Code-aided Bayesian Parameter Estimation for Multi-Carrier Systems

M. Guenach, F. Simoens, H. Wymeersch*, H. Steendam, M. Moeneclaey

Department of Telecommunications and Information Processing

Ghent University,

St.-Pietersnieuwstraat 41, B-9000 GENT, Belgium

TEL: +32 9 264 8900; Fax: +32 9 264 4295

E-mail: {guenach,fsimoens,hs,mm}@telin.UGent.be

*LIDS Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA

Email: hwymeers@mit.edu

March 30, 2006

Abstract

In this paper, we present a theoretical framework for code-aided synchronization and channel estimation for multi-carrier (MC) systems. Our endeavor is to exploit all available information from the received signal for the estimation of the unknown parameters. By means of the Expectation-Maximization (EM) formalism, we will show how soft information of the coded symbols computed by the detector can be combined with pilot symbols in a systematic and efficient manner in order to improve the overall system performance. The resulting code-aided algorithm comes with a negligible increase in processing demands compared to its data-aided counterpart, since the synchronizer and estimator can be embedded in the iterative 'turbo' detector. For sake of simplicity, we develop our estimation framework for a simple orthogonal frequency division multiplexing (OFDM) transceiver. As far as the estimation is concerned, extension to other multi-carrier set-ups using more advanced detectors is conceptually very straightforward. Computer simulations show that the proposed algorithm outperforms the conventional data-aided estimation algorithms.

Index terms: OFDM, synchronization, turbo estimation, turbo synchronization

1   Introduction

Orthogonal frequency division multiplexing (OFDM) is put forward as the obvious modulation scheme for current and next-generation wireless and wireline communications [1,2]. Its capability to accommodate high data rate transmissions on frequency selective channels while allowing for a simple equalization, makes it a very promising technology.

On the other hand, OFDM-based transmission systems are very sensitive to synchronization errors [3,4]. Accurate timing and frequency synchronization is required to maintain orthogonality among subcarriers and to avoid inter-symbol-interference (ISI). Additionally, the channel impulse response (CIR) must be known to coherently detect the data per subcarrier. When we consider the massive amount of research devoted to these problems (see e.g. [4–13] and references therein), it becomes clear that synchronization and channel estimation are critical issues. The conventional estimation techniques are either *data-aided* (i.e., exploiting training symbols in the time- or frequency domain) [5–10] or *blind* (e.g., exploiting the presence of the cyclic prefix) [11–13].

With the advent of powerful error-correcting codes (the likes of turbo-codes), these conventional techniques cannot always be applied successfully. Powerful codes lead to a combination of low BER at low SNR: a challenging starting point to perform estimation. At these low SNRs, blind techniques are completely unreliable. On the other hand, data-aided algorithms require an unreasonable amount of power and bandwidth to be reserved for training. This has spurred several research groups to consider 'code-aided' or 'code-aware' estimation algorithms. These algorithms iterate between data detection and estimation, thus improving both the estimates of synchronization parameters and CIR, while simultaneously performing increasingly reliable data detection. Such techniques are often inspired by the turbo-principle [14] or the Expectation-Maximization (EM) algorithm [15,16]. In [17], an EM-based semi-blind technique is described that performs code-aided estimation of the CIR per Multi-Carrier (MC) symbol. The same idea was applied in the frequency domain in [18] for a multi-antenna, multi-user system. The EM algorithm was again considered in [19] for estimation of the CIR for a time-varying MIMO-OFDM scenario. Finally, [20] proposes an ad-hoc code-aided channel estimator for time-varying OFDM systems. The reader will note that code-aided estimation of synchronization parameters has received little interest.

In the current paper, we address code-aided frequency synchronization, frame synchronization and channel estimation for multi-carrier communication schemes. Our analysis is carried out for a plain single-user multi-carrier communication scheme. Nevertheless, as we will elaborate upon, the actual configuration is of marginal importance as far as the estimation is

concerned. The proposed estimator can easily be extended to other multi-carrier set-ups.

We compare a Bayesian estimation approach based on the Maximum A Posteriori (MAP) principle with a non-Bayesian estimation approach based on the Maximum Likelihood (ML) principle. We resort to the EM algorithm to exploit information from the pilot symbols and coded data symbols in a systematic fashion.

This paper is organized as follows. The system model is presented in section 2, including a brief description of the detector. We continue with a brief description of ML, MAP and EM estimation algorithms in section 3. Then we apply these algorithms to the problem at hand and derive an EM-based estimator, paying special attention to issues related to computational complexity. Numerical results are provided in section 5 before we draw conclusions in section 6.

2 OFDM system model

2.1 Transmitter

To develop our code-aided estimation framework, we consider the simple transmitter scheme depicted in Fig. 1. We assemble a (turbo-)encoder, block-interleaver and symbol mapper to set up a so-called Bit-Interleaved Coded Modulation (BICM) scheme. The BICM scheme transforms N_b information bits $b_0, \ldots, b_{N_b-1}$ into N_d coded symbols which belong to a unit-energy 2^q-point constellation Ω. The coded symbol sequence, denoted $\mathbf{a} = [a_0, \ldots, a_{N_d-1}]^T$, is multiplexed with N_t pilots symbols. The p-th element d_p of the overall block of pilot and data symbols $\mathbf{d} = \left[p_0, \ldots, p_{N_t-1}, \mathbf{a^T}\right]^T$, is mapped to the carrier n_p where $0 \leq n_p \leq N - 1$ with $N = N_t + N_a$. This operation is known as frequency interleaving. We end up with a block of N interleaved symbols which is applied to an N-point Inverse Discrete Fourier Transform[1] (IDFT). A ν-point cyclic prefix (CP) is pre-appended, resulting in $N + \nu$ time-domain samples $[s_{-\nu}, \ldots, s_{-1}, s_0, \ldots, s_{N-1}]^T$ where $s_m = s_{m+N}$, for $m = -\nu, \ldots, -1$. Hence, the length of the time-domain sequence is equal to $N_T = N + \nu$. We can write the m-th time-domain sample $(m = -\nu, \ldots, N - 1)$ as

$$s_m = \sqrt{\frac{E_s}{N + \nu}} \sum_{p=0}^{N-1} d_p e^{j2\pi n_p m/N} \tag{1}$$

where E_s denotes the average energy per symbol d_p. Note that by construction, each OFDM symbol encompasses a single codeword. We further define the OFDM symbol period T and the

---

[1]Obviously, the DFT and IDFT operations are implemented through a Fast Fourier Transform.

notation	meaning
m	time-domain index: $-\nu \leq m < N$
p	data symbol index (for a given OFDM symbol): $0 \leq p < N$
ξ	EM iteration index: $\xi \geq 1$

Table 1: Notations

sampling period $T_s = T/N_T$. Finally, the signal is shaped with a normalized transmit filter and transmitted over the channel.

2.2 Receiver

The transmitted signal propagates through a channel with overall Channel Impulse Response (CIR) $h_{ch}(t)$. This CIR incorporates the transmit filter, physical propagation channel and receive filter (e.g., a filter matched to the transmitter pulse). We assume a quasi-static block-fading channel that remains constant during each burst but can vary independently from burst to burst. The CIR is assumed to have a delay spread no greater than LT_s: $h_{ch}(t) = 0$ for $t < 0$ and for $t \geq LT_s$. Additionally, the signal arrives with a certain delay τ and is affected by a CFO f_o and is corrupted by thermal noise. The propagation delay τ belongs to some interval $[0, \tau_{max}]$, while the CFO depends on the speed of the transmitter and any possible mismatch between the transmit and receive oscillators. We assume the carrier frequency offset (CFO) to be small, compared to the bandwidth of the receiver's matched filter. Hence, the received signal is given by

$$r(t) = s(t) + w(t) \tag{2}$$

where $w(t)$ is the baseband representation of the Additive White Gaussian Noise (AWGN) with power spectral density $N_0/2$ per real dimension. In (2), the signal $s(t)$ is given by

$$s(t) = \sum_{m=-\nu}^{N-1} s_m h_{ch}(t - mT_s - \tau) e^{j2\pi f_o t}. \tag{3}$$

The receiver is fully digital and samples the received signal $r(t)$ at a rate $1/T_s$ resulting in a sequence of samples $r(lT_s)$, $l = -\nu, \ldots, +\infty$. Let us define the normalized CFO $F = f_o T_s N$. Following [5], we break up τ as

$$\tau = \Delta T_s - \delta T_s \tag{4}$$

with $\Delta \in \{0, 1, \ldots, \Delta_{max} \doteq \lceil \tau_{max}/T_s \rceil\}$ and $\delta \in [0, 1[$. Defining $h(t) = h_{ch}(t + \delta T_s)$, we can express the samples $r(lT_s)$ as a function of $h(t)$ as

$$r(lT_s) = \sum_{m=-\nu}^{N-1} s_m h(lT_s - mT_s - \Delta T_s) e^{j2\pi Fl/N} + w(lT_s). \tag{5}$$

Since $T = N_T T_s$ is a multiple of T_s, the channel is fully characterized by the vector

$$\mathbf{h} = [h(0), h(T_s), \ldots, h((L-1)T_s)]^T. \tag{6}$$

The channel taps are assumed to be gaussian distributed with $L \times L$ covariance matrix $C_{\mathbf{h}}$. When $\nu \geq L - 1$, no interference between successive MC symbols occurs. As $\Delta \in \{0, 1, \ldots, \Delta_{max}\}$, samples taken prior to $t = -\nu T_s$ and later than $t = (T + (\Delta_{max} + L - \nu - 2)T_s)$ do not depend on the current frame. Hence, the following sequence of $N_T + \Delta_{max} + L - 1$ time-domain samples is sufficient for data detection and estimation:

$$\mathbf{r} = [r(-\nu T_s), \ldots, r(T + (\Delta_{max} + L - \nu - 2)T_s)]^T. \tag{7}$$

The ultimate goal of the receiver is to recover the transmitted information bits $\mathbf{b}$. In order to do so, knowledge of the delay shift Δ, the (normalized) CFO F and the channel taps $\mathbf{h}$ is required.

The conceptual block diagram of the receiver is shown in Fig. 2. The receiver employs an estimation algorithm (e.g., by exploiting the pilot symbols $[p_0, \ldots, p_{N_t-1}]^T$) to obtain estimates $(\hat{\Delta}, \hat{F}, \hat{\mathbf{h}})$ for the timing offset Δ, the CFO F and the channel taps $\mathbf{h}$. After synchronization, the CP is stripped and the remaining samples applied to a DFT. The output of the DFT at subcarrier n_p is given by

$$z_p = \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} r(\hat{\Delta} T_s + lT_s) e^{j2\pi ln_p/N} e^{-j2\pi \hat{F}(\hat{\Delta}+l)}. \tag{8}$$

The detector assumes that the channel and synchronization parameters are perfectly known and equal to their estimated values. By equalization of the samples $\{z_{p+N_t}\}_{p=0,\ldots,N_a-1}$, we create a decision variable for a_p:

$$y_p = \alpha_{p+N_t} z_{p+N_t} \tag{9}$$

$$= a_p \mu_p + v_p \tag{10}$$

where α_p is the coefficient of the one-tap equalizer at subcarrier n_p. Based on (10), the receiver

is able to compute the likelihood $p(y_p|a_p)$ of each transmitted symbol. These likelihoods are fed to a block that performs soft demodulation and decoding, operating according to the turbo principle [21, 22]. For our purposes, this block is considered as a black box that computes a posteriori probabilities (APPs), possibly in an iterative fashion. Upon completion, the APPs of b_m (i.e, $p\left(b_m \left| \mathbf{r}, \hat{\mathbf{h}}, \hat{\Delta}, \hat{F}\right.\right)$) $m = 0, \ldots, N_b - 1$ are used to make final decisions on the information bits. At the same time, this block also computes the APPs of the coded symbols $p\left(a_m \left| \mathbf{r}, \hat{\mathbf{h}}, \hat{\Delta}, \hat{F}\right.\right)$. Note that, while we considered a specific OFDM configuration, the estimation framework that we are about to develop can be applied to any detection scheme as long as a soft-decoding algorithm is available.

3 Estimation algorithms

It is clear that the detector described in the previous section requires the estimates of the delay/frame shift[2] Δ, the CFO F and the channel taps $\mathbf{h}$. The estimation should be carried out without knowledge of the transmitted symbols. Hence, for the estimation purpose, the unknown symbols are *nuisance* parameters. In this section we introduce the basic algorithms that allow us to tackle this estimation problem. We start off with the standard ML and MAP estimators, and end with the introduction of the (Bayesian) EM algorithm.

3.1 ML and MAP estimation

We consider two possible techniques to compute the estimates of one or more unknown parameters: ML or MAP estimation. Computing the MAP or ML estimates entails two - at first sight simple - steps. In the first step, the cost functions $p\left(r\left|\theta\right.\right)$ (ML) or $p\left(\theta|r\right)$ (MAP) are determined and evaluated. In the second step, these functions are maximized with respect to the unknown parameters. Depending on the situation, both steps can be cumbersome. As we consider the estimation of a parameter(set) θ from an observation r in the presence of a (discrete) nuisance parameter s, these two steps lead up to the following estimates

$$\hat{\theta}_{ML} = \arg\max_{\theta} \sum_{s} p\left(r\left|\theta, s\right.\right) p\left(s\left|\theta\right.\right) \tag{11}$$

$$\hat{\theta}_{MAP} = \arg\max_{\theta} \sum_{s} p\left(r, \theta\left|s\right.\right) p\left(s\right) \tag{12}$$

When dealing with a large or unstructured parameter space, the (multidimensional) maximization (second step) is often difficult. This problem, can be alleviated using signal decomposition

---

[2]In correspondence with technical literature, we will name the process of determining Δ *frame synchronization.*

methods [23]. Only too often, however, the nuisance parameters itself form the major source of concern. When the nuisance parameter space is extensive, the summation involved in the evaluation of (11) or (12) tends to be intractable. To circumvent this problem, we will resort to the iterative expectation-maximization (EM) algorithm.

3.2 EM estimation

The Expectation-Maximization (EM) algorithm is a method that iteratively solves the ML (11) or MAP (12) problem [15]. We will describe the MAP version of the EM algorithm (also known as the Bayesian EM algorithm) [24, 25]. We will point out where the ML solution differs from the MAP solution.

In accordance with the EM formalism, we define the *complete data* z, which is related to the observation r through some mapping $r = g(z)$. The EM algorithm starts from an initial estimate of θ (say, $\hat{\theta}(0)$) and iteratively computes new estimates. At iteration ξ, the EM algorithm consists of two steps: given the current estimate $\hat{\theta}(\xi)$, we first take the expectation of the log-likelihood function of the complete data, given the observation $\mathbf{r}$ and the current estimate of θ:

$$\text{E-step: } Q\left(\theta | \hat{\theta}(\xi)\right) = E_z\left[\log p(z, \theta) \middle| r; \hat{\theta}(\xi)\right]. \tag{13}$$

Note that $\log p(z, \theta)$ should be replaced with $\log p(z|\theta)$ to obtain the ML solution. When z is well-chosen, evaluation of (13) is fairly simple compared to the computations required in (11) or (12). In the second step, we maximize $Q\left(\theta | \hat{\theta}(\xi)\right)$ with respect to θ to find a new estimate:

$$\text{M-step: } \hat{\theta}(\xi + 1) = \arg\max_\theta \left\{Q\left(\theta | \hat{\theta}(\xi)\right)\right\}. \tag{14}$$

Convergence of the EM algorithm is guaranteed in a sense that subsequent estimates have a non-decreasing a posteriori probability:

$$p\left(\hat{\theta}(\xi + 1) | r\right) \geq p\left(\hat{\theta}(\xi) | r\right) \tag{15}$$

for $\xi \geq 0$. In the ML version of the EM algorithm, the a posteriori probabilities in (15) are replaced with likelihoods. Any value $\hat{\theta}$ for which $\hat{\theta} = \arg\max_\theta Q\left(\theta | \hat{\theta}\right)$ is called a *solution* of the EM algorithm. One of these solutions is the MAP (or ML) estimate. In order to achieve convergence to the true MAP (or ML) estimate, a good initial estimate of θ is required [25].

4 Code-aided Channel estimation and synchronization

In this section, we will derive the ML and MAP estimators for the code-aided estimation problem at hand. Consistent with the notations from previous section, we consider the estimation parameters $\theta = [\mathbf{h}, \Delta, F]$, the observation $r = \mathbf{r}$ and the nuisance parameters $s = \mathbf{s}$. As the direct computation of the ML and MAP estimates turns out to be infeasible, we propose a new EM based estimator.

4.1 ML and MAP estimation

In order to derive the ML or MAP cost functions, we rewrite the observation model in a more convenient form. We start from the observation $\mathbf{r} = [r\left(-\nu T_s\right), \ldots, r\left(T + \left(\Delta_{max} + L - \nu - 2\right) T_s\right)]^T$ from (7). Note that the length of this vector is independent of Δ. We introduce the vector $\mathbf{s}$ of length $(N_T + 2L - 2)$ constructed by concatenating all time-domain samples, padded with $L - 1$ zeros at the beginning and end of the vector, leading to

$$\mathbf{s} \doteq \left[\mathbf{0}_{1 \times (L-1)} \, \mathbf{s}_{-\nu} \, \ldots \, \mathbf{s}_{N-1} \ldots \, \mathbf{0}_{1 \times (L-1)}\right] \tag{16}$$

where $\mathbf{0}_{X \times Y}$ is an $X \times Y$ matrix consisting of all zeros. Note that $\mathbf{s}$ depends solely on the pilot symbols and coded symbols $\mathbf{d}$.

Let us now define the $(N_T + L - 1) \times L$ Toeplitz matrix $\mathbf{S}$ as follows: the n-th row of $\mathbf{S}$ is obtained by time-reversing the n-th until the $(n + L - 1)$-th sample of $\mathbf{s}$. For instance, the first row is given by $\left[s_{-\nu}, \mathbf{0}_{1 \times (L-1)}\right]$, the second row by $\left[s_{-\nu+1}, s_{-\nu}, \mathbf{0}_{1 \times (L-2)}\right]$ and so forth. Note that we can write $\mathbf{S}$ as the sum of two Toeplitz matrices of size $(N_T + L - 1) \times L$: $\mathbf{S} = \mathbf{S}_P + \mathbf{S}_A$, where $\mathbf{S}_P$ contains only the pilot symbols and $\mathbf{S}_A$ contains only the data symbols.

Finally, we define an $(N_T + \Delta_{max} + L - 1) \times L$ matrix $\mathbf{S}_\Delta$ as

$$\mathbf{S}_\Delta = \begin{bmatrix} \mathbf{0}_{\Delta \times L} \\ \mathbf{S} \\ \mathbf{0}_{(\Delta_{max} - \Delta + L - 1) \times L} \end{bmatrix}. \tag{17}$$

These transformations enable us to write the following simple relationship between $\mathbf{r}$, $\mathbf{S}_\Delta$ and $\mathbf{h}$:

$$\mathbf{r} = \mathbf{F}\,\mathbf{S}_\Delta \mathbf{h} + \mathbf{w} \tag{18}$$

where $\mathbf{F}$ is a diagonal matrix with $\mathbf{F}_{kk} = e^{j2\pi F k/N}$, $k = -\nu, \ldots, N_T + \Delta_{max} + L - \nu - 2$ and the vector $\mathbf{w}$ embeds the zero-mean AWGN with variance σ_n^2 per real dimension. Note that

by substituting $\mathbf{S} = \mathbf{S}_P + \mathbf{S}_A$ into (17), we can break up $\mathbf{S}_\Delta = \mathbf{S}_{\Delta,P} + \mathbf{S}_{\Delta,A}$. Based on the observation model (18), we can easily derive the ML and MAP cost functions.

The ML estimate of the delay shift and channel taps is obtained by maximizing the log-likelihood function

$$\left[\hat{\Delta}_{ML}, \hat{F}_{ML}, \hat{\mathbf{h}}_{ML}\right] = \arg \max_{\Delta, F, \mathbf{h}} \left\{\log p\left(\mathbf{r} \,|\, \Delta, F, \mathbf{h}\right)\right\} \tag{19}$$

where

$$p\left(\mathbf{r} \,|\, \Delta, F, \mathbf{h}\right) \propto \sum_{\mathbf{s}} p\left(\mathbf{r} \,|\, \Delta, F, \mathbf{h}, \mathbf{s}\right) p\left(\mathbf{s}\right) \tag{20}$$

and

$$p\left(\mathbf{r} \,|\, \Delta, F, \mathbf{h}, \mathbf{s}\right) \propto \exp\left(-\frac{1}{2\sigma_n^2} \left\|\mathbf{r} - \mathbf{F}\,\mathbf{S}_\Delta \mathbf{h}\right\|^2\right). \tag{21}$$

The MAP estimate, on the other hand, is given by

$$\begin{aligned}
\left[\hat{\Delta}_{MAP}, \hat{F}_{MAP}, \hat{\mathbf{h}}_{MAP}\right] &= \arg \max_{\Delta, F, \mathbf{h}} \left\{\log p\left(\mathbf{r}, \Delta, F, \mathbf{h}\right)\right\}, \\
&= \arg \max_{\Delta, F, \mathbf{h}} \left\{\log p\left(\mathbf{r} \,|\, \Delta, F, \mathbf{h}\right) + \log p\left(\mathbf{h}\right)\right\}.
\end{aligned} \tag{22}$$

We assume a conjugate prior [26] distribution of the channel taps. In this present case, the conjugate prior is a gaussian distribution

$$p\left(\mathbf{h}\right) \propto \exp\left(-\mathbf{h}^H C_{\mathbf{h}}^{-1} \mathbf{h}\right). \tag{23}$$

Note that in general the frequency offset F and delay Δ are uniformly distributed and impart no additional prior information.

As we already anticipated, the ML or MAP estimates cannot be recovered from (19) or (22) with reasonable complexity due to the presence of nuisance parameters. The summation in (20) requires the evaluation of (21) over all 2^{N_b} possible codewords, which unfortunately turns out to be infeasible. It is obvious that the summation in (20) would disappear if the unknown data was ignored and only pilot symbols where applied for estimation purpose. However, in the low SNR regimes where current state-of-the-art coding schemes prosper, too many pilot symbols are required to acquire reliable estimates. As this results in a significant loss in terms of power and bandwidth, there is great interest in developing algorithms that are also able to exploit the unknown data symbols for estimation purpose. In the following section, we adopt the EM algorithm introduced in section 3.2.

4.2 EM estimation

For the practical implementation of the EM algorithm, we will restrict ourselves again to the MAP scenario. The ML solution is touched upon at the end of this paragraph. Let us take as complete data $\mathbf{z} = [\mathbf{r}, \mathbf{s}]$. In that case, since $\mathbf{s}$ and θ are independent:

$$\log p (\mathbf{z}, \theta) = \log p (\mathbf{r}, \theta | \mathbf{s}) + \log p (\mathbf{s}) \tag{24}$$

so that (13) becomes

$$Q \left(\theta | \hat{\theta} (\xi) \right) \propto E_{\mathbf{s}} \left[\log p (\mathbf{r}, \theta | \mathbf{s}) | \mathbf{r}, \hat{\theta} (\xi) \right], \tag{25}$$

where we ignored terms that do not depend on θ. From (21) and (22) we have

$$\log p (\mathbf{r}, \theta | \mathbf{s}) \propto -\frac{1}{2\sigma_n^2} \mathbf{h}^H \mathbf{S}_\Delta^H \mathbf{S}_\Delta \mathbf{h} + \frac{1}{\sigma_n^2} \Re \left(\mathbf{r}^H \mathbf{F} \mathbf{S}_\Delta \mathbf{h} \right) - \mathbf{h}^H C_{\mathbf{h}}^{-1} \mathbf{h} \tag{26}$$

so that

$$Q \left(\theta | \hat{\theta} (\xi) \right) = -\mathbf{h}^H \left(\widetilde{\mathbf{S}_\Delta^H \mathbf{S}_\Delta} + 2\sigma_n^2 C_{\mathbf{h}}^{-1} \right) \mathbf{h} + 2\Re \left(\mathbf{r}^H \mathbf{F} \widetilde{\mathbf{S}}_\Delta \mathbf{h} \right) \tag{27}$$

where, due to the linearity of the expectation operator, $\widetilde{\mathbf{S}}_\Delta = E_{\mathbf{s}} \left[\mathbf{S}_\Delta | \mathbf{r}, \hat{\theta} (\xi) \right]$ is obtained by replacing each entry s_m with the corresponding a posteriori expectation $E \left[s_m | \mathbf{r}, \hat{\theta} (\xi) \right]$. Similarly, $\widetilde{\mathbf{S}_\Delta^H \mathbf{S}_\Delta} = E_{\mathbf{s}} \left[\mathbf{S}_\Delta^H \mathbf{S}_\Delta \middle| \mathbf{r}, \hat{\theta} (\xi) \right]$ is obtained by replacing the entries $s_m (s_{m'})^*$ with $E \left[s_m (s_{m'})^* | \mathbf{r}, \hat{\theta} (\xi) \right]$. Equation (1) tells us that

$$s_m = \sqrt{\frac{E_s}{N + \nu}} \sum_{p=0}^{N-1} d_p e^{j2\pi n_p m / N} \tag{28}$$

so that

$$E \left[s_m | \mathbf{r}, \hat{\theta} (\xi) \right] = \sqrt{\frac{E_s}{N + \nu}} \sum_{p=0}^{N-1} E \left[d_p | \mathbf{r}, \hat{\theta} (\xi) \right] e^{j2\pi n_p m / N}. \tag{29}$$

From section 2, we know that the detector computes the APPs of the coded symbols $p \left(d_p \middle| \mathbf{r}, \hat{\theta} (\xi) \right)$. The a posteriori expectation $E \left[d_p | \mathbf{r}, \hat{\theta} (\xi) \right]$ is obtained as

$$E \left[d_p | \mathbf{r}, \hat{\theta} (\xi) \right] = \sum_{\omega \in \Omega} \omega \times p \left(d_p = \omega \middle| \mathbf{r}, \hat{\theta} (\xi) \right) \tag{30}$$

which can be interpreted as a soft symbol decision: it is a weighted average of all possible constellation points. Note that if d_p corresponds to a pilot symbol, (30) is replaced by the actual value of this pilot symbol. As $E \left[s_m (s_{m'})^* \middle| \mathbf{r}, \hat{\theta} (\xi) \right]$ is a function of $E \left[d_p (d_{p'})^* \middle| \mathbf{r}, \hat{\theta} (\xi) \right]$

it cannot be computed exactly based solely on the (marginal) APPs. However, thanks to the presence of the interleaver, the coded symbols can be assumed to be essentially uncorrelated so that

$$E\left[d_p \left(d_{p'}\right)^* \Big| \mathbf{r}, \hat{\theta}\left(\xi\right)\right] \approx E\left[d_p \Big| \mathbf{r}, \hat{\theta}\left(\xi\right)\right] E\left[\left(d_{p'}\right)^* \Big| \mathbf{r}, \hat{\theta}\left(\xi\right)\right] \tag{31}$$

when $p' \neq p$. Additionally, if we extend the previous relation to $p = p'$, the computation of $\widetilde{\mathbf{S}_\Delta^H \mathbf{S}_\Delta}$ can further be simplified as

$$\widetilde{\mathbf{S}_\Delta^H \mathbf{S}_\Delta} \approx \widetilde{\mathbf{S}}_\Delta^H \widetilde{\mathbf{S}}_\Delta. \tag{32}$$

Finally, the updated estimates of the delay shift, the CFO and the channel taps are given by

$$\left[\hat{\Delta}\left(\xi+1\right), \hat{F}\left(\xi+1\right)\right] = \arg\max_{\Delta,F} \left\{\Re\left(\mathbf{r}^H \mathbf{F} \widetilde{\mathbf{S}}_\Delta \left(\widetilde{\mathbf{S}^H \mathbf{S}} + 2\sigma_n^2 C_{\mathbf{h}}^{-1}\right)^{-1} \widetilde{\mathbf{S}}_\Delta^H \mathbf{F}^H \mathbf{r}\right)\right\}. \tag{33}$$

and

$$\hat{\mathbf{h}}\left(\xi+1\right) = \left(\widetilde{\mathbf{S}^H \mathbf{S}} + 2\sigma_n^2 C_{\mathbf{h}}^{-1}\right)^{-1} \widetilde{\mathbf{S}}_{\hat{\Delta}(\xi+1)}^H \hat{\mathbf{F}}(\xi+1)^H \mathbf{r} \tag{34}$$

where $\widetilde{\mathbf{S}^H \mathbf{S}} = E_{\mathbf{s}}\left[\mathbf{S}^H \mathbf{S} \Big| \mathbf{r}, \hat{\theta}\left(\xi\right)\right]$. Making use of the approximation (32), this leads to a very elegant interpretation: the EM-based algorithms are formally obtained by replacing in the corresponding DA algorithms, pilot symbols with a posteriori symbol expectations.

Note that the 2-dimensional maximization in (33) can be reduced to two 1-dimensional maximizations when the CFO is small [5]. Therefore, the low-complexity update rules for the timing and the CFO are

$$\left[\hat{\Delta}\left(\xi+1\right)\right] = \arg\max_{\Delta} \left\{\Re\left(\mathbf{r}^H \hat{\mathbf{F}}\left(\xi\right) \widetilde{\mathbf{S}}_\Delta \left(\widetilde{\mathbf{S}^H \mathbf{S}} + 2\sigma_n^2 C_{\mathbf{h}}^{-1}\right)^{-1} \widetilde{\mathbf{S}}_\Delta^H \hat{\mathbf{F}}^H\left(\xi\right) \mathbf{r}\right)\right\} \tag{35}$$

and

$$\left[\hat{F}\left(\xi+1\right)\right] = \arg\max_{F} \left\{\Re\left(\mathbf{r}^H \mathbf{F} \widetilde{\mathbf{S}}_{\hat{\Delta}(\xi+1),} \left(\widetilde{\mathbf{S}^H \mathbf{S}} + 2\sigma_n^2 C_{\mathbf{h}}^{-1}\right)^{-1} \widetilde{\mathbf{S}}_{\hat{\Delta}(\xi+1),}^H \mathbf{F}^H \mathbf{r}\right)\right\}. \tag{36}$$

Note that ML estimation is obtained by replacing $\log p\left(\mathbf{r}, \theta | \mathbf{s}\right)$ with $\log p\left(\mathbf{r} | \theta, \mathbf{s}\right)$ in (25). This boils down to setting $C_{\mathbf{h}}^{-1} = 0$ in equations (33)-(36).

computation	complexity
$\tilde{\mathbf{S}}_\Delta$	$\mathcal{O}\left(N\log_2 N\right)$
$\widetilde{\mathbf{S}^H\mathbf{S}}$	$\mathcal{O}\left(N_T L\right)$
update of $\hat{\Delta}$	$\mathcal{O}\left(N_T L\Delta_{max}\right)$
update of $\hat{\mathbf{h}}$	$\mathcal{O}\left(N_T L\right)$

Table 2: Computational complexity, assuming $\widetilde{\mathbf{S}^H\mathbf{S}}$ is a diagonal matrix. We have omitted the complexity related to updating $\hat{F}$ as it depends on the specific numerical technique used to solve (36).

4.3   Implementation aspects

The proposed EM estimator can be modified in several ways:

1. The inversion of the matrix $\widetilde{\mathbf{S}^H\mathbf{S}}$ in (33)-(34) can be simplified when the symbol sequence is sufficiently long, such that $\widetilde{\mathbf{S}^H\mathbf{S}}$ can be approximated by a diagonal matrix. A detailed view of the computational complexity in the general case is given in Table 2. Observe that the total computational cost is dominated by the Δ-update.

2. When the estimates of Δ and F are sufficiently accurate and no further updates are required, all computations (with the exception of (34)) can take place in the frequency-domain.

3. In case the delay Δ and the CFO F are perfectly known, the code-aided estimation algorithm of the channel impulse response can take place completely in the frequency domain, thus avoiding FFT operations at each EM iteration. This requires a specific frequency-domain observation model [27]. This is not pursued in the current paper.

Even with these complexity-reducing modifications, the complexity of the EM-based estimator may still be unacceptably high: let us denote by T_{EM} the time (in seconds) to compute the cost-function (27) and perform a single update to the estimates of F, Δ and $\mathbf{h}$ and by T_{detect} the time to detect the data, given an estimate[3] of F, Δ and $\mathbf{h}$. When I_{EM} EM iterations are performed, the total computation time is roughly $T_{tot} = I_{EM}T_{EM} + (I_{EM}+1)T_{detect}$, so that the overhead related to estimation is given by

$$O_{EM} = \frac{T_{tot} - T_{detect}}{T_{detect}} = I_{EM}\left(1 + \frac{T_{EM}}{T_{detect}}\right)$$

which is (at least) an I_{EM}-fold complexity increase as compared to a conventional system with (non-iterative) estimation followed by data detection (corresponding to an overhead of

---

[3]Hence, T_{detect} represents the processing time for a perfectly synchronized detector with perfect channel knowledge.

0). Currently, many state-of-the-art detectors or decoders operate according to some iterative procedure. Each iteration within the detector may correspond for instance to a demapping iteration or decoding iteration or both. In either way, we may write $T_{detect} = I_D T_D$, where I_D is the number of iterations performed within the detector, and T_D the time to perform a single iteration. We now perform the following modifications: for each EM iteration, we perform $I_D^* < I_D$ iterations within the detector, but we maintain state information[4] from one iteration to the next. This is known as *embedded* estimation [28]. In this case the total computation time is roughly $T_{tot} = I_{EM} T_{EM} + (I_{EM} + 1) I_D^* T_D$. Suppose we perform roughly $I_{EM} = I_D - 1$ EM iterations, then the overhead related to EM estimation is now

$$O_{EM} = \frac{I_{EM}}{I_{EM} + 1} \times \frac{T_{EM}}{T_D} + I_D^* - 1$$

In general, we have that $T_{EM} \ll T_D$ since T_D usually involves a complex decoding step. Hence, if we set $I_D^* = 1$, which corresponds to one detection iteration per EM iteration, the overhead caused by the estimation is marginal. Therefore, embedded EM estimation is especially well-suited to detectors which are themselves iterative.

5 Numerical Results

5.1 Simulation parameters

To validate the proposed algorithms, we have carried out Monte Carlo simulations. We consider a $R = 1/3$ rate turbo-code consisting of two constituent systematic recursive rate one-half convolutional codes with generators $(21, 37)_8$. We consider a block of $N = 512$ 16-QAM symbols made of N_a unknown data and N_t pilot symbols. The channel has length $L = 15$ and was generated with independent components, each being a zero-mean complex Gaussian random variable with an exponential power delay profile [5]:

$$E\left[|h(l)|^2\right] = 0.58 \exp(-2l/5), \, l = 0, \ldots, L - 1. \tag{37}$$

Hence, the energy of the channel is concentrated mainly in the first few channel taps. Note also that $C_{\mathbf{h}}$ is diagonal, with its diagonal elements given by (37). To avoid ISI, a cyclic prefix of length $\nu = 16$ is employed. We have set $\Delta_{max} = 40$ and kept actual delay fixed $\Delta = 23$. In the numerical results, we compare code-aided or *soft decision directed* (SDD) estimation with *data-aided* (DA) estimation. When considering code-aided or SDD estimation, the estimation is fully embedded in the iterative detector, i.e., we set $I_D^* = 1$ as explained in section 4.3.

---

[4]Sometimes known as *extrinsic* information.

5.2 Channel estimation

First, we consider the estimation of the channel impulse response $\mathbf{h}$, assuming perfect knowledge of F and Δ. In Fig. 3, we compare the mean squared error (MSE) of the iterative EM algorithm for both ML and MAP estimation as function of the number of iterations within the EM algorithm. The first iteration corresponds to a pure data-aided estimator. The signal to noise ratio is set equal to $E_b/N_0 = 12\,\mathrm{dB}$ and the different curves correspond to different number of pilot symbols applied to provide initial estimates. Clearly, the initial estimate is rather crucial to prevent the EM algorithm from converging to a local maximum of the cost function rather than to the global maximum (ML or MAP solution). We observe a significant improvement resulting from iterating between detector and estimator. The MAP estimator exhibits a lower MSE than the ML estimator, since the former takes prior information into account about the channel taps distribution, whereas the latter ignores this information. This also translates into the BER plots from Fig. 4, which show the significant performance advantage of the MAP estimator compared to the ML estimator.

5.3 Carrier frequency offset estimation

Next we consider the carrier frequency offset (CFO) estimation, assuming perfect knowledge of the other parameters. In the simulation results, the initial estimate of the CFO is set equal to zero, while the actual *normalized* CFO varies from zero to 0.5, i.e. the CFO takes on values up to one half of the carrier spacing. Fig. 5 and 6 show plots of the MSE and BER, respectively, as a function of the actual normalized CFO for $E_b/N_0 = 12\,\mathrm{dB}$. It is well-known that multi-carrier systems are very sensitive to residual CFO's [3,4]. This is confirmed in Fig. 6. The EM algorithm, on the other hand, can recover the CFO in a satisfactory manner as long as the CFO is smaller than 0.4 times the carrier spacing.

The improvement obtained from the iterative estimation may be even more pronounced in Fig. 7, where the MSE is plotted as a function of the SNR, when all unknown parameters are jointly estimated. In particular at high SNR values, the difference between no CFO correction and EM estimation is substantial.

5.4 Frame synchronization

Finally, we consider frame synchronization, jointly with channel and CFO estimation. The MSE of the normalized delay is shown in Fig. 8. Again the iterative estimator turns out to be superior to its data-aided counterpart. The MSE of the delay renders only a partial view of the behavior of $\hat{\Delta}$. Therefore, we take a look at the average ML cost function for the timing delay Δ

at $E_b/N_0 = 12\,\mathrm{dB}$, illustrated in Fig. 9. For both data-aided estimation and EM estimation, the cost function is maximized at $\Delta = 23$, which corresponds to the actual value of Δ. Furthermore, we observe that the peak is much more pronounced when approaching from $\Delta > 23$, compared to $\Delta < 23$. Hence, the probability to have a positive delay error, i.e. $\epsilon_\Delta \doteq \hat{\Delta} - \Delta > 0$, will be lower than the probability to encounter $\epsilon_\Delta < 0$. Fig. 10 depicts the probability mass function (pmf) of ϵ_Δ after estimation and confirms the latter conjecture. Note that even at $E_b/N_0 = 12\,\mathrm{dB}$, the frame synchronizer will make the wrong decision in about 50% and 2% of the time for the data-aided and EM estimator, respectively. Fortunately, the situation where $\epsilon_\Delta < 0$ is not very critical. In fact, if $L - \nu < \epsilon_\Delta < 0$, the frame synchronization error will not destroy the orthogonality among the subcarriers [5]. Furthermore, when $\nu \ll \epsilon_\Delta \leq L - \nu$, no large degradation is expected, since the first few channel taps carry most of the energy. On the other hand, when $\epsilon_\Delta > 0$, the estimate of $\mathbf{h}$ will not capture the dominant components, which can have a severe impact on the detection performance. From Fig. 10, we see that the latter situation occurs rarely with the data-aided and EM-based estimators.

To conclude the simulation results, we show the BER resulting from the joint ML estimation of all parameters, in Fig. 11. As expected, the DA estimator gives rise to large degradations, whereas the EM based estimator yields a close to optimal performance.

6   Conclusions

We have presented a novel Bayesian code-aided estimation algorithm for joint channel estimation and synchronization for multi-carrier systems. Based on the EM algorithm, the receiver iterates between data detection and estimation, with the exchange of soft information in the form of a posteriori probabilities. Compared to a conventional data-aided algorithm, the code-aided algorithm results in impressive gains in terms of mean squared error and BER performance. Although the complexity of this estimator is large, we have described how the computational load may be reduced in a practical set-up. For iterative detectors, the proposed algorithm can be embedded in the iterative detection, which only gives rise to a fairly small overhead. The proposed algorithm can easily be extended to any other (multi-user) multi-carrier system, as long as a soft detector is available.

Acknowledgement

References

[1] J.A.C. Bingman. "Multicarrier modulation: an idea whose time has come". *IEEE Comm. Magazine*, 28(5):5–14, May 1990.

[2] Nanda, S.; Walton, R.; Ketchum, J.; Wallace, M.; Howard, S.;. "A high-performance MIMO OFDM wireless LAN". *IEEE Comm. Magazine*, 43(1):53–60, January 2005.

[3] T. Pollet, M. Van Bladel and M. Moeneclaey. "BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise". *IEEE Trans. Comm.*, 43(234):191–193, feb-mar-apr 1995.

[4] H. Steendam and M. Moeneclaey. "Synchronization sensitivity of multicarrier systems". *European Transactions on Communications, ETT special issue on Multi-Carrier Spread Spectrum*, 52(5):834–844, May 2004.

[5] M. Morelli. "Timing and frequency synchronization for the uplink of an OFDMA system". *Transactions on Communications*, 52:pp.296–306, February 2004.

[6] T. M. Schmidl and D. C. Cox. "Robust frequency and timing synchronization for OFDM". *IEEE Transactions on Communicaions*, 45:pp.1613–1621, December 1997.

[7] M. Morelli and U. Mengali. "A comparison of pilot-aided channel estimation methods for OFDM systems". *IEEE Trans. on Signal Processing*, 49(12):3065–3073, December 2001.

[8] Speth, M.; Fechtel, S.A.; Fock, G.; Meyr, H.;. "Optimum receiver design for wireless broad-band systems using OFDM. I". *IEEE Trans. Comm*, 47(11):1668–1677, November 1999.

[9] Speth, M.; Fechtel, S.; Fock, G.; Meyr, H. "Optimum receiver design for OFDM-based broadband transmission .II. A case study". *IEEE Trans. Comm*, 49(4):571–579, April 2001.

[10] E.G. Larsson, Guoqing Liu, Jian Li and G.B. Giannakis. "Joint symbol timing and channel estimation for OFDM based WLANs". *IEEE Comm. Letters*, 5(8):325–327, August 2001.

[11] Yingwei Yao and G.B. Giannakis. "Blind carrier frequency offset estimation in SISO, MIMO, and multiuser OFDM Systems". *IEEE Trans. Comm*, 53(1), January 2005.

[12] M. Tanda. "Blind symbol-timing and frequency offset estimation in OFDM systems with real data symbols". *IEEE Trans. Comm*, 52(10):1609–1612, October 2004.

[13] M.C. Necker and G.L. Stuber. "Totally blind channel estimation for OFDM on fast varying mobile radio channels". *IEEE Trans. on Wireless Comm.*, 3(5):1514–1525, September 2004.

[14] C. Berrou, A. Glavieux and P. Thitimajshima. "Near Shannon limit error-correcting coding and decoding: turbo codes". In *IEEE ICC'94*, pages 1064–1070, 1993.

[15] A. P. Dempster, N. M. Laird and D. B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm". *Journal of the Royal Statistical Society*, 39(1):1–38, 1977. Series B.

[16] N. Noels, V. Lottici, A. Dejonghe, H. Steendam, M. Moeneclaey, M. Luise , L. Vandendorpe. "A Theoretical Framework for Soft Information Based Synchronization in Iterative (Turbo) Receivers". *"EURASIP Journal on Wireless Communications and Networking JWCN, Special issue on Advanced Signal Processing Algorithms for Wireless Communications"*, 2005. Accepted for publication.

[17] G.A Al-Rawi, T.Y. Al-Naffouri, A. Bahai and J. Cioffi. "Exploiting error-control coding and cyclic prefix in channel estimation for coded OFDM systems". *IEEE Comm. Letters*, 7(7):388–390, July 2003.

[18] Aldana, C.H.; de Carvalho, E.; Cioffi, J.M. "Channel estimation for multicarrier multiple input single output systems using the EM algorithm". *IEEE Trans. on Signal Processing*, 51(12):3280–3292, December 2003.

[19] Al-Naffouri, T.Y.; Awoniyi, O.; Oteri, O.; Paulraj, A.;. "Receiver design for MIMO-OFDM transmission over time variant channels". *IEEE Globecom*, 4(29):436–440, December 2004.

[20] Seung Young Park; Yeun Gu Kim; Chung Gu Kang;. "Iterative receiver for joint detection and channel estimation in OFDM systems under mobile radio channels". *IEEE Trans. on Vehicular Technology*, 53(2):450–460, March 2004.

[21] S. ten Brink, J. Speidel and R.-H.Yan. "Iterative demapping and decoding for multilevel modulation". In *IEEE Global Communications Conference (Globecom)*, pages 579–584, Sydney, Australia, November 1998.

[22] F. Kschischang, B. Frey and H.-A. Loeliger. "Factor graphs and the sum-product algorithm". *IEEE Trans. Inform. Theory*, 47(2):pp.498–519, February 2001.

[23] J.A. Fessler and A.O. Hero. "Space alternating generalized expectation maximization algorithm". *IEEE Transactions on Communications*, 42(10):pp.2664–2677, 1994.

[24] Gallo, A.S.; Vitetta, G.M.; Chiavaccini, E. "A BEM-based algorithm for soft-in soft-output detection of co-channel signals". *IEEE Transactions on Wireless Communications*, 3(5):1533–1542, September 2004.

[25] H. Wymeersch. *"Sofware Radio Algorithms for Coded Transmission"*. PhD thesis, Ghent University, Gent, Belgium, 2005.

[26] J.M. Bernardo and A.F.M. Smith. *"Bayesian Theory"*. Wiley, 1994.

[27] H. Wymeersch, F. Simoens and M. Moeneclaey. "Code-aided channel tracking for OFDM". In *IEEE International Symposium on Turbo Codes & Related Topics*, Munich, April 2006.

[28] V. Lottici and M. Luise. "Embedding carrier phase recovery into iterative decoding of turbo-coded linear modulations". *IEEE Transactions on Communications*, 52(4):661–669, April 2004.
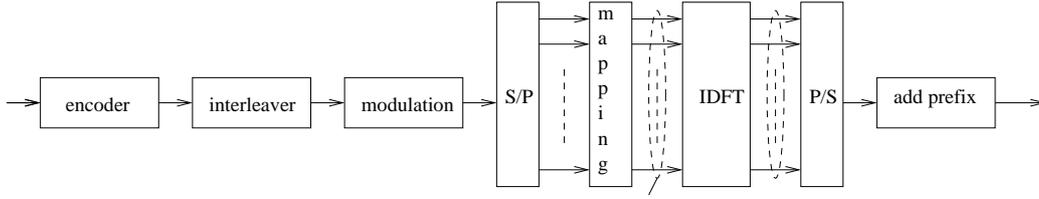
Figure 1: Multi-carrier transmitter.
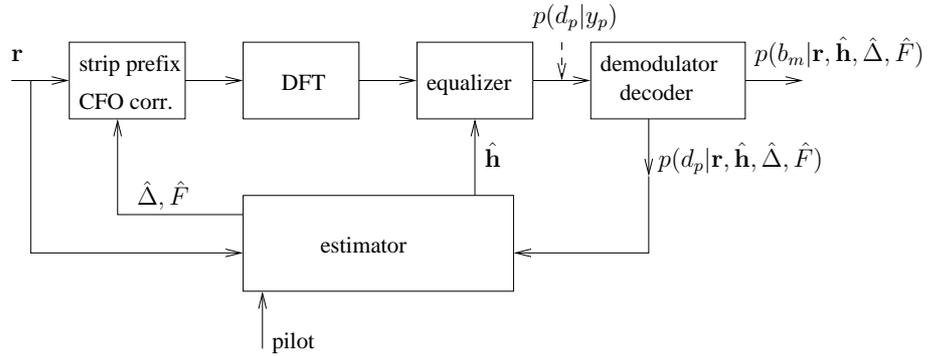
Figure 2: Multi-carrier detector and estimator.
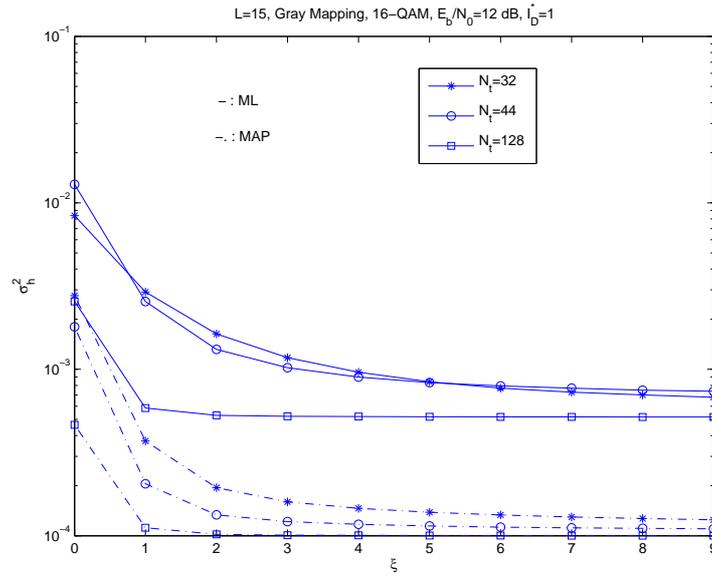
Figure 3: Channel Impulse Response estimation: MSE performance as function of the iteration index ξ ($\xi = 0$ corresponds to pure data-aided estimation) at $E_b/N_0 = 12$ dB. Perfect frame/frequency synchronization is assumed.
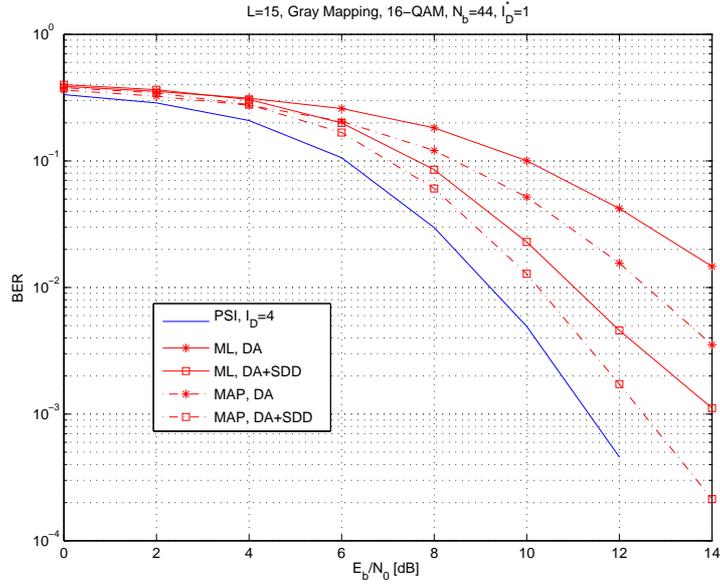
Figure 4: Channel Impulse Response estimation: BER performance as function of E_b/N_0. Perfect frame/frequency synchronization is assumed. (PSI stands for Perfect State Information, DA/SDD stands for Data-Aided/Soft Decision Directed)
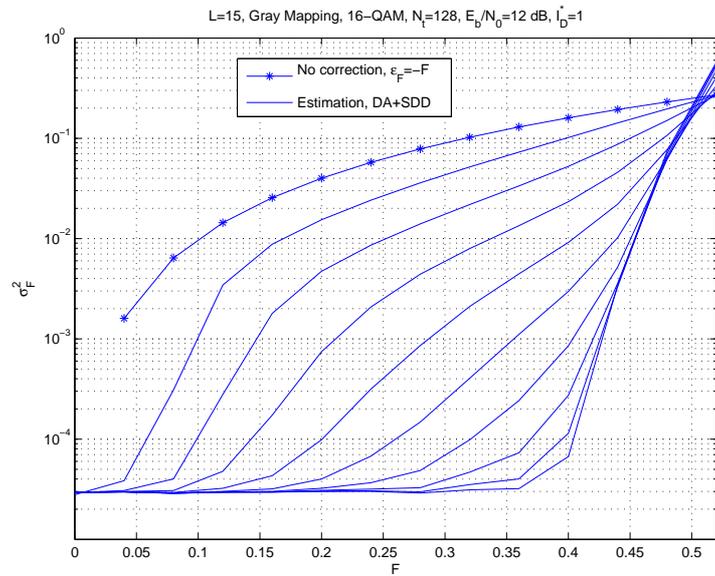
Figure 5: Frequency synchronization: variance of frequency estimation error at $E_b/N_0 = 12$ dB using $\hat{F}(\xi = 0) = 0$ as a frequency initial estimate. Perfect frame synchronization and channel knowledge are assumed.
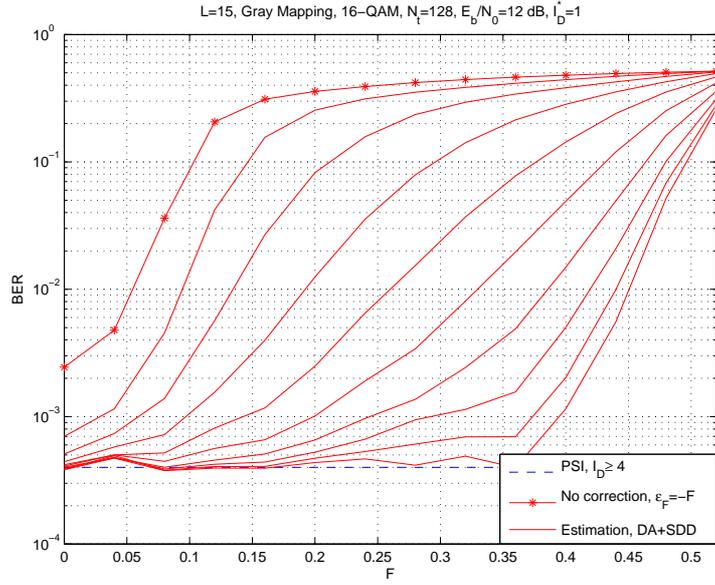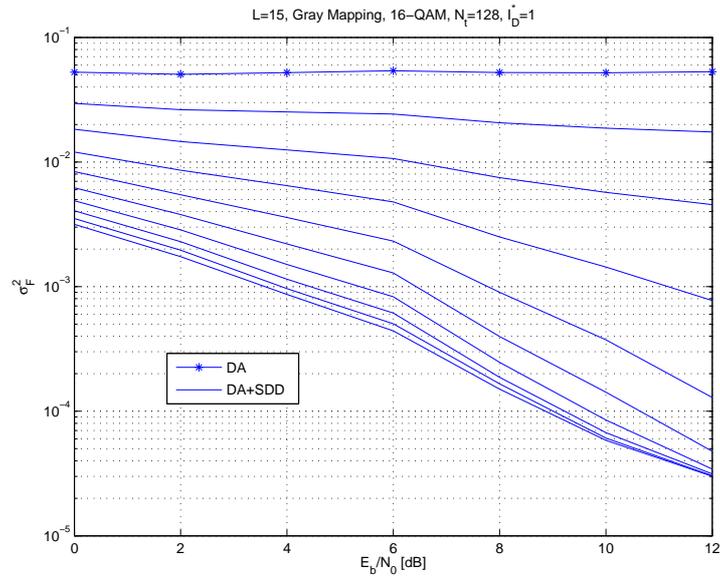
Figure 6: Frequency synchronization: BER at $E_b/N_0 = 12$ dB using $\hat{F}(\xi = 0) = 0$ as a frequency initial estimate. Perfect frame synchronization and channel knowledge are assumed.

Figure 7: Joint frame-frequency synchronization and channel estimation: MSE using $\hat{F}(\xi = 0) = 0$ as a frequency initial estimate and F is selected randomly such that $|F| < 0.4$.
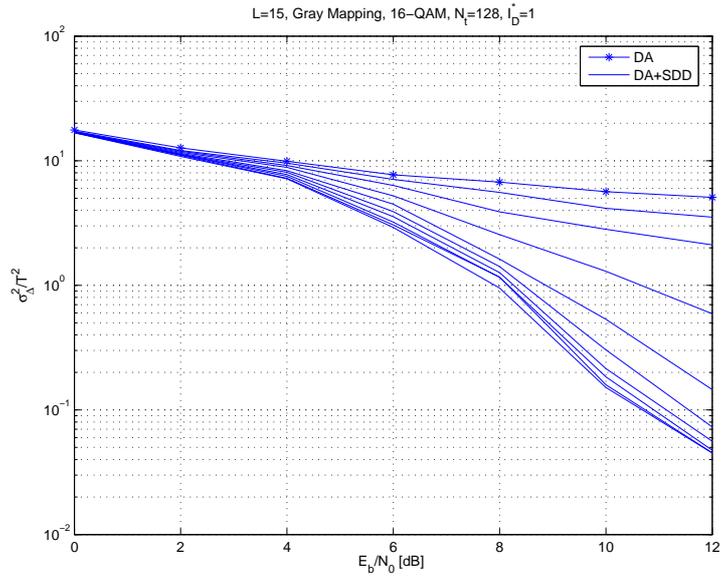
Figure 8: Joint frame-frequency synchronization and channel estimation: variance of timing estimation error at $E_b/N_0 = 12$ dB.
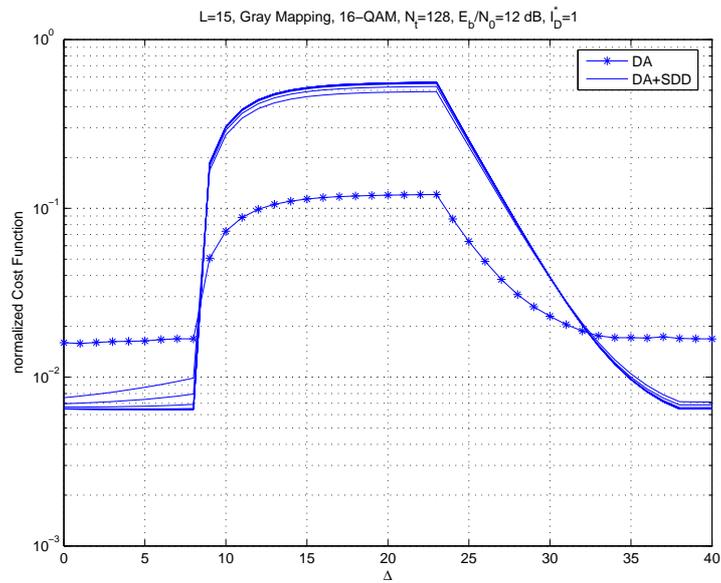

Figure 9: Joint frame-frequency synchronization and channel estimation: trial value of Δ vs. ML timing cost function at $E_b/N_0 = 12$ dB. The true frame shift is $\Delta = 23$.
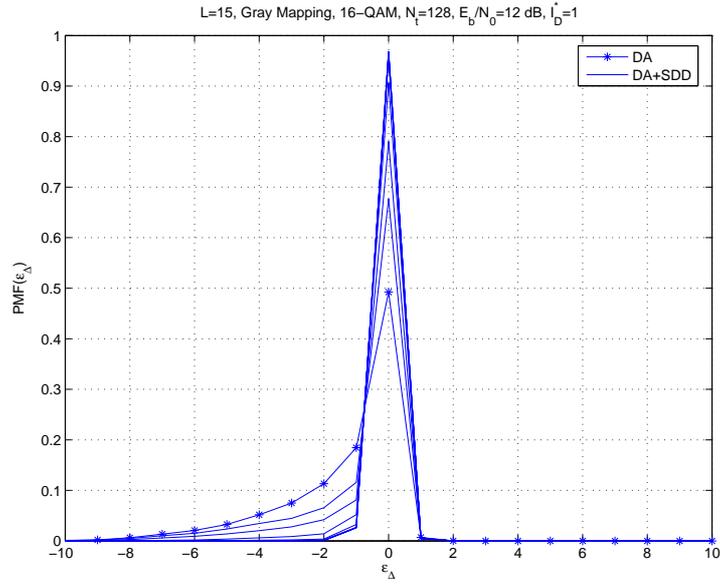
Figure 10: Joint frame-frequency synchronization and channel estimation: pmf of timing estimation error at $E_b/N_0 = 12$ dB.
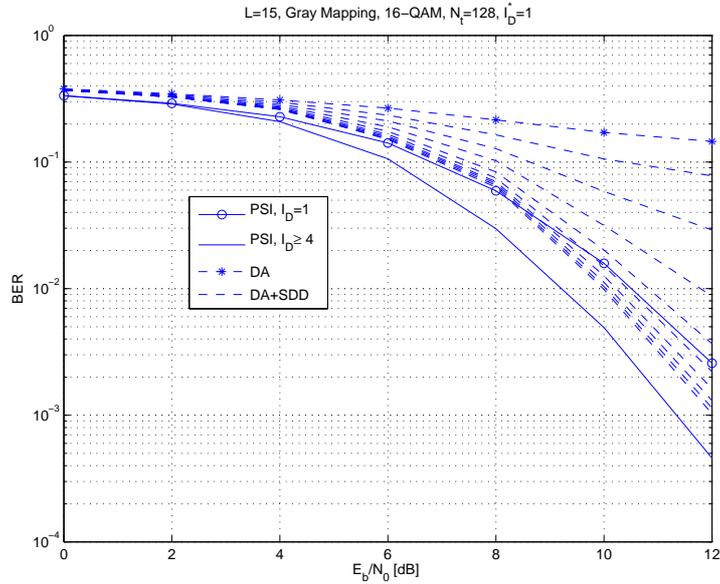

Figure 11: Joint frame-frequency synchronization and channel estimation: BER performance using $\hat{F}(\xi = 0) = 0$ as a frequency initial estimate and F is selected randomly such that $|F| < 0.4$.