p 227-238, Feb 1997.
- [34] S.-C. Rezeanu, R.E. Ziemer, "Joint maximum-likelihood data and bit synchronization epoch estimation for burst direct sequence spread-spectrum transmission," Proc. of IEEE MILCOM, v. 3, pp. 801-805, 1996.
- [35] O. Rioul and P. Duhamel, "Fast algorithms for discrete and continuous wavelet transforms", IEEE Trans. IT, 38(2):569-586, Mar. 1992.
- [36] , I. Sharfer, "Recursive algorithms for digital communication using the discrete wavelet transform", Ph.D. thesis, University of Michigan, Ann Arbor, 1996.
- [37] , I. Sharfer and A.O. Hero, "Iterative maximum likelihood sequence estimation for CDMA systems using grouped ascent and the DWT," I. Sharfer and A.O. Hero, Proceedings of 1997 IEEE Workshop on Signal Processing Advances in Communications, Paris, pp. 137-140, April 1997.
- [38] D. Slepian, "On bandwidth", Proc. IEEE, 64(3):292-300, Mar. 1976.
- [39] M.A. Tanner, *Tools for Statistical Inference : Methods for the exploration of posterior distributions and likelihood functions*, New York: Springer-Verlag, 1993.

- [40] J.R. Thompson, R.A. Tapia, *Nonparametric function estimation, modeling, and simulation*, Philadelphia, PA: SIAM, 1990.
- [41] M. Vetterli, J. Kovacevic, *Wavelets and sub-band coding*, Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [42] M. V. Wickerhauser, *Adapted wavelet analysis from theory to software*, Wellesley, MA: Peters, 1994.
- [43] G. W. Wornell, "Emerging applications of multirate signal processing and wavelets in digital communications", Proc. IEEE, 84(4):586-603, Apr. 1996.
- [44] K.M. Wong, J. Wu, T.N. Davidson and Q. Jin, "Wavelet packet division multiplexing and wavelet packet design under timing error effects", to appear in IEEE Trans. on Signal Proc., 1998.
- [45] J. Wu, Q. Jin, and K.M. Wong, "Multiplexing based on wavelet packets", in Wavelet Applications II, H.H. Szu, Ed., Proc. of the SPIE, pp. 315-326, 1995.

Appendix

The following proposition follows from the definition of B in (22) and assumption (i) in that section :

Proposition 1: If $|n_1 - n_2| \geq 2B$ then $\nu_m^{(i)}(n_1, n_2) = 0$ for all $m \in \mathbb{Z}$ and index sets $W^{(i)} \subset \{(j, k) : j \leq q, k \in \mathbb{Z}\}$.

proof: The claim will be proved by showing that $d_{jk,m}(n_1, n_2)$ are identically zero. In view of the definition (21) it suffices to show that either $c_{jk,m}(n_1)$ or $c_{jk,m}(n_2)$ are identically zero.

Suppose that $n_2 > n_1$ and the pair (j, k) is chosen such that the lower end of the supporting interval of $\psi_{jk}(t)$ overlaps with that of the n_1 -th symbol, i.e.

$$2^j k - 2^{j-1} T_\psi \leq n_1 T_b + \frac{T_p + T_0}{2}.$$

This implies that $w_{jk}(n_1; t)$ is possibly non-zero, so its Fourier coefficients $c_{jk,m}(n_1)$ are not identically zero. Now consider the right endpoint of the supporting interval of $\psi_{jk}(t)$

$$\begin{aligned}
2^j k + 2^{j-1} T_\psi &\leq n_1 T_b + \frac{T_p + T_0}{2} + 2 \cdot 2^{j-1} T_\psi \\
&\leq n_1 T_b + \frac{T_p + T_0}{2} + 2^q T_\psi \\
&\leq n_2 T_b - 2B T_b + \frac{T_p + T_0}{2} + 2^q T_\psi \\
&\leq n_2 T_b - 2 \left(\frac{T_p + T_0}{2} + 2^{q-1} T_\psi \right) + \frac{T_p + T_0}{2} + 2^q T_\psi \\
&\leq n_2 T_b - \frac{T_p + T_0}{2}.
\end{aligned} \tag{38}$$

The third inequality above follows from the assumption $n_1 \leq n_2 - 2B$, and the fourth one follows from (22). This implies that $w_{jk}(n_2; d)$ is identically zero because $\psi_{jk}(t)$ and $p(t - n_2 T_b - d)$ have non-overlapping supporting intervals. Therefore $c_{jk,m}(n_2)$ are identically zero, and it follows that $d_{jk,m}(n_1, n_2)$ are also zero as claimed. A similar proof applies to the case where $n_1 > n_2$.

Next we show that $\nu_m^{(i)}(n_1, n_2)$ is zero if the most recent sample in the wavelet index set $W^{(i)}$ is localized at the n th symbol and n is sufficiently smaller than either n_1 or n_2 . For a set of wavelet indices $W^{(i)}$ recall the definitions of $n_-^{(i)}$ and $n_+^{(i)}$ in (23).

Proposition 2: If $n_1 \geq n_-^{(i)} + B$ or $n_2 \geq n_-^{(i)} + B$ then $\nu_m^{(i)}(n_1, n_2) = 0$ for all $m \in \mathbb{Z}$. If $n \geq n_-^{(i)} + B$ then $\mu_m^{(i)}(n) = 0$ for all $m \in \mathbb{Z}$.

proof: To show the first part, assume $n_1 \geq n_-^{(i)} + B$, then for all $(j, k) \in W^{(i)}$

$$\begin{aligned}
2^j k + 2^{j-1} T_\psi &\leq 2^j k + 2^{q-1} T_\psi \\
&\leq n_-^{(i)} T_b + 2^{q-1} T_\psi \\
&\leq (n_1 - B) T_b + 2^{q-1} T_\psi \\
&\leq n_1 T_b - \frac{T_p + T_0}{2},
\end{aligned}$$

where the second inequality follows from the definition of $n_-^{(i)}$ and the third from the assumption on n_1 . It follows that $c_{jk,m}(n_1) = 0$ for all $m \in \mathbb{Z}$, which implies $\nu_m^{(i)}(n_1, n_2) = 0$. The rest of the claim follows similarly.

The next proposition states conditions under which $\mu_m^{(i)}(n)$ and $\nu_m^{(i)}(n_1, n_2)$ do not change for an increasing sequence of wavelet index sets $W^{(i)}$.

Proposition 3: If $n_1 \leq n_+^{(i)} - B$ or $n_2 \leq n_+^{(i)} - B$ for some index set $W^{(i)}$ then $\nu_m^{(p)}(n_1, n_2) =$

$\nu_m^{(i)}(n_1, n_2)$ for all $p \geq i$ and $m \in \mathbb{Z}$. If $n \leq n_+^{(i)} - B$ for some wavelet index set $W^{(i)}$ then $\mu_m^{(p)}(n) = \mu_m^{(i)}(n)$ for all $p \geq i$ and $m \in \mathbb{Z}$.

proof: From the definition (21) we have

$$\nu_m^{(p)}(n_1, n_2) = \nu_m^{(i)}(n_1, n_2) + \sum_{jk \in W^{(p)} - W^{(i)}} d_{jk,m}(n_1, n_2).$$

Consider $d_{jk,m}(n_1, n_2)$ for all $(j, k) \in W^{(p)} - W^{(i)}$ and assume that $p > i$ and the difference set is not empty. Then, the left endpoint of the supporting interval of $\psi_{jk}(t)$ satisfies

$$\begin{aligned} 2^j k - 2^{j-1} T_\psi &\geq 2^j k - 2^{q-1} T_\psi \\ &\geq n_+^{(i)} T_b - 2^{q-1} T_\psi \\ &\geq n_1 T_b + B T_b - 2^{q-1} T_\psi \\ &\geq n_1 T_b + \frac{T_p + T_0}{2}, \end{aligned}$$

where the second inequality follows from the definition of $n_+^{(i)}$ and assumption (iii), and the third from the assumption on n_1 . The last inequality shows that the right endpoint of $p(t - n_1 T_b - d)$ is smaller or equal to the left endpoint of $\psi_{jk}(t)$ whenever $|d| \leq T_0/2$. Consequently, $c_{jk,m}(n_1)$ is zero for all $m \in \mathbb{Z}$, which implies that $d_{jk,m}(n_1, n_2)$ is zero for $(j, k) \in W^{(p)} - W^{(i)}$. This proves the claim. The proofs for the other cases are similar.

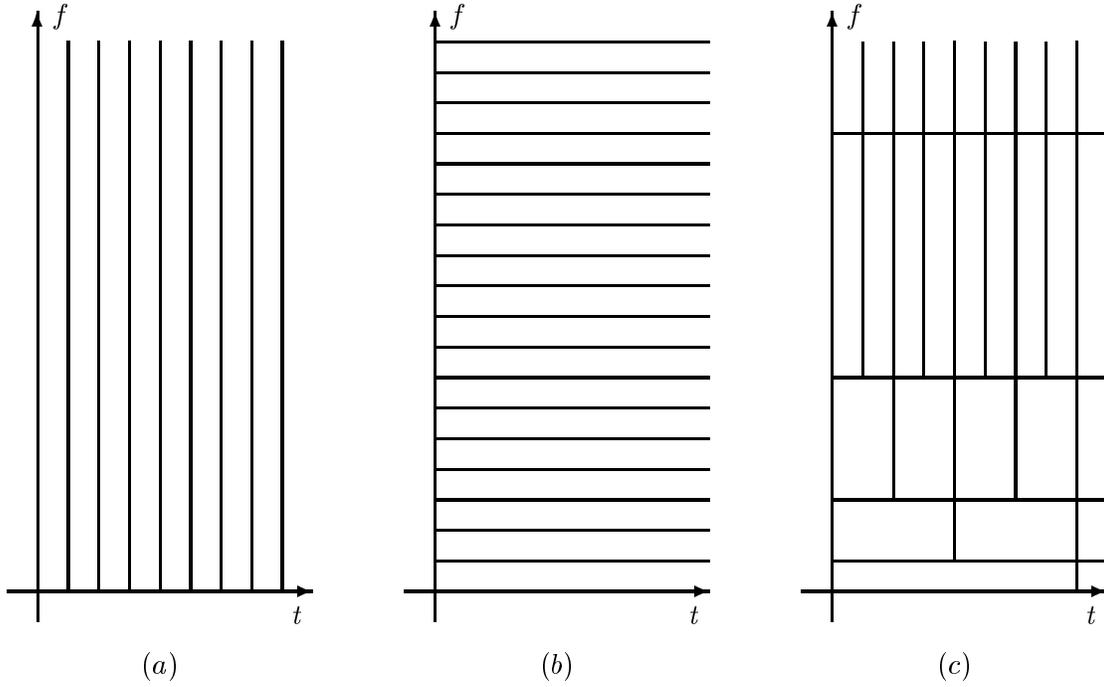


Figure 1: Tiling of the time-frequency plane, (a) Dirac basis, (b) Fourier basis, and (c) wavelet basis.

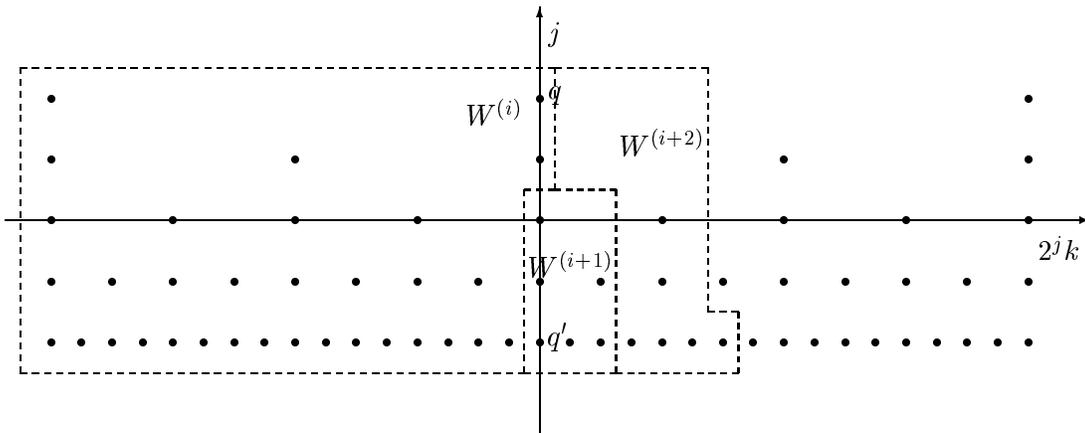


Figure 2: Wavelet index sets.

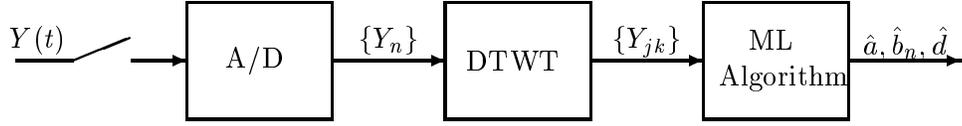


Figure 3: Block diagram of ML receiver.

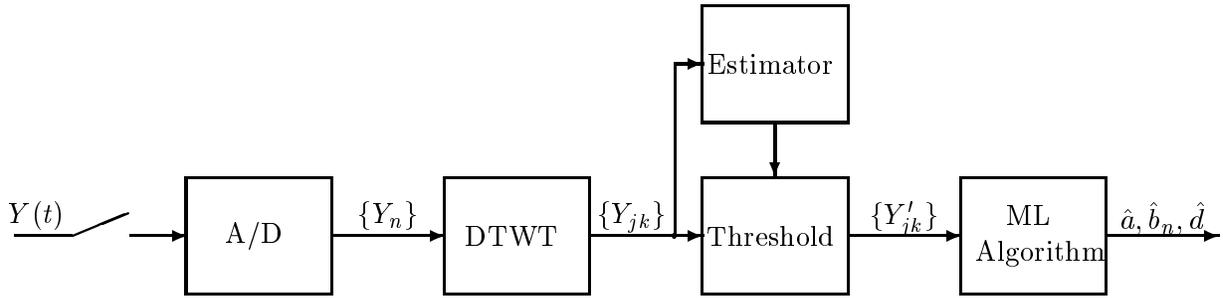


Figure 4: Block diagram of wavelet thresholding receiver.

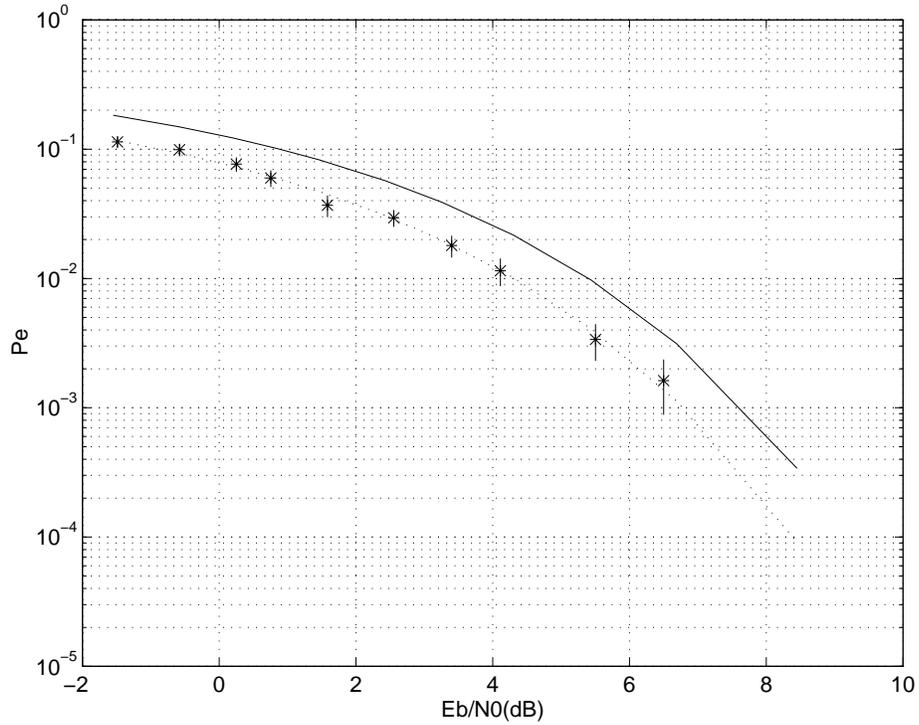


Figure 5: Bit error probability vs. SNR, stars: simulation results; solid: DLL performance; dotted: PSK error bound.

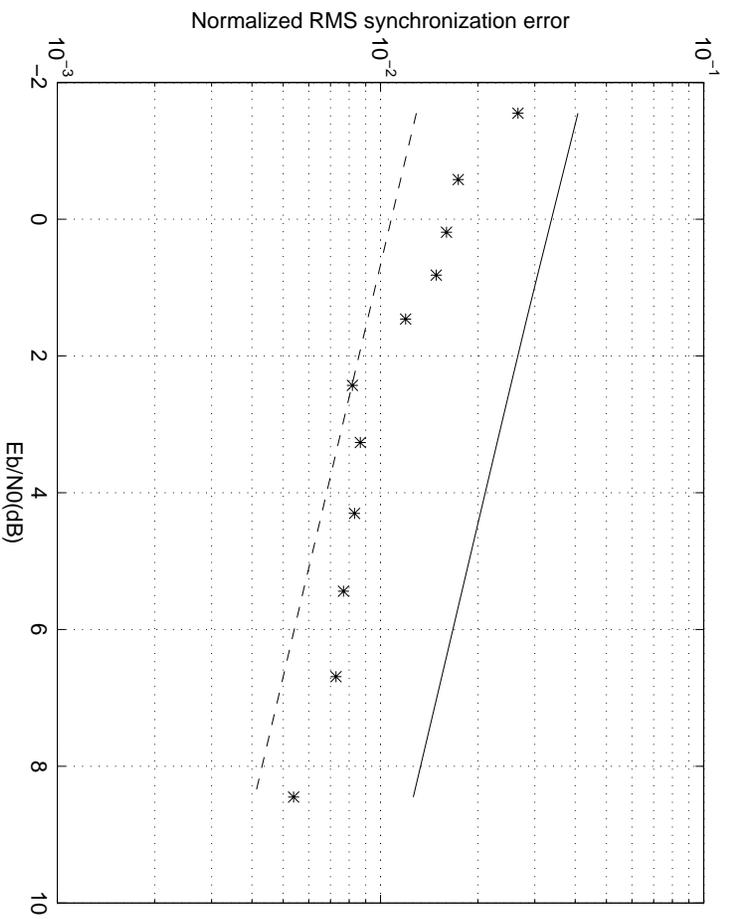


Figure 6: Normalized RMS synchronization error vs. SNR, stars: simulation results; solid: DLL performance; dashed: CR bound.

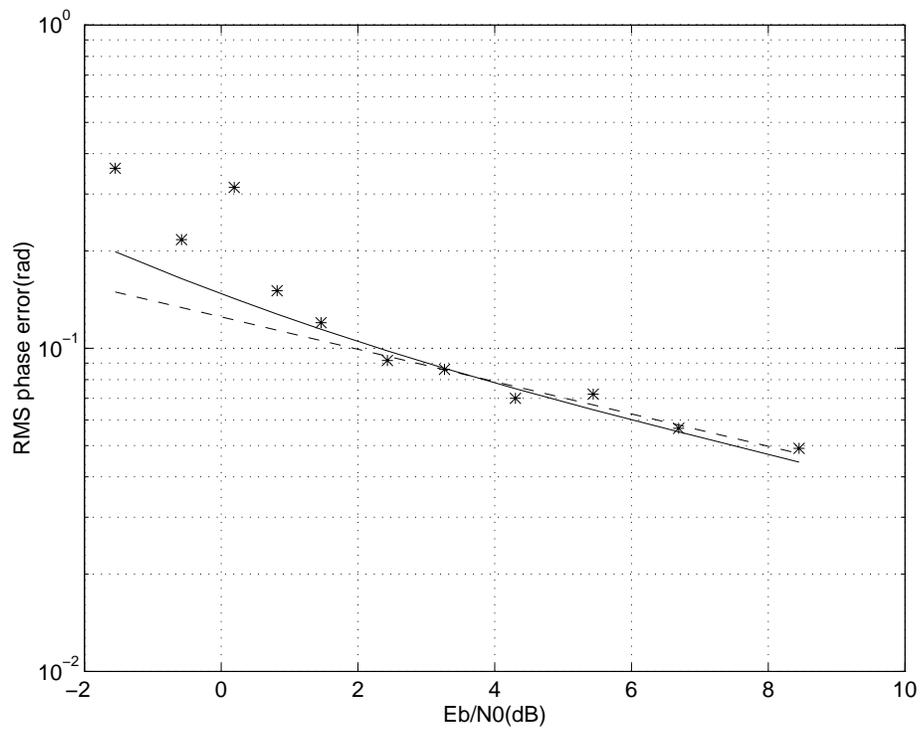


Figure 7: RMS phase error, stars: simulation results; solid: DAL performance; dashed: CR bound.

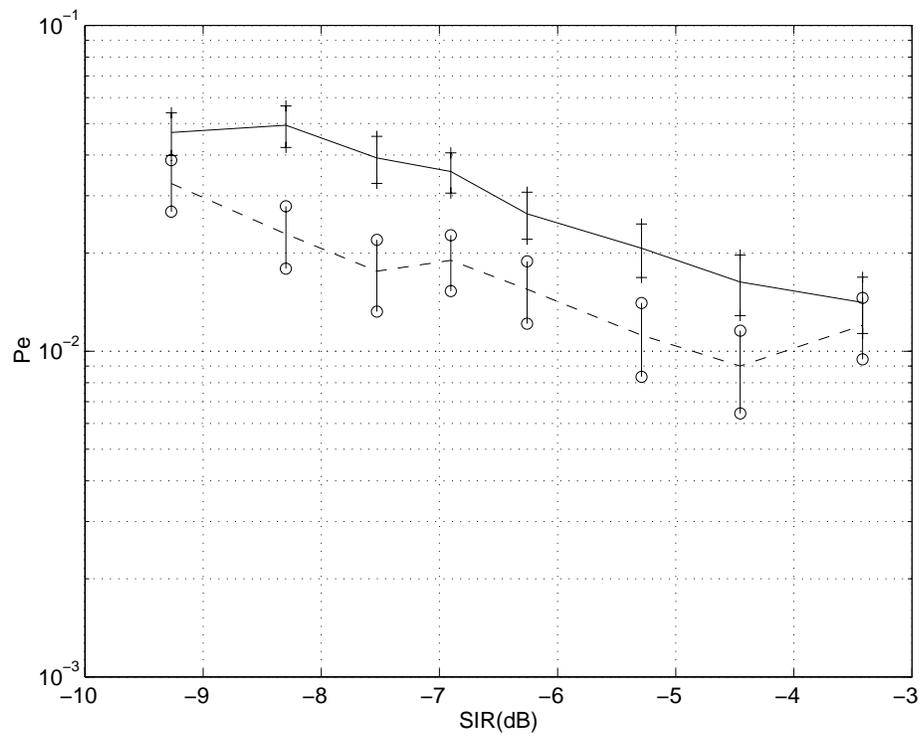


Figure 8: Impulse noise bit error probability vs. SIR, solid: conventional receiver; dashed: wavelet thresholding receiver; Confidence intervals are marked with '+' and 'o' respectively.