A Maximum Likelihood Digital Receiver Using the EM Algorithm and the Discrete Wavelet Transform

Ilan Sharfer and Alfred O. Hero III

ABSTRACT

A Maximum Likelihood (ML) method for joint estimation of amplitude, phase, time delay, and data demodulation on a single user direct sequence spread spectrum communication system is developed. Direct maximization of the likelihood function is analytically intractable, so we consider a recursive estimation algorithm. The Expectation Maximization (EM) algorithm has been used in similar problems, however, it suffers from slow rate of convergence and it is not computationally efficient. Recently, a variant of the EM algorithm, called Space Alternating Generalized EM (SAGE), has been derived. In this work we apply the SAGE algorithm to the sequence estimation problem in a way which results in simple sequential updates of all the estimated parameters. An important feature of the proposed algorithm is the use of a discrete wavelet decomposition of the received signal as a sufficient statistic. The consequence is that while all the information is still available to the receiver, the complicated multi-parameter estimation problem is considerably simplified. Computer simulations of the single user system show that the algorithm has fast convergence, and essentially achieves the theoretically optimal performance.

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

I. INTRODUCTION

The emergence of new technologies of multi-user wireless communication systems, mobile radio, and personal communication networks, require advanced signal processing methods for improved efficiency and reliability. These types of systems operate in harsh environmental conditions which introduce adverse effects on the transmitted signal such as time varying channel response, multipath, fading, interference, and loss of synchronization. In order to optimally decode the desired information the receiver should benefit from knowledge of the nuisance parameters associated with the received signal which typically consist of the amplitude, phase, and time delay. Traditionally, these parameters are estimated by a suboptimal combination of several techniques, each specialized to a particular parameter. For example, carrier phase and time delay estimation are usually done with a Phase Locked Loop (PLL) and a Delay Locked Loop (DLL) respectively [12]. The PLL and DLL techniques are equivalent to the optimal Maximum Likelihood (ML) or Maximum Aposteriori Probability (MAP) estimators, under the assumptions that the data symbols are known, and that for the PLL and DLL the symbol timing and carrier phase are known, respectively. However, the overall estimator performance is suboptimal, and it may become unacceptable at or near the system threshold. It is intuitively clear that an optimal receiver which jointly estimates the nuisance parameters as well as the data symbols will have a superior performance. A considerable amount of research has gone into improving the performance of these basic synchronization techniques, e.g. by using decision feedback in a Data-Aided Loop (DAL) [11], or by deriving non data-aided estimation structures [14]. More recently digital implementations of pseudo-noise (PN) code tracking algorithms using the extended Kalman filter have been proposed [10]. The problem of maximum likelihood sequence estimation for the intersymbol interference (ISI) channel has been treated in the classical work [8] using a whitened matched filter followed by a Viterbi decoding algorithm. The estimation problem becomes even more difficult in a multi-user communication system such as the Code Division Multiple Access (CDMA) system, which motivated this paper.

In this work we focus on the problem of joint ML estimation of all unknown parameters given the received signal for a single user system. Our recursive solution can easily be generalized to the multi-user case, but this will be treated in a follow up paper. Direct maximization of the likelihood function is analytically intractable even for the single user and the Additive White Gaussian Noise (AWGN) channel. Recursive ML estimation algorithms such as the EM algorithm have been used in similar problems [9], [5]. The Space Alternating Generalized EM (SAGE) algorithm which has recently been developed in [6] is a variant of the EM algorithm [4]; both are recursive algorithms which generate a sequence of parameter estimates whose likelihood increases monotonically. The SAGE algorithm is more flexible than the EM algorithm, because it is possible to update a subset of the parameters while fixing the rest of the parameters. The algorithm is formulated by identifying a hidden data space which is admissible with respect to that subset. This additional "degree of freedom" is exploited to derive an algorithm which updates the parameter estimates either analytically or numerically, and has a fast convergence. In this work we will employ the SAGE idea to derive a simple and computationally efficient single user symbol detection algorithm in the presence of amplitude, carrier phase, and symbol timing uncertainty.

One of the basic concepts in estimation theory is that of a sufficient statistic. For example, if we assume that almost all possible realizations of a parameter dependent signal are elements in $L^2(\mathbb{R})$, i.e. they are square integrable, then any reversible transformation applied to the received signal gives a sufficient statistic for that parameter. This key fact was the reason for looking into other signal representations besides the usual time domain representation. The principle of using sufficient statistics in a similar communication problem has been discussed in [15]. We have chosen to represent the signal by its expansion on an orthonormal wavelet basis. This choice has been motivated by the excellent time-frequency localization properties of many wavelet families [2], and the multi resolution property of the wavelet representation. As will be seen, by identifying the wavelet coefficients as part of the hidden data space, the SAGE algorithm can be implemented without cycling through the data symbol sequence. Thus, once the SAGE algorithm has been formulated in terms of the received signal decomposition in the discrete wavelet transform domain, we obtain a causal, fully digital receiver which makes a "single pass" over the data, without requiring any buffering or delayed processing. Wavelet based signal processing techniques have seen a rapid growth recently, and have demonstrated their usefulness in seismology, image processing, and data compression as well as in many other applications.

This paper is organized as follows. In Section 2 we define the general multi-user system model, and briefly review the SAGE algorithm and the concept of a hidden data space. In Section 3 we outline a strategy of choosing the hidden data spaces, and specialize the algorithm to the single user case. In Section 4 we describe a Fourier based method for solving the maximization step of the algorithm numerically. The detailed recursive implementation of the algorithm is given in Section 5. We conclude with simulation results of a single-user system and performance comparison with other techniques.

II. SYSTEM MODEL AND SAGE ALGORITHM REVIEW

We consider the following general L user complex baseband CDMA model:

$$y(t) = \sum_{l=0}^{L-1} s_l(t) + u(t) \qquad -\infty < t < \infty\,, \tag{1}$$

where $s_l(t)$ is the data modulated signal transmitted by the l-th user,

$$s_l(t) = a_l \sum_{n=0}^{N-1} b_{n,l} p_l(t - nT_p - d_l). \tag{2}$$

The unknown parameters of user $l = 0, \ldots, L-1$ are the complex gain $a_l \in \mathbb{C} - \{0\}$, the data symbols $b_{n,l} \in \{e^{i\omega} : \omega \in [-\pi, \pi)\}, n = 0, \ldots, N-1$, and the time delay $d_l \in [-T_d/2, T_d/2]$. We will use the notation $I_N = \{0, \ldots, N-1\}$, and $I_L = \{0, \ldots, L-1\}$ for the data symbol and user index sets respectively. The data symbol model corresponds to phase modulation formats such as BPSK or QPSK. The signaling waveforms $p_l(t)$ are known PN codes. For simplicity we assume that the period of the PN code is equal to the data symbol time interval T_p. Notice that we do not assume that $p_l(t)$ has a compact support, so the algorithm can be applied to ISI channels. Further assumptions about the support of $p_l(t)$ will be given in section 5. The noise $u(t)$ is a complex white Gaussian process with power spectral density $N_0/2$. Following is a brief review of the SAGE algorithm [6].

Let θ be a parameter taking values in a parameter space Θ, which is a subset of the p dimensional Euclidean space $\mathbb{R}^p$ ($\mathbb{C}^p$). The goal is to find the penalized maximum likelihood estimate of θ given the observation $Y = y$:

$$\hat{\theta} = \operatorname*{argmax}_{\theta \in \Theta} \{\log p(y; \theta) - P(\theta)\},$$

where $P(\theta)$ is a suitably defined penalty function. The maximum of the penalized log likelihood is known as the penalized maximum likelihood (PML) estimate. Penalties are frequently introduced to regularized the estimate [18], promote faster convergence [7], or take advantage of prior information [17]. In this work the penalty $P(\theta)$ will be used to enforce a global constraint on updating the locally defined symbol timing parameters.

We define an index set S to be a non-empty subset of the set of integers $I_p = \{1, \cdots, p\}$, and we denote its complement by $\bar{S}$. Corresponding to these index sets we define θ_S and $\theta_{\bar{S}}$ as the elements of θ whose indices are in S and $\bar{S}$ respectively. In the SAGE algorithm the maximization of the penalized log likelihood function is replaced by a maximization of a sequence of objectives functions $\{\phi^{S^{(i)}}(\theta_S; \theta^{(i-1)})\}_{i=1,2,\ldots}$. For this purpose a hypothetical random vector X^S, called the *hidden data space*, is defined such that it satisfies the following admissibility condition

$$p(y, x; \theta) = p(y|x; \theta_{\bar{S}}) p(x; \theta), \quad x \in X^S, y \in Y, \tag{3}$$

i.e. the conditional pdf of Y given X^S does not depend on θ_S. This condition includes the EM complete data space requirement (see [4]) as a special case, when S includes all the elements of θ, i.e. $S = I_p$ and $\theta_S = \phi$ (the empty set). Each iteration of the penalized SAGE algorithm consists of an "E step" and an "M step". First SAGE computes the E step quantities:

$$\phi^S(\theta_S; \theta') = Q^S(\theta_S; \theta') - P(\theta_S, \theta'_{\bar{S}}) \tag{4}$$
$$Q^S(\theta_S; \theta') \equiv E\{\log p(X^S; \theta_S, \theta'_{\bar{S}} | Y = y; \theta')\},$$

where θ' is the current estimate of θ. The maximization of (4), called the M step, completes an iteration of the SAGE algorithm. In this way, starting from an initial parameter guess $\theta^{(0)}$, the algorithm generates a sequence of estimates $\{\theta^{(i)} : i = 0, 1, \ldots\}$ by alternating between different hidden data spaces $X^{S^{(i)}}$, computing $\phi^{S^{(i)}}(\theta_S; \theta^{(i-1)})$ using (4), and maximizing it with respect to θ_S. The SAGE algorithm is guaranteed to increase the penalized log likelihood function at each iteration:

$$\log p(y; \theta^{(i)}) - P(\theta^{(i)}) \geq \log p(y; \theta^{(i-1)}) - P(\theta^{(i-1)}),$$

for $i = 1, 2, \ldots$ [6].

III. SINGLE-USER SAGE ALGORITHM

We start the development of the algorithm by transforming the uncountable *continuous* time domain observation model (1) into an equivalent countable *discrete* observation model. Let $\{\psi_{jk}; j,k \in \mathbb{Z}\}$ be a real orthonormal wavelet basis derived from a multiresolution analysis of $L^2(\mathbb{R})$ [13]. This orthonormal basis is obtained by imparting scaling and translations to a fixed function $\psi(t)$, i.e.

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

There are several bases which can be used, such as the Daubechies wavelets, Battle-Lemarie wavelets, wavelet packets [3], as well as other orthonormal bases, e.g. local cosine bases. From a practical point of view the compactly supported Daubechies wavelet basis [1] is preferred because the wavelet coefficients can be computed in real time, although other non-compact wavelet bases offer better localization in the frequency domain [2], which implies a more stable reconstruction in the presence of a non-ideal sampling process. The received signal $y(t)$ is decomposed in this basis, yielding a set of random variables:

$$Y_{jk} = \langle y, \psi_{jk}\rangle \quad j,k \in \mathbb{Z},$$

where $\langle f, g\rangle$ denotes the inner product $\int f(t)g^*(t)dt$. Since the basis $\{\psi_{jk}\}$ is complete, it is possible to reconstruct the time signal $y(t)$ from the set of wavelet coefficients $Y = \{Y_{jk} : j,k \in \mathbb{Z}\}$ using the synthesis formula of the discrete wavelet transform, so Y is a sufficient statistic for the parameters. Given that $u(t)$ is a white complex Gaussian process, it follows from (1) and (2) that Y is a set of independent jointly Gaussian complex random variables with mean

$$E\{Y_{jk}\} = \sum_{l=0}^{L-1} a_l \sum_{n=0}^{N-1} b_{n,l}\, w_{jk}(n,l;d_l) \quad j,k \in \mathbb{Z},$$

and variance $N_0/2$, where $w_{jk}(n,l;d)$ are the wavelet coefficients of $p_l(t - nT_p - d)$,

$$w_{jk}(n,l;d) = \int p_l(t - nT_p - d)\psi_{jk}(t)dt \quad j,k \in \mathbb{Z}. \tag{5}$$

For brevity we will usually drop the dependence of (5) on the delay parameter d, thus the random variables Y_{jk} can also be written as:

$$Y_{jk} = \sum_{l=0}^{L-1} a_l \sum_{n=0}^{N-1} b_{n,l}\, w_{jk}(n,l) + U_{jk} \qquad j,k \in \mathbb{Z}, \tag{6}$$

where U_{jk}, the projections of $u(t)$ on ψ_{jk}, are independent zero mean complex Gaussian random variables, with variance $N_0/2$. Notice that the time delay parameters d_l implicitly appear in (6) through the wavelet coefficients of $p_l(t-nT_p-d_l)$. The new observation model (6) has several advantages over the original model (1). First, it allows a simplification in the the choice of the hidden data spaces. It also produces a multi-linear model in which the M step of the algorithm is considerably simplified, as opposed to the non linear updates implied by (1) and (2). Finally, the new model is discrete which implies that the likelihood function is given by a summation formula rather than an integral. The multi-user parameter which needs to be estimated is defined as $\theta = [\theta_0, \ldots, \theta_{L-1}]$ where

$$\theta_l = [a_l, w_l, b_{n,l} : n \in I_N],$$
$$w_l = \{w_{jk}(n,l) : n \in I_N,\ j,k \in \mathbb{Z}\}.$$

We observe that the parameter w_l lies on a one-dimensional manifold parametrized by the time delay d, and embedded in an infinite dimensional space. Formally, the parameter $\theta_l, l = 0, \ldots, L-1$, takes values in the parameter space Θ_l given by:

$$\Theta_l = \Theta_a \times \Theta_b^N \times \Theta_{w,l}, \tag{7}$$

where $\Theta_a = \mathbb{C} - \{0\}$, $\Theta_b = \{e^{i\omega} : \omega \in [-\pi,\pi)\}$, and $\Theta_{w,l}$ are defined by:

$$\Theta_{w,l} = \bigcup_{d\in[-T_d/2,T_d/2)} \Theta_{w,l,d},$$

$$\Theta_{w,l,d} = \prod_{j,k\in\mathbb{Z}} \left\{ \prod_{n=0}^{N-1} \langle p_l(\cdot - nT_p - d), \psi_{jk}(\cdot)\rangle, 0^N \right\}. \tag{8}$$

The 0^N element has been included in each set of the Cartesian product (8) in order to allow certain elements $w_{jk}(n,l)$ in w_l to be initialized to 0. This will be exploited by the SAGE algorithm as will be shown shortly.

We now outline the strategy of choosing the hidden data spaces. Let $W \subset \mathbb{Z}^2$ be a non empty set of wavelet indices, and let $\bar{W}$ denote its complement. Define the following data spaces $X_W^S = \{X_{jk}^S : (j,k) \in W\}$:

i. If $\theta_S = w_l, l \in I_L$,

$$X_{jk}^S = a_l \sum_{n=0}^{N-1} b_{n,l} w_{jk}(n,l) + U_{jk}. \tag{9}$$

ii. If $\theta_S = b_{n,l}, n \in I_N, l \in I_L$,

$$X_{jk}^S = a_l b_{n,l} w_{jk}(n,l) + U_{jk}. \tag{10}$$

iii. If $\theta_S = a_l, l \in I_L$,

$$X_{jk}^S = a_l \sum_{n=0}^{N-1} b_{n,l} w_{jk}(n,l) + U_{jk}. \tag{11}$$

We collectively refer to (9)-(11) as the hidden data space X_W^S. The subscript W implies that these hidden data spaces are also indexed by some arbitrary set W of wavelet indices. Note that X_W^S is not physically observable. None of the data spaces defined above is a complete data space for θ in the sense of [4], but as Proposition 1 below shows they are valid hidden data spaces with respect to the corresponding subsets of θ. To illustrate why this is the case, consider for example case (i) where $\theta_S = w_l, l \in I_L$. In view of (6),

$$Y_{jk} = a_l \sum_{n=0}^{N-1} b_{n,l} w_{jk}(n,l) + U_{jk}$$
$$+ \sum_{p=0,p\neq l}^{L-1} a_p \sum_{n=0}^{N-1} b_{n,p} w_{jk}(n,p), \quad j,k \in \mathbb{Z}.$$

Hence, Y_{jk} can be written explicitly in terms of X_{jk}^S:

$$Y_{jk} = \begin{cases} X_{jk}^S + V_{jk}^0, & (j,k) \in W \\ V_{jk}^1, & (j,k) \in \bar{W}, \end{cases} \tag{12}$$

where V_{jk}^0 is a *deterministic* quantity, and X_{jk}^S, V_{jk}^1 are independent complex Gaussian random variables:

$$X_{jk}^S \sim \mathcal{N}(a_l \sum_{n=0}^{N-1} b_{n,l} w_{jk}(n,l), N_0/2), \qquad (j,k) \in W,$$
$$V_{jk}^1 \sim \mathcal{N}(\sum_{l=0}^{L-1} a_l \sum_{n=0}^{N-1} b_{n,l} w_{jk}(n,l), N_0/2), \qquad (j,k) \in \bar{W}. \tag{13}$$

The usefulness of the wavelet index set W will now be explained. Suppose that the parameter θ is chosen such that

$$w_{jk}(n,l) = 0 \quad (j,k) \in \bar{W}, n \in I_N, l \in I_L. \tag{14}$$

The condition (14) and the additive decomposition (12) can be used to establish the fact that X_W^S is a hidden data space, similar to [6, sec. *IV*].

Proposition 1: The data spaces X_W^S given by (9),(10), and (11) are hidden data spaces with respect to w_l, $b_{n,l}$, and a_l respectively.
Proof: The admissibility condition requires that $p(y|x_W^S;\theta)$ does not depend on θ_S. Consider $\theta_S = w_l, l \in I_L$. Based on the previous discussion, and particularly on (12) and (13) the above is found to be

$$p(y|x_W^S;\theta)$$
$$= C \exp\big[-\frac{2}{N_0} \sum_{(j,k)\in\bar{W}} |y_{jk} - \sum_{l=0}^{L-1} a_l \sum_{n=0}^{N-1} b_{n,l} w_{jk}(n,l)|^2\big]$$
$$= C \exp\big[-\frac{2}{N_0} \sum_{(j,k)\in\bar{W}} |y_{jk}|^2\big], \tag{15}$$

where C is a constant which does not depend on θ_S. This proves the claim for $\theta_S = w_l$. The other cases follow similarly.

The selection of update order in the hidden data spaces is arbitrary in general. In this application we consider a sequence of hidden data spaces $\{X_{W^{(i)}}^S\}$ for a sequence of wavelet index sets $W^{(i)}, i \in \mathbb{N}$ which satisfies the following conditions:

$$(i) \qquad W^{(i)} \subset W^{(i+1)}, i \in \mathbb{N}$$
$$(ii) \qquad \bigcup_{i\in\mathbb{N}} W^{(i)} = \mathbb{Z}^2. \tag{16}$$

As will be shown subsequently, this strategy results in recursive parameter updates; in each iteration a new set of observations $\{Y_{jk}\}$ is incorporated and used to modify the current parameter estimates. In this way the algorithm makes only a "single pass" over the entire data set Y. Notice that the parameter updates are performed by cycling through an *increasing* sequence of hidden data spaces, rather than cycling through a *fixed* set of hidden data spaces as in [6], [7]. This fundamental difference is a consequence of the multi resolution property of the wavelet representation.

So far we have explained how the hidden data spaces are chosen in the multi-user case. For the rest of the paper we will consider only the single user case, so all user indices will dropped. Thus the gain, data symbols, and wavelet coefficients will be denoted by a, b_n, and w respectively. Stated below is the single user SAGE algorithm, in a non data recursive form. The algorithm was obtained by computing the objective $Q^S(\theta_S;\theta')$ of (4) for each of the possible hidden data spaces X_W^S given by (9), (10), or (11), and then maximizing it over the respective parameter set θ_S. The problem of choosing a penalty will be discussed in the next section. The value of the parameter estimate at the end of the ith iteration is denoted by $a^{(i)}$, $b_n^{(i)}$, and $w^{(i)}$.

Non-Recursive Single User SAGE Algorithm

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

i. Choose a wavelet index set $W^{(i+1)}$, a parameter index set S, and a hidden data space $X_{W^{(i+1)}}^S$.
ii. If $\theta_S = w$ then:
- E step: Compute

$$Q^S(\theta_S;\theta^{(i)})$$
$$= \frac{2}{N_0}\big[a^{(i)*} \sum_{(j,k)\in W^{(i+1)}} y_{jk} \sum_{n=0}^{N-1} b_n^{(i)*} w_{jk}^*(n)$$
$$+ a^{(i)} \sum_{(j,k)\in W^{(i+1)}} y_{jk}^* \sum_{n=0}^{N-1} b_n^{(i)} w_{jk}(n)$$
$$- |a^{(i)}|^2 \sum_{(j,k)\in W^{(i+1)}} |\sum_{n=0}^{N-1} b_n^{(i)} w_{jk}(n)|^2\big].$$

- M step: Maximize (17) over $\{w \in \Theta_w\}$ which satisfy (14) for $W = W^{(i+1)}$ to yield $w^{(i+1)}$.

iii. If $\theta_S = b_n, n \in I_N$ then:

- E step: Compute

$$Q^S(\theta_S;\theta^{(i)}) = \frac{2}{N_0}\Big[b_n a^{(i)} \sum_{(j,k)\in W^{(i+1)}} w_{jk}^{(i+1)}(n)(y_{jk}^* - I_{jk}^*(n)) + b_n^* a^{(i)*} \sum_{(j,k)\in W^{(i+1)}} w_{jk}^{(i+1)*}(n)(y_{jk} - I_{jk}(n))\Big],$$

where $I_{jk}(n)$ is an intersymbol interference term, caused by the wavelet expansion and not by the channel, given by

$$I_{jk}(n) = a^{(i)} \sum_{m=0,m\neq n}^{N-1} b_m^{(i)} w_{jk}^{(i+1)}(m). \tag{17}$$

- M step: Hard decision (for BPSK)

$$b_n^{(i+1)} = \mathrm{sign}(\mathrm{Re}\,[a^{(i)} \sum_{(j,k)\in W^{(i+1)}} w_{jk}^{(i+1)}(n)(y_{jk}^* - I_{jk}^*(n))]).$$

iv. If $\theta_S = a$ then:
- E step: Compute

$$Q^S(\theta_S;\theta^{(i)}) = \frac{2}{N_0}\Big[a^* \sum_{(j,k)\in W^{(i+1)}} y_{jk} \sum_{n=0}^{N-1} b_n^{(i+1)*} w_{jk}^{(i+1)*}(n) + a \sum_{(j,k)\in W^{(i+1)}} y_{jk}^* \sum_{n=0}^{N-1} b_n^{(i+1)} w_{jk}^{(i+1)}(n) - |a|^2 \sum_{(j,k)\in W^{(i+1)}} |\sum_{n=0}^{N-1} b_n^{(i+1)} w_{jk}^{(i+1)}(n)|^2\Big].$$

- M step:

$$a^{(i+1)} = \sum_{(j,k)\in W^{(i+1)}} y_{jk} \sum_{n=0}^{N-1} b_n^{(i+1)*} w_{jk}^{(i+1)*}(n) \Big/ \sum_{(j,k)\in W^{(i+1)}} |\sum_{n=0}^{N-1} b_n^{(i+1)} w_{jk}^{(i+1)}(n)|^2 \tag{18}$$

We make several comments. First, the SAGE algorithm above is written in order of execution, so the effect of updating w is taken into account when computing the E step objectives for the data symbols b_n, and similarly for the gain a. Second, the algorithm is very time consuming because computation of the E steps (see (17), (17), and (18)) requires summation over the wavelet index set $W^{(i+1)}$, and over the data symbols which are very large sets in general. Finally, while the M steps for b_n and a are explicit and very simple, the M step for w requires a line search over the parameter set Θ_w (8).

IV. Numerical Solution of the M Step

We now describe a method by which the M steps of the SAGE algorithm can be performed via finding the roots of a polynomial equation. The benefits of this method will become clear: a computationally intensive line search for the maximum of (17) is eliminated, and the numerical solution can be found by standard polynomial rooting algorithms which are fast and reliable.

First, recall that the parameters $w_{jk}(n)$ depend implicitly on the time delay parameter d (see (5)). This dependency is captured from the definition of $w_{jk}(n)$ as the correlation between $p(s)$ and $\psi_{jk}(s)$ at lag d:

$$w_{jk}(n;d) = \int p(s-d) 2^{-j/2} \psi(2^{-j}s - (k - 2^{-j}nT_p))\,ds. \tag{19}$$

Transforming (19) to the frequency domain we obtain

$$W_{jk}(n;\omega) = \mathcal{F}\{w_{jk}(n;t)\} = 2^{j/2} e^{-i\omega k 2^j} e^{-i\omega n T_p} \Psi(2^j\omega) P(-\omega),$$

where $\Psi(\omega)$ and $P(\omega)$ are the Fourier transforms of $\psi(t)$ and $p(t)$ respectively.

Suppose that we are interested in the correlation (19) on some interval $d \in [-T_0/2, T_0/2]$. This interval should be larger than the apriori time delay uncertainty T_d, that is $T_0 \geq T_d$. We expand $w_{jk}(n;t)$ in a Fourier series on the interval $[-T_0/2, T_0/2]$ obtaining the Fourier coefficients:

$$c_{jk,m}(n) = \frac{1}{\sqrt{T_0}} \int_{-T_0/2}^{T_0/2} w_{jk}(n;t) e^{-im\omega_0 t} dt = \sqrt{T_0} \int_{-\infty}^{\infty} v_{T_0}(t) w_{jk}(n;t) e^{-im\omega_0 t} dt,$$

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 we have

$$c_{jk,m}(n) = \frac{\sqrt{T_0}}{2\pi} W_{jk}(n;\omega) * V_{T_0}(\omega)|_{\omega = m\omega_0}, \tag{20}$$

where $V_{T_0}(\omega) = \sin(\omega T_0/2)/(\omega T_0/2)$ is the Fourier transform of $v_{T_0}(t)$. The Fourier series expansion of $w_{jk}(n;t)$ can be written in the following z notation as

$$w_{jk}(n;t) = \sum_{m\in\mathbb{Z}} c_{jk,m}(n) z^m, \tag{21}$$

with $z = e^{i\omega_0 t}$ and $c_{jk,m}(n)$ given in (20). The Fourier coefficients $c_{jk,m}(n)$ are conjugate symmetric,

$$c_{jk,m}(n) = c_{jk,-m}^*(n) \qquad m \in \mathbb{Z},$$

because $w_{jk}(n;t)$ is a real function. We observe that the time interval T_0 is arbitrary, but in practice it has to be chosen sufficiently large with respect to T_d in order to avoid convergence problems near the edges due to the Gibbs phenomenon - note that $w_{jk}(n;t)$ is not generally periodic in t. We also note that the number of significant Fourier coefficients $c_{jk,m}(n)$ depends on the bandwidth of $P(\omega)$, which for PN sequences is proportional to the number of chips per symbol interval (also called the *processing gain*). With the representation (21) we can express the objective (17) in the form of a polynomial in z. We illustrate this with a simple example of a single symbol transmitted by the user ($N = 1$). In this case it can be easily verified that the E step objective in (17) can be written as

$$Q^S(\theta_S;\theta') = \sum_{m\in\mathbb{Z}} q_m z^m \,, \tag{22}$$

where

$$q_m = \frac{2}{N_0}[-|a'|^2 \sum_{jk\in W}\sum_{p\in\mathbb{Z}} c_{jk,p}(n) c_{jk,m-p}(n) + a' b'_n \sum_{jk\in W} y^*_{jk} c_{jk,m}(n) + a'^* {b'_n}^* \sum_{jk\in W} y_{jk} c_{jk,m}(n)] \,,$$

and for the single symbol case $n = 0$. The conjugate symmetry of $c_{jk,m}(n)$ implies that $q_m = q^*_{-m}$. A similar polynomial is obtained in the general multi-symbol and multi-user cases. If the summation in (21) is truncated at $|m| \le M$, then (22) is a polynomial of degree $4M$. We also observe that for the E step in (17) the objective is a polynomial of degree 2. Therefore, we have shown that the objectives for the time delay and symbol parameters can be written as conjugate symmetric polynomials in z. This polynomial structure will permit us to introduce a prior or a penalty on d and b_n.

So far we have developed a generic form of the non penalized objective. Next, we need to include penalty functions $P(\theta)$. We will define these penalties to correspond to the priors on d and b_n. Since the objective is a complex function of z evaluated on the unit circle, the prior statistics can be written as $\log \Pi(z)$, where $\Pi(z)$ is any complex function of z satisfying,

$$\begin{aligned} &(i) \quad \Pi(z) \ge 0 \quad ; \quad z = e^{i\omega} \\ &(ii) \quad \int_{z=e^{i\omega}} \Pi(z)\, dz < \infty \,. \end{aligned} \tag{23}$$

It is not necessary to normalize $\Pi(z)$ to integrate to 1 since this will not effect the maximization. A very useful choice of $\Pi(z)$ is one of the form:

$$\Pi(z) = \frac{G(z)G^*(z^{*-1})}{H(z)H^*(z^{*-1})} \,, \tag{24}$$

where $G(z)$ and $H(z)$ are polynomials over $\mathbb{C}$,

$$G(z) = \sum_{n=0}^{D_g} g_n z^n \,;\; H(z) = \sum_{n=0}^{D_h} h_n z^n \,. \tag{25}$$

The condition (23) implies that $\Pi(z)$ cannot have poles on the unit circle, but it may have zeros. The class of functions (24) is quite large, and for sufficiently large D_g and D_h it can approximate any reasonable pdf up to an arbitrarily small error. Therefore, in view of (22) the *penalized* objective can be put in the form

$$\begin{aligned} \phi(z) &= Q(z) + \log \Pi(z), \\ Q(z) &= F(z) + F^*(z^{*-1}), \end{aligned} \tag{26}$$

where $F(z)$ is a polynomial over $\mathbb{C}$,

$$F(z) = \sum_{n=0}^{D_f} f_n z^n, \qquad D_f \ge 1\,. \tag{27}$$

Notice that $\phi(z)$ in (26) is a real function on the unit circle $z = e^{i\omega}$, differentiable with respect to ω except possibly at a finite number of negative valued singularities. Therefore, $\phi(z)$ has a maximum which occurs on a stationary point of $\phi(z)$ on the unit circle. It follows that the M step is implemented by differentiation of (26) with respect to z, finding the roots of the resulting polynomial which lie on the unit circle, and selecting the one which maximizes the penalized objective (26) to be the next parameter estimate. This conclusion is summarized below.

Proposition 2: Suppose $\Pi(z)$ and $F(z)$ are given by (24), (25), and (27). Then the polynomial equation

$$\Pi(z)[F'(z) + {F^*}'(z^{*-1})] + \Pi'(z) = 0 \tag{28}$$

has at least one solution $z = \lambda$ such that: (i) $|\lambda| = 1$, (ii) the function $\phi(z)$ in (26) is maximized on the unit circle at $z = \lambda$.

The polynomial equation (28) has a fixed degree $D = 2(D_f + D_g + D_h)$, which does not depend on the size of the hidden data space $X^S_{W^{(i)}}$. This is a very attractive feature of the algorithm, because it has a fixed computational complexity per iteration, assuming that the root finding algorithm takes advantage of the previous solutions to the same polynomial equation, and that the roots remain relatively stable from one iteration to the next.

We now indicate how the prior density $\Pi(z)$ might be chosen. In the case of the data bit parameter $b_n = e^{i\omega}$, and assuming BPSK modulation, $\Pi(z)$ should have two peaks at $\omega = 0$ and $\omega = \pi$. The width of these peaks can be chosen according to the statistics of the phase jitter due to short term phase instability of the RF oscillators in the transmitter and the receiver. Alternatively, hard decision decoding of the data bits will be optimal if no penalty on b_n is incorporated. In the case of the time delay parameter d, the prior density should be positive and flat on $z = e^{i\omega}$, $|\omega| \le \omega_0 T_d/2$ and close to 0 elsewhere. These priors can be implemented by applying standard filter design methods [16] to the transfer function $G(z)/H(z)$ in (24).

Next we show that it is possible to compute the data dependent coefficients $\{q_m\}$ in (23) recursively, due to the use of an increasing sequence of wavelet index sets $W^{(i)}$, and the adoption of a suitable parameter update strategy.

V. RECURSIVE IMPLEMENTATION OF THE ALGORITHM

In Section 3 we have derived the single user SAGE algorithm in a general non-recursive form. We now develop an efficient recursive method, by which the objectives of the E steps can be computed in real time. Straightforward calculation of the objectives requires a large amount of arithmetic operations per iteration which is proportional to the square of the number of data symbols N. The solution we propose significantly reduces the amount of computations. We define the following variables which will become useful in the development of the algorithm.

$$\mu_m^W(n) = \sum_{jk\in W} y_{jk} c^*_{jk,m}(n) \quad m \in \mathbb{Z}, n \in I_N,$$

$$\nu_m^W(n_1,n_2) = \sum_{jk\in W} d_{jk,m}(n_1,n_2)$$

$$m \in \mathbb{Z}, n_1, n_2 \in I_N, \tag{29}$$

where

$$d_{jk,m}(n_1,n_2) = \sum_{p\in\mathbb{Z}} c_{jk,p}(n_1) c^*_{jk,p-m}(n_2). \tag{30}$$

It is easy to verify that $\nu_m^W(n_1,n_2)$ satisfies the following symmetry properties

$$\nu_m^W(n_1,n_2) = \nu_m^W(n_2,n_1)$$

$$\nu_{-m}^W(n_1,n_2) = \nu_m^{W*}(n_1,n_2).$$

Notice that $\mu_m^W(n)$ represents matched filtering of $\{Y_{jk}\}$ to the Fourier coefficients of the correlation $w_{jk}(n;t)$, while $\nu_m^W(n_1,n_2)$ is a deterministic quantity and represents the energy term. Next, the following assumptions will be made:

i. The signaling waveform $p(t)$ and the basic wavelet $\psi(t)$ are both compactly supported on $[-T_s/2, T_s/2]$ and $[-T_\psi/2, T_\psi/2]$ respectively, where $T_s \geq T_p$.

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 time scale is normalized such that $T_p = r2^q$, where T_p is the data symbol interval, q is the largest scale of the decomposition, and r is a positive integer.

iv. The wavelet index sets $W^{(i)}$ are chosen such 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 too restrictive, will now be explained. The first assumption is needed to eliminate the terms $\nu_m^W(n_1,n_2)$ when n_1 and n_2 are widely separated, or in other words, that the intersymbol interference is localized in a finite range. The second and third assumptions are required so that the constants $c_{jk,m}(n)$ and $d_{jk,m}(n_1,n_2)$ can be stored in a table of finite size. The fourth assumption together with the condition (16) means that the wavelet coefficients are processed over all scales simultaneously, by an increasing order of the localization $2^j k$.

Next we derive conditions under which the variables $\nu_m^W(n_1,n_2)$ are zero, so that double summation terms in the objectives (17) and (18) are simplified considerably. Define:

$$B_j = \min\{k : r2^j k \geq \frac{T_s+T_0}{2} + T_\psi 2^{j-1}, k \in \mathbb{Z}\},$$

$$j \leq q. \tag{31}$$

We note that since only scales $j \leq q$ are used, all Fourier coefficients $c_{jk,m}(n)$ can be uniquely mapped into $c_{jk,m}(0)$ because

$$c_{jk,m}(n) = c_{jk-2^{-(j-q)}rn,m}(0) \quad m \in \mathbb{Z}, \tag{32}$$

which follows from (19). Similarly we get

$$d_{jk,m}(n_1,n_2) = \sum_{p\in\mathbb{Z}} c_{jk-2^{-(j-q)}rn_1,p}(0) c^*_{jk-2^{-(j-q)}rn_2,p-m}(0) \quad m \in \mathbb{Z}. \tag{33}$$

It follows from (31) that $c_{jk,m}(0) = 0$ for all $m \in \mathbb{Z}$ if $|k| \geq rB_j$ because $\psi_{jk}(\cdot)$ and $p(\cdot - t)$ have non-overlapping supports whenever $|t| \leq T_0/2$. This fact implies:

Proposition 3: If $|n_1 - n_2| \geq 2B_q$ then $\nu_m^W(n_1,n_2) = 0$ for all $m \in \mathbb{Z}$ and index sets $W \subset \{(j,k) : j \leq q, k \in \mathbb{Z}\}$.

proof: The claim for ν will be proved by showing that $d_{jk,m}(n_1,n_2) = 0$. Write (33) as

$$d_{jk,m}(n_1,n_2) = \sum_{p\in\mathbb{Z}} c_{jk_1,p}(0) c^*_{jk_2,p-m}(0),$$

where $k_1 = k - 2^{-(j-q)}rn_1$ and $k_2 = k - 2^{-(j-q)}rn_2$. It follows that

$$|k_1 - k_2| = 2^{-(j-q)} r |n_1 - n_2|$$

$$\geq 2rB_q 2^{-(j-q)}.$$

From the definition of B_j we conclude that

$$2^{j-q} B_j \geq \left(2^{-q}\frac{(T_s+T_0)}{2} + 2^{j-q}\frac{T_\psi}{2}\right)/r, \quad j \leq q. \tag{34}$$

For $j \leq q$ the right hand side of (34) decreases as j decreases, while the left hand side does not increase because B_j is chosen to minimize the inequality in (31) on increasingly denser sets. Thus we have shown that $B_q \geq 2^{j-q} B_j$ for $j \leq q$, so

$$|k_1 - k_2| \geq 2rB_j.$$

This implies that either $c_{jk_1,m}(0) = 0$ or $c_{jk_2,m}(0) = 0$ for all $m \in \mathbb{Z}$, so $d_{jk,m}(n_1,n_2) = 0$ as claimed.

We remark here that there are at most $2rB_j - 1$ non zero $c_{jk,m}(n)$ coefficients at the jth scale for all n at a fixed value of m. Therefore all of these coefficients can be stored in a table whose size does not exceed

$$\sum_{j=q'}^{q} 2rB_j - 1.$$

Similarly, in the case of the correlations $d_{jk,m}(n_1,n_2)$, there can be at most $(2rB_j - 1)rB_j$ possible nonzero values at the jth scale for a fixed m (taking into account the symmetry $d_{jk,m}(n_1,n_2) = d_{jk,m}(n_2,n_1)$), so the size of the required table does not exceed

$$\sum_{j=q'}^{q} (2rB_j - 1)rB_j,$$

for a fixed value of m.

Next we show that $\nu_m^W(n_1,n_2)$ is 0 if the most recent sample in the wavelet index set W is localized at the nth symbol and n is sufficiently smaller than either n_1 or n_2. We define

$$\begin{aligned} n^W &= \min\{n : rn \geq 2^{(j-q)}k,\ (j,k) \in W,\ n \in \mathbb{Z}\} \\ m^W &= \max\{n : rn \leq 2^{(j-q)}k,\ (j,k) \in W,\ n \in \mathbb{Z}\} \end{aligned}$$

Proposition 4: If $n_1 \geq n^W + B_q$ or $n_2 \geq n^W + B_q$ then $\nu_m^W(n_1,n_2) = 0$ for all $m \in \mathbb{Z}$. If $n \geq n^W + B_q$ then $\mu_m^W(n) = 0$ for all $m \in \mathbb{Z}$.
proof: To show the first part, assume $n_1 \geq n^W + B_q$, then

$$\begin{aligned} k - 2^{-(j-q)}rn_1 &\leq n^W 2^{-(j-q)}r - 2^{-(j-q)}r(n^W + B_q) \\ &= -2^{-(j-q)}rB_q \\ &\leq -rB_j, \end{aligned}$$

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

The next proposition states conditions under which $\mu_m^W(n)$ and $\nu_m^W(n_1,n_2)$ do not change for an increasing sequence of wavelet index sets $W^{(i)}$. For clarity the set qualifier W will be omitted, e.g. we will write $n^{(i)}$ and $\nu_m^{(i)}(n_1,n_2)$ instead of $n^{W^{(i)}}$ and $\nu_m^{W^{(i)}}(n_1,n_2)$ respectively.

Proposition 5: If $n_1 \leq m^{(i)} - B_q$ or $n_2 \leq m^{(i)} - B_q$ 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 $n \leq m^{(i)} - B_q$ 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 (29) 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 $(j,k) \in W^{(p)} - W^{(i)}$ and assume $p > i$ so the above set is not empty. Then

$$\begin{aligned} k - 2^{-(j-q)}rn_1 &\geq 2^{-j}(2^j k - 2^q r(m^{(i)} - B_q)) \\ &= 2^{-(j-q)}rB_q + 2^{-j}(2^j k - 2^q rm^{(i)}) \end{aligned}$$

We need to show that the last term above is non-negative. From assumption (iv) we get

$$\begin{aligned} 2^{j-q}k &\geq \max\{2^{j-q}k : (j,k) \in W^{(i)}\} \\ &\geq rm^{(i)}, \end{aligned}$$

so the last term is non-negative. The first term $2^{-(j-q)}B_q$ has been shown to be larger or equal to B_j, so $c_{jk,m}(n_1) = 0$ for all $m \in \mathbb{Z}$. This implies that $d_{jk,m}(n_1,n_2)$ is zero for $(j,k) \in W^{(p)} - W^{(i)}$ which proves the claim. The proofs for n_2 and the case of μ are similar.

An additional assumption about the symbol updates is required is order to derive an efficient algorithm for computation of the objectives. We define the following symbol index sets:

$$\begin{aligned} T^{(i)} &= \{n : m^{(i-1)} - B_q < n < n^{(i)} + B_q\} \\ U^{(i)} &= \{n : m^{(i-1)} - B_q < n \leq m^{(i)} - B_q\} \end{aligned} \tag{35}$$

Notice that the set $T^{(i)}$ is not empty because $n^{(i)} \geq m^{(i-1)}$ and $B_q \geq 1$, while the set $U^{(i)}$ may be empty.

v. At the ith iteration the algorithm updates the data symbol parameters $\{b_n : n \in T^{(i)}\}$.

This assumption is justified in practice because the objective for symbols not in this set does not change as a result of the most recent samples in $W^{(i)} - W^{(i-1)}$, so the only change in the objective of these symbols is a result of updating other parameters, which has a secondary effect. The w and a parameters on the other hand are always updated because they have a global effect on the likelihood.

The recursive implementation of the algorithm is given below. We assume that the algorithm is in the $i+1$st iteration, which means that the objectives for the wavelet index set $W^{(i+1)}$ are computed, and the initial state of the parameter estimates at the end of the ith iteration is denoted by the superscript (i), e.g. $z^{(i)}, b_n^{(i)}$, and $a^{(i)}$ for the time delay, data symbols, and gain respectively. The wavelet coefficients parameter w is updated first, then the data symbols b_n in increasing order, and finally the gain a. In each case the E step and the M step are indicated.

For $i = 0, 1, \ldots$

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

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

$$\mu_m^{(i+1)}(n) = \mu_m^{(i)}(n) + \sum_{jk \in D^{(i+1)}} y_{jk} c^*_{jk,m}(n), \quad m \in \mathbb{Z}, \quad n \in T^{(i+1)},$$

$$\nu_m^{(i+1)}(n_1,n_2) = \nu_m^{(i)}(n_1,n_2) + \sum_{jk \in D^{(i+1)}} d_{jk,m}(n_1,n_2),$$

$$m \in \mathbb{Z}, n_1, n_2 \in T^{(i+1)}, |n_1 - n_2| < 2B_q, \quad (36)$$

where $D^{(i+1)} = W^{(i+1)} - W^{(i)}$. The propositions stated above and the definition of the set $T^{(i+1)}$ guarantee that $\mu^{(i+1)}$ and $\nu^{(i+1)}$ are either zero or remain fixed for all other symbol indices not included in (36).

3. E step: For $\theta_S = w$ compute recursively the following:

$$\alpha_m^{(i+1)} = \alpha_m^{(i)} + \sum_{\substack{(n_1,n_2) \in V^{(i)} \\ |n_1-n_2|<2B_q}} b_{n_1}^{(i)} b_{n_2}^{(i)*} \nu_m^{(i)}(n_1,n_2),$$

$$\beta_m^{(i+1)} = \beta_m^{(i)} + \sum_{n \in U^{(i)}} b_n^{(i)*} \mu_m^{(i)}(n), \quad (37)$$

where

$$V^{(i)} = U^{(i)} \times T^{(i)} \cup T^{(i)} \times U^{(i)}.$$

Notice that $\alpha^{(i)}$ and $\beta^{(i)}$ accumulate the part of the objective which will not be changed, because $b_n^{(i)}$ remain fixed in all successive iterations. Next the objective is calculated by including the missing terms in each coefficient:

$$\gamma_m^{(i+1)} = |a^{(i)}|^2 \big[\alpha_m^{(i+1)} + \sum_{\substack{n_1,n_2 \in T^{(i+1)} \\ |n_1-n_2|<2B_q}} b_{n_1}^{(i)} b_{n_2}^{(i)*} \nu_m^{(i+1)}(n_1,n_2)\big],$$

$$\delta_m^{(i+1)} = a_l^{(i)*} \big[\beta_m^{(i+1)} + \sum_{n \in T^{(i+1)}} b_n^{(i)*} \mu_m^{(i+1)}(n)\big]. \quad (38)$$

Finally the penalized objective for $z = e^{i\omega_0 t}$ is given by:

$$\phi^{(i+1)}(z) = \frac{2}{N_0} \sum_{m \in \mathbb{Z}} (\delta_{-m}^{(i+1)} + \delta_m^{(i+1)*} - \gamma_m^{(i+1)}) z^m + \log \Pi_d(z). \quad (39)$$

M step: maximize the penalized objective,

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

4. E step: For $\theta_S = b_n, n \in T^{(i+1)}$ compute recursively the following:

$$\epsilon_m^{(i+1)}(n) = \epsilon_m^{(i)}(n) + \sum_{\substack{n_2 \in U^{(i)} \\ |n-n_2|<2B_q}} b_{n_2}^{(i)*} \nu_m^{(i)}(n,n_1). \quad (40)$$

Next the missing terms are included:

$$\varepsilon^{(i+1)}(n) = a^{(i)} \sum_{m \in \mathbb{Z}} \big[\mu_m^{(i+1)*}(n) - a^{(i)*} \big(\epsilon_m^{(i+1)}(n) + \sum_{\substack{n_2 \in T^{(i+1)} \\ n_2<n}} b_{n_2}^{(i+1)*} \nu_m^{(i+1)}(n,n_2) + \sum_{\substack{n_2 \in T^{(i+1)} \\ n_2>n}} b_{n_2}^{(i)*} \nu_m^{(i+1)}(n,n_2)\big)\big] z^{(i+1)m}. \quad (41)$$

Notice that symbols with index smaller than n have already assumed the updated value. Finally the objective for $\{b_n\}$ is given by:

$$Q^{(i+1)}(b_n) = \frac{2}{N_0} [b_n \varepsilon^{(i+1)}(n) + b_n^{-1} \varepsilon^{(i+1)*}(n)]. \quad (42)$$

M step:

- *Hard* decision: $b_n^{(i+1)} = \operatorname{sign}[\operatorname{Re}(\varepsilon^{(i+1)}(n))]$.
- *Soft* decision: maximize the penalized objective,

$$\phi^{(i+1)}(b_n) = Q^{(i+1)}(b_n) + \log \Pi_b(b_n),$$
$$b_n^{(i+1)} = \underset{b_n=e^{i\omega}}{\operatorname{argmax}} \{\phi^{(i+1)}(b_n)\}.$$

5. E step: For $\theta_S = a$ compute recursively the following:

$$\zeta^{(i+1)} = \sum_{m \in \mathbb{Z}} z^{(i+1)m} \big(\beta_{-m}^{(i+1)} + \sum_{n \in T^{(i+1)}} b_n^{(i+1)*} \mu_{-m}^{(i+1)}(n)\big),$$

$$\eta^{(i+1)} = \frac{2}{N_0} \sum_{m \in \mathbb{Z}} \big(\alpha_m^{(i+1)} + \sum_{\substack{n_1,n_2 \in T^{(i+1)} \\ |n_1-n_2|<2B_q}} b_{n_1}^{(i+1)} b_{n_2}^{(i+1)*} \nu_m^{(i+1)}(n_1,n_2)\big) z^{(i+1)m}. \quad (43)$$

The objective for a is given by:

$$Q^{(i+1)}(a) = a^* \zeta^{(i+1)} + a \zeta^{(i+1)*} - |a|^2 \eta^{(i+1)}. \quad (44)$$

M step: maximize the objective analytically,

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

VI. SIMULATIONS

The SAGE algorithm 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 mean squared error of the time delay and phase estimates.

In this paper only a small subset of all possible variations of system parameters has been tested. We have restricted our attention to an uncoded single-user system at the low SNR range. The results reported below were obtained from

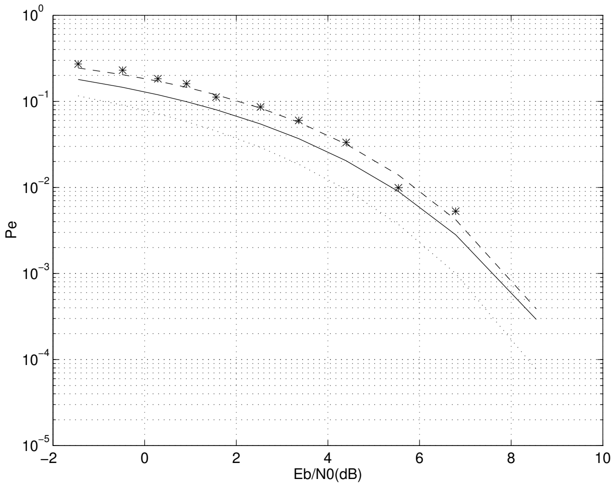

Fig. 1. Bit error probability vs. SNR, stars: simulation results; solid: DLL performance; dashed: DPSK error bound; dotted: PSK error bound.

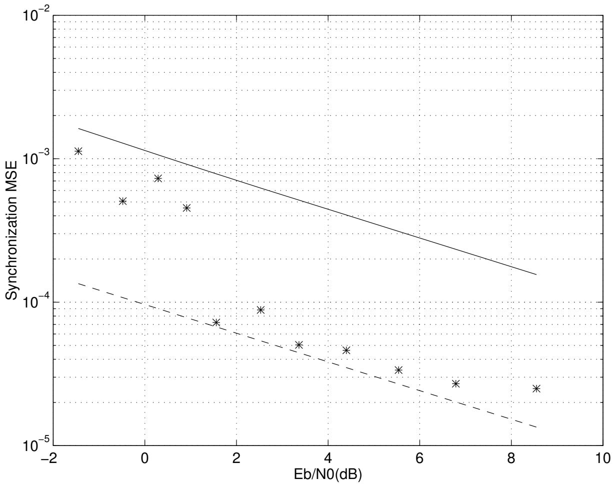

Fig. 2. Synchronization mean squared error vs. SNR, stars: simulation results; solid: DLL performance; dashed: CR bound.

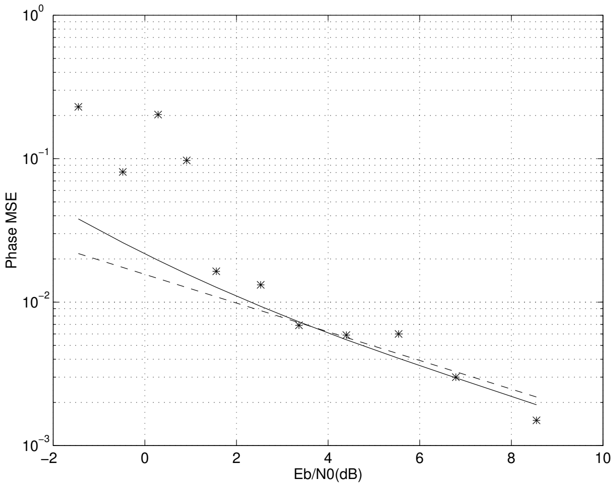

Fig. 3. Phase mean squared error, stars: simulation results; solid: DAL performance; dashed: CR bound.

100 Monte-Carlo simulations at each SNR value in the range of -1dB to 8dB. The signal parameters were chosen randomly and independently of each other. The data bits $\{b_n\}$ were selected as either $+1$ or -1. The gain a was of a fixed amplitude and random phase uniformly distributed in $[-\pi, \pi]$, and the time delay d was uniformly distributed in $[-T_p/2, T_p/2]$. In each run a 7 chip PN code was phase modulated by a 63 bit long sequence. The transmitted signal was then passed through a band limited AWGN channel, $|f| \leq 1/2T_c$, where T_c is the chip time. The algorithm was allowed to achieve timing and phase synchronization during the transmission of the first 31 bits. During the second part of the transmission, the time delay and gain estimates were kept fixed, and only the remaining bits were estimated. Thus, the bit error probability estimates were taken only during the synchronized state of the algorithm. The data bits were differentially decoded due to phase ambiguity. The Daubechies wavelets [1] 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 algorithm step, i.e. the size of $W^{(i)} - W^{(i-1)}$ was 7, and the number of wavelet samples per bit interval was approximately 31. The time delay penalty $\Pi_d(z)$ was designed by a third order Chebyshev filter. We remark that the SAGE algorithm may converge to a local maximum of the likelihood. This can be easily detected by a failure to synchronize within one chip time of the true value of d or by observing an objective value in (39) which is too small. Synchronization failures were rejected so as not to affect the estimates. The simulation results are shown in the figures below.

Figure 1 shows the bit error probability. We observe that the simulation results closely match the theoretical bound for a DPSK decoder. 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 [12]. Notice that the DLL performance has degraded by more than 1dB with respect to the ideal PSK error bound, while the SAGE performance is essentially optimal. Figure 2 shows the synchronization performance of the SAGE algorithm. The SAGE algorithm achieves a very small synchronization error at low SNR which explains its nearly optimal detection 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. Finally, Figure 3 compares the phase error variance of the SAGE algorithm with that of a Data Aided Loop (DAL). The CR bound on phase error variance is shown for reference. 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 [11]). The two systems have a similar performance, but it should be noted that the DAL relies on a perfectly synchronized reference.

VII. CONCLUSIONS

In this paper we have introduced a new method for joint timing and phase synchronization and optimal ML detection of transmitted symbols in a single user receiver. This method was derived by applying the principles of the SAGE algorithm. The use of the wavelet and Fourier transforms and a choice of parameter update strategy are the keys to the its simple recursive form. The flexibility in choosing several system parameters such as the wavelet basis, the step size of the algorithm, and the penalty functions, makes it suitable for a variety of applications depending on technological feasibility and cost considerations. We point out that the problem of selecting an appropriate wavelet basis has not been considered in this work. There exist however several algorithms which perform such a selection under various optimality criteria. We also observe that the algorithm does not deal effectively with situations in which the gain and delay parameters are time varying, so finding an adaptive version of the algorithm could be very useful both from a practical and a theoretical point of view. Finally, the multi-user problem will be dealt with in a forthcoming paper.

References

[1] I. Daubechies, "Orthonormal bases of compactly supported wavelets", Comm. Pure Appl. Math., 41 (1988), no. 7, 909–996.

[2] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis", IEEE Trans. IT, 36(5):961-1005, Sep. 1990.

[3] I. Daubechies, *Ten lectures on wavelets*, Philadelphia, PA: SIAM, 1992.

[4] 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.

[5] M. Feder, E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm", IEEE Trans. ASSP, 36(4):477-489, Apr. 1988.

[6] J.A. Fessler, A.O. Hero, "Space-alternating generalized expectation-maximization algorithm", IEEE Trans. ASSP, 42(10):2664-2677, Oct. 1994.

[7] 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.

[8] 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.

[9] C.N. Georghiades, D.L. Snyder, "The expectation-maximization algorithm for symbol unsynchronized sequence detection", IEEE Trans. COM, 39(1):54-61, Jan. 1991.

[10] 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.

[11] W.C. Lindsey, M.K. Simon, "Data-aided carrier tracking loops", IEEE Trans. COM, 19(2):157-168, Apr. 1971.

[12] W.C. Lindsey, M.K. Simon, *Telecommunication systems engineering*, Englewood Cliffs, NJ: Prentice-Hall, 1973.

[13] S.G. Mallat, "Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$", Trans. Amer. Math. Soc., 315(1):69–87, 1989.

[14] 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.

[15] 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.

[16] A.V. Oppenheim, R.W. Schafer, *Discrete-time signal processing*, Englewood Cliffs, NJ: Prentice Hall, 1989.

[17] M.A. Tanner, *Tools for Statistical Inference : Methods for the exploration of posterior distributions and likelihood functions*, New York: Springer-Verlag, 1993.

[18] J.R. Thompson, R.A. Tapia, *Nonparametric function estimation, modeling, and simulation*, Philadelphia, PA: SIAM, 1990.