Turbo synchronization : an EM algorithm interpretation

N. Noels*, C. Herzet†, A. Dejonghe†, V. Lottici‡, H. Steendam*,
M. Moeneclaey*, M. Luise‡ and L. Vandendorpe†

*DIGICOM Research Group, TELIN Department, Ghent University, St. Pietersnieuwstraat 41, B9000 Gent, Belgium
†Communications Laboratory, Université catholique de Louvain, Pl. du Levant 2, B1348 Louvain-la-Neuve, Belgium
‡Department of Information Engineering, University of Pisa, Via Diotisalvi 2, I56122 Pisa, Italy

***Abstract*—This paper is devoted to turbo synchronization, that is to say the use of soft information to estimate parameters like carrier phase, frequency offset or timing within a turbo receiver. It is shown how maximum-likelihood estimation of those synchronization parameters can be implemented by means of the iterative expectation-maximization (EM) algorithm [1]. Then we show that the EM algorithm iterations can be combined with those of a turbo receiver. This leads to a general theoretical framework for turbo synchronization. The soft decision-directed ad-hoc algorithm proposed in [2] for carrier phase recovery turns out to be a particular instance of this implementation. The proposed mathematical framework is illustrated by simulations reported for the particular case of carrier phase estimation combined with iterative demodulation and decoding [3].**

I. Introduction

THANKS to the turbo codes, the performance of coded transmission systems has never been so close to the Shannon limit. On top of that, the associated idea of iterative decoding has more recently been extended to other receiver functions. This led to the so-called turbo principle which enables to perform suboptimal joint detection and decoding in a number of transmission configurations, through the iterative exchange of soft information between soft-input/soft-output (SISO) stages. See [4] for a review of some existing turbo receivers.

In addition to detection and decoding, a receiver has also to perform synchronization i.e. to estimate a number of parameters like carrier phase, frequency offset, timing offset... Synchronization for turbo receivers is an even more challenging task since such systems usually operate at low SNR values. In this context, it is relevant to try using the available soft information provided by the turbo detector to help the synchronizers. This approach will be referred to as turbo synchronization in the sequel. The idea has already been applied in a number of contributions. References [2] and [5], for instance, propose to combine soft decision-directed carrier phase estimation with turbo decoding. In [6] a carrier phase recovery method embedded in the MAP decoder and exploiting the extrinsic information generated at each iteration is reported. However, these schemes do often not rely on any theoretical basis.

This work has been supported by the Interuniversity Attraction Poles Programme - Belgium State - Federal Office for Scientific, Technical and Cultural Affairs. A. Dejonghe thanks the Belgian "F.N.R.S." for its financial support.

The purpose of this paper is to give a mathematical interpretation to turbo synchronization by means of the expectation-maximization (EM) algorithm [1]. Such an algorithm has been applied to various problems. In [7] for instance, the missing data is modeled by a Markov chain and the EM algorithm is used for channel and noise variance estimation in combination with optimal BCJR-based detection. This latter approach has been extended to turbo receivers in [8] (see also references therein) and [9] for channel and noise variance estimation in CDMA and MIMO context respectively.

In the present paper we will focus on the specific problem of synchronization. Section II will give a general formulation of iterative ML estimation of unknown parameters in the presence of nuisance parameters by means of the EM algorithm. The particular issue of synchronization for digital data-modulated passband signal will then be addressed in section III. This implementation will be extended to the turbo context in section IV by showing that the EM algorithm iterations (for parameters estimation) can be combined with those of a turbo receiver (for symbol detection/decoding). This will lead to a general theoretical framework for turbo synchronization. In particular, it will turn out that algorithms introduced in an ad-hoc fashion, such as the blind soft-decision directed carrier phase synchronizer recently proposed in [2], actually corresponds to a particular instance of the general scheme proposed here. In order to illustrate the mathematical considerations, simulation results will be provided in section V for the particular case of carrier phase recovery in a bit-interleaved coded modulation (BICM) transmission system with iterative demodulation and decoding at the receiver [3].

II. ML Estimation in the Presence of a Nuisance Vector

Let $\mathbf{r}$ denote a random vector obtained by expanding the received modulated-signal $r(t)$ onto a suitable basis and let $\mathbf{b}$ indicate a deterministic vector of parameters to be estimated from the observation of the received vector $\mathbf{r}$. Assume that $\mathbf{r}$ also depends on a random discrete-valued nuisance parameter vector $\mathbf{a}$ independent of $\mathbf{b}$ and with a priori probability density function $p(\mathbf{a})$. The problem addressed in this section is to find the ML estimate $\hat{\mathbf{b}}$ of $\mathbf{b}$ that is to say the solution of

$$\hat{\mathbf{b}} = \arg\max_{\tilde{\mathbf{b}}}\{\ln p(\mathbf{r}|\tilde{\mathbf{b}})\}, \qquad (1)$$

where

$$p(\mathbf{r}|\tilde{\mathbf{b}}) = \int_{\mathbf{a}} p(\mathbf{r}|\mathbf{a},\tilde{\mathbf{b}})\, p(\mathbf{a})\, d\mathbf{a} \quad (2)$$

and $\tilde{\mathbf{b}}$ is a trial value of $\mathbf{b}$. Most of the time there is unfortunately no analytical solution to such a problem. In this case one has to use iterative numerical methods in order to find the solution of (1). One of these iterative methods is referred to as the expectation-maximization (EM) algorithm.

Using the formalism of this algorithm, let us set $\mathbf{r}$ as the *incomplete* data set and $\mathbf{z} \triangleq [\mathbf{r},\mathbf{a}]$ as the *complete* data set. The EM algorithm states that the sequence $\hat{\mathbf{b}}^{(n)}$ defined by

$$\begin{aligned} \mathcal{Q}(\tilde{\mathbf{b}},\hat{\mathbf{b}}^{(n-1)}) &= \int_{\mathbf{z}} p(\mathbf{z}|\mathbf{r},\hat{\mathbf{b}}^{(\mathbf{n-1})})\ \ln p(\mathbf{z}|\tilde{\mathbf{b}})\, d\mathbf{z} && (3)\\ \hat{\mathbf{b}}^{(n)} &= \arg\max_{\tilde{\mathbf{b}}}\{\mathcal{Q}(\tilde{\mathbf{b}},\hat{\mathbf{b}}^{(n-1)})\} && (4)\end{aligned}$$

converges under fairly general conditions [10] towards the ML estimate (1). Using now the Bayes rule and taking into account the independence of $\mathbf{a}$ and $\mathbf{b}$ we may write

$$p(\mathbf{z}|\tilde{\mathbf{b}}) = p(\mathbf{r},\mathbf{a}|\tilde{\mathbf{b}}) = p(\mathbf{r}|\mathbf{a},\tilde{\mathbf{b}})\,p(\mathbf{a}|\tilde{\mathbf{b}}) = p(\mathbf{r}|\mathbf{a},\tilde{\mathbf{b}})\,p(\mathbf{a}).$$

Substitution of this result into (3) yields

$$\begin{aligned} \mathcal{Q}(\tilde{\mathbf{b}},\hat{\mathbf{b}}^{(n-1)}) &= \int_{\mathbf{a}} p(\mathbf{a}|\mathbf{r},\hat{\mathbf{b}}^{(n-1)})\ \ln p(\mathbf{r}|\mathbf{a},\tilde{\mathbf{b}})\, d\mathbf{a}\\ &+ \int_{\mathbf{a}} p(\mathbf{a}|\mathbf{r},\hat{\mathbf{b}}^{(n-1)})\ \ln p(\mathbf{a})\, d\mathbf{a}. && (5)\end{aligned}$$

The second term in (5) does not depend on $\tilde{\mathbf{b}}$ and does not affect the maximization operation in (4). It can therefore be dropped. So, for the particular case of parameter $\mathbf{b}$ estimation in the presence of independent nuisance vector $\mathbf{a}$, the $\mathcal{Q}$-function has the following structure

$$\mathcal{Q}(\tilde{\mathbf{b}},\hat{\mathbf{b}}^{(n-1)}) = \int_{\mathbf{a}} p(\mathbf{a}|\mathbf{r},\hat{\mathbf{b}}^{(\mathbf{n-1})})\ \ln p(\mathbf{r}|\mathbf{a},\tilde{\mathbf{b}})\, d\mathbf{a}. \quad (6)$$

The solution of (1) can then be found iteratively by only using posterior probabilities $p(\mathbf{a}|\mathbf{r},\hat{\mathbf{b}}^{(n)})$ and the log-likelihood function $\ln p(\mathbf{r}|\mathbf{a},\tilde{\mathbf{b}})$.

III. EM Algorithm and Synchronization

In this section we will apply the general framework of the previous section to the particular case of synchronization for a digital data-modulated bandpass signal. In this context, the nuisance parameter vector $\mathbf{a}$ contains the values of the K unknown transmitted data symbols $(a_0, a_1, ..., a_{K-1}) \in \mathcal{A}^K$ where $\mathcal{A}$ is the constellation alphabet. The parameter vector $\mathbf{b}$ to be estimated contains the channel gain A, the symbol timing τ, the carrier frequency offset ν and the phase offset θ. For the sake of simplicity we will consider an AWGN channel in the sequel. The baseband received signal $r(t)$ can then be written as

$$r(t) = A \sum_{k=0}^{K-1} a_k\, p(t-kT-\tau)\, e^{j(2\pi\nu t+\theta)} + w(t), \quad (7)$$

where T is the symbol period, $p(t)$ is a unit energy square-root raised-cosine pulse and $w(t)$ is complex-valued AWGN with power spectral density $2N_0$.

Neglecting terms independent of $\mathbf{b}$, the log-likelihood function $\ln p(\mathbf{r}|\mathbf{a},\tilde{\mathbf{b}})$ present in (6) can be written as

$$\begin{aligned} \ln p(\mathbf{r}|\mathbf{a},\tilde{\mathbf{b}}) &= -2\tilde{A}\,\mathrm{Re}\{\sum_{k=0}^{K-1} a_k^*\, y_k(\tilde{\nu},\tilde{\tau})\, e^{-j\tilde{\theta}}\}\\ &+ \tilde{A}^2 \sum_{k=0}^{K-1} |a_k|^2, && (8)\end{aligned}$$

where

$$y_k(\tilde{\nu},\tilde{\tau}) \triangleq \int_{-\infty}^{+\infty} r(t)\, e^{-j(2\pi\tilde{\nu}t)}\, p(t-kT-\tilde{\tau})\, dt.$$

Let us define for each transmitted symbol a_k

$$\begin{aligned} \eta_k(\mathbf{r},\hat{\mathbf{b}}^{(n-1)}) &\triangleq \int_{\mathbf{a}\in\mathcal{A}^K} a_k\ p(\mathbf{a}|\mathbf{r},\hat{\mathbf{b}}^{(n-1)})\, d\mathbf{a}\\ &= \sum_{a\in\mathcal{A}} a_k\ p(a_k = a|\mathbf{r},\hat{\mathbf{b}}^{(n-1)}), && (9)\\ \rho_k(\mathbf{r},\hat{\mathbf{b}}^{(n-1)}) &\triangleq \int_{\mathbf{a}\in\mathcal{A}^K} |a_k|^2\, p(\mathbf{a}|\mathbf{r},\hat{\mathbf{b}}^{(n-1)})\, d\mathbf{a}\\ &= \sum_{a\in\mathcal{A}} |a_k|^2 p(a_k = a|\mathbf{r},\hat{\mathbf{b}}^{(n-1)}). && (10)\end{aligned}$$

Using these definitions and replacing $\ln p(\mathbf{r}|\mathbf{a},\tilde{\mathbf{b}})$ by (8) in (6), we get

$$\begin{aligned} \mathcal{Q}(\tilde{\mathbf{b}},\hat{\mathbf{b}}^{(n-1)}) = &-2\tilde{A}\,\mathrm{Re}\{\sum_{k=0}^{K-1} \eta_k^*(\mathbf{r},\hat{\mathbf{b}}^{(n-1)})\, y_k(\tilde{\nu},\tilde{\tau})\, e^{-j\tilde{\theta}}\}\\ &+ \tilde{A}^2 \sum_{k=0}^{K-1} \rho_k(\mathbf{r},\hat{\mathbf{b}}^{(n-1)}). && (11)\end{aligned}$$

Due to the particular structure of $\mathcal{Q}(\tilde{\mathbf{b}},\hat{\mathbf{b}}^{(n-1)})$ the synchronization parameter joint maximization problem of (4) leads to a partially decoupled solution

$$[\hat{\nu}^{(n)},\hat{\tau}^{(n)}] = \arg\max_{\tilde{\nu},\tilde{\tau}}\{|\sum_{k=0}^{K-1} \eta_k^*(\mathbf{r},\hat{\mathbf{b}}^{(n-1)})\, y_k(\tilde{\nu},\tilde{\tau})|\} \quad (12)$$

$$\hat{\theta}^{(n)} = \arg\{\sum_{k=0}^{K-1} \eta_k^*(\mathbf{r},\hat{\mathbf{b}}^{(n-1)})\, y_k(\hat{\nu}^{(n)},\hat{\tau}^{(n)})\} \quad (13)$$

$$\hat{A}^{(n)} = \frac{|\sum_{k=0}^{K-1} \eta_k^*(\mathbf{r},\hat{\mathbf{b}}^{(n-1)})\, y_k(\hat{\nu}^{(n)},\hat{\tau}^{(n)})|}{\sum_{k=0}^{K-1} \rho_k(\mathbf{r},\hat{\mathbf{b}}^{(n-1)})}. \quad (14)$$

Note that thanks to relations (9) and (10) the a posteriori average values $\eta_k(\mathbf{r},\hat{\mathbf{b}}^{(n-1)})$ and $\rho_k(\mathbf{r},\hat{\mathbf{b}}^{(n-1)})$ can be computed from the marginals $p(a_k|\mathbf{r},\hat{\mathbf{b}}^{(n-1)})$ only. In other words,

Algorithm 1:

```
1  b̂^(0) = b_0;
   for n = 1 → N do
2      Reset turbo system;
3      repeat
           Perform 1 turbo iteration;
       until soft information convergence;
4      Computation b̂^(n);
   end
```

for the particular case of synchronization for a digital data-modulated bandpass signal, the implementation of an iterative ML estimation algorithm only requires the evaluation of the marginal posterior probabilities $p(a_k|\mathbf{r}, \hat{\mathbf{b}}^{(n-1)})$. As we will see in the next sections, this important feature makes EM-like algorithms well-suited to synchronization in turbo receivers.

IV. EM Algorithm and Turbo-System

Since the advent of turbo-codes, the turbo principle has received a lot of attention. Many simulation results tend to show that turbo receivers can reach iteratively ML performance on information bit estimation. Consequently, although never proved, one usually assumes that a turbo receiver can provide, after convergence of the iterative process, soft information which equals channel symbol posterior probabilities. We will adopt the same assumption in the sequel.

As shown in section III, the estimation of synchronization parameter $\mathbf{b}$ via the EM algorithm only requires the knowledge of marginal posterior probabilities $p(a_k|\mathbf{r}, \hat{\mathbf{b}}^{(n-1)})$. This makes synchronization via the EM algorithm and turbo receivers complementary since the marginal probabilities needed by the first one can be provided by the second one. This leads to the turbo synchronization algorithm described in Algorithm 1. N denotes the number of performed EM iterations. Step 2 means that extrinsic information in the turbo receiver is reinitialized to zero at each EM iteration. Step 3 means that the turbo system will iterate until soft information reaches a steady-state value. These last operations enable to assume that the soft information provided by the turbo receiver will be a good approximation of marginal posterior probabilities $p(a_k|\mathbf{r}, \hat{\mathbf{b}}^{(n)})$. Step 4 is simply performed by resolving (12), (13) and (14). Note that this agorithm will dramatically increase the computational complexity as compared to a perfectly synchronized receiver since the turbo system is required to converge at each EM iteration.

In order to deal with such a problem we propose the approximate implementation described in Algorithm 2. The turbo system is no longer reinitialized and only one turbo iteration is now performed at each EM iteration. In other words, the synchronization iterations (EM algorithm) are merged with the detection/decoding ones (turbo receiver). The only extra computional load with respect to a perfectly synchronized turbo detector is then the evaluation of equations (12), (13) and (14).

Algorithm 2:

```
1  b̂^(0) = b_0;
   for n = 1 → N do
2      Perform 1 turbo iteration;
3      Computation b̂^(n);
   end
```

This leads to a very slight complexity increase in comparison with the computational requirements of the detection/decoding algorithms implemented in the turbo receiver. Note however that, strictly speaking, Algorithm 2 is no longer an EM algorithm. Indeed, performing only one turbo iteration at each EM iteration may lead to poor approximations of marginal posterior probabilities $p(a_k|\mathbf{r}, \hat{\mathbf{b}}^{(n)})$, especially for the first EM iterations. Moreover, non-resetting extrinsinc information implies that the soft information computed by the turbo system will be a function of all the previous estimates $\hat{\mathbf{b}}^{(n-1)}, \hat{\mathbf{b}}^{(n-2)}...\hat{\mathbf{b}}^{(0)}$ and not only of $\hat{\mathbf{b}}^{(n-1)}$ as required by the EM algorithm. Consequences of both these approximations will be studied through simulation results in the next section.

Algorithm 2 actually corresponds to the one introduced in an ad hoc fashion by Lottici [2] for the particular case of phase estimation in a turbo-coded scheme. We have provided here a mathematical interpretation of Lottici's algorithm by showing that it may actually be regarded as an EM algorithm approximation. Note that the very general framework considered in section III shows that this algorithm may be properly extended to the estimation of a wide range of parameters (frequency offset, timing offset...) in various transmission schemes (MIMO, CDMA...).

V. Simulation Results

This section is devoted to the study of the two synchronization algorithms described in the previous section. More particularly we will focus on algorithms performance for carrier phase synchronization.

The considered transmitter is a BICM scheme: it is made up with a binary encoder and a constellation mapper separated by a bit interleaver. At the receiver a turbo-demodulation scheme is considered. Such a system introduced in [3] performs iterative joint demodulation and decoding through the exchange of extrinsic information between a SISO demodulator and a SISO decoder. We consider a rate-$\frac{1}{2}$ non-systematic convolutional encoder with generators $(g_1, g_2) = (5, 7)_8$ and use 8-PSK modulation with set partitioning mapping. The carrier frequency offset ν, the timing offset τ and the channel gain A are assumed to be known. The phase offset θ is unknown and needs to be estimated. Simulations have been run with frames of 2000 8-PSK symbols and 10 EM iterations have been performed. The interleaver is totally random and a different permutation is used at each frame. In all simulations we totalized 200 frame errors at the last iteration. For Algorithm 1, 5 turbo-demodulation iterations have been performed at each EM step.

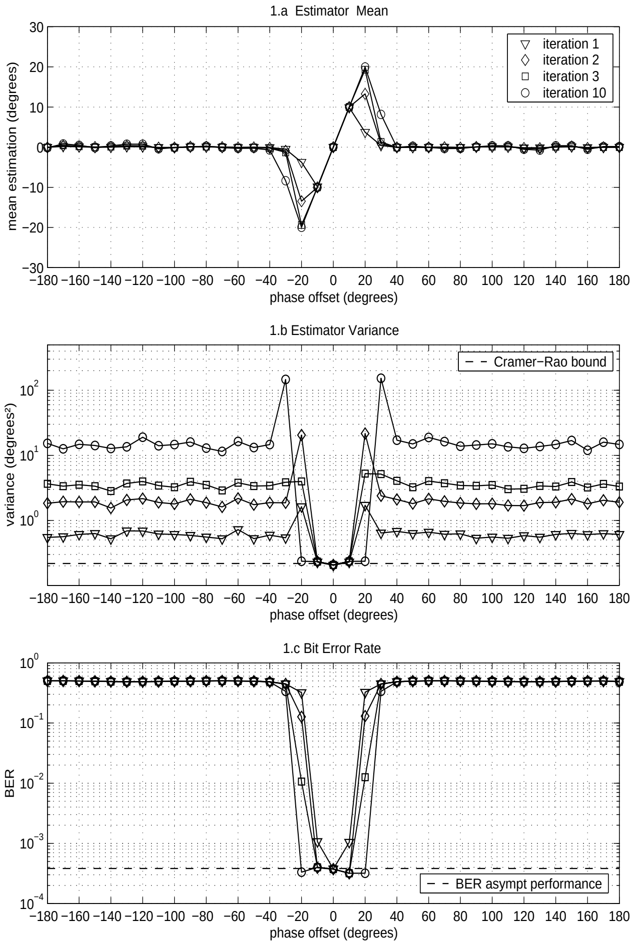

Fig. 1. Algorithm 1 - Estimator mean, estimator variance and BER performance versus carrier phase offset for $E_b/N_0 = 4dB$

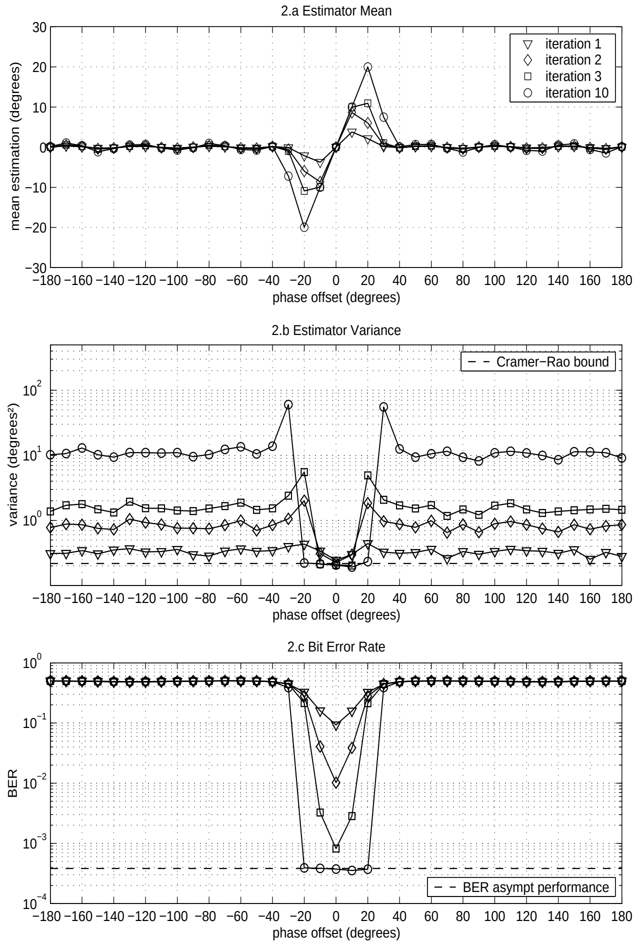

Fig. 2. Algorithm 2 - Estimator mean, estimator variance and BER performance versus carrier phase offset for $E_b/N_0 = 4dB$

Fig. 1 and Fig. 2 represent the algorithms performance for $E_b/N_0 = 4$dB and for a phase offset θ ranging from -180 to 180 degrees. More particularly, Fig. 1.a and 2.a show the mean estimated phase value with Algorithm 1 and 2 respectively. The estimation variance is illustrated in Fig. 1.b for Algorithm 1 and in Fig. 2.b for Algorithm 2. In these figures, the Cramer-Rao bound curve (corresponding to known data) is plotted in dashed line. Fig. 1.c and Fig. 2.c represent the bit error rate (BER) reached by the system with both algorithms. The BER asymptotic performance curve of the considered iterative demodulation and decoding scheme, i.e. the one obtained considering perfectly synchronized turbo-demodulator with perfect soft information fed back by the decoder, is drawn in dashed line. Note that the results are provided at an E_b/N_0-ratio for which the perfectly synchronized turbo-demodulator reaches asymptotic performance.

Algorithm 1 and 2 exhibit the same performance at EM iteration 10. The acquisition region i.e. the phase set for which the receiver is able to recover correctly the carrier phase ranges from -20 to 20 degrees for both methods. In this region the estimator variance meets the Cramer-Rao bound and the BER reaches the perfectly-synchronized system ML performance on bit detection. Note however that Algorithm 1 enables faster convergence in term of EM iterations than Algorithm 2. This observation may be explained by the crude approximation made on the posterior probabilities in Algorithm 2. Indeed, soft information provided by the SISO decoder in the first iterations is a rough approximation of the true marginal posterior probabilities, leading to a rather high BER as shown in Fig. 2.c. The phase estimates computed with the Algorithm 2 will then rely in the first iterations on bit estimates less reliable than with Algorithm 1. Note however that, even if more EM steps are needed, Algorithm 2 will reach the same performance as Algorithm 1 at considerably reduced complexity and delay.

Note also that the soft information provided by the turbo system depends on the code used for the transmission. So, unlike Non Data-Aided (NDA) estimators, the considered algorithms benefit from the code trellis structure which constrains the number of possible transmitted channel symbol sequences. This is illustrated by the non-periodicity of curves in Fig. 1 and

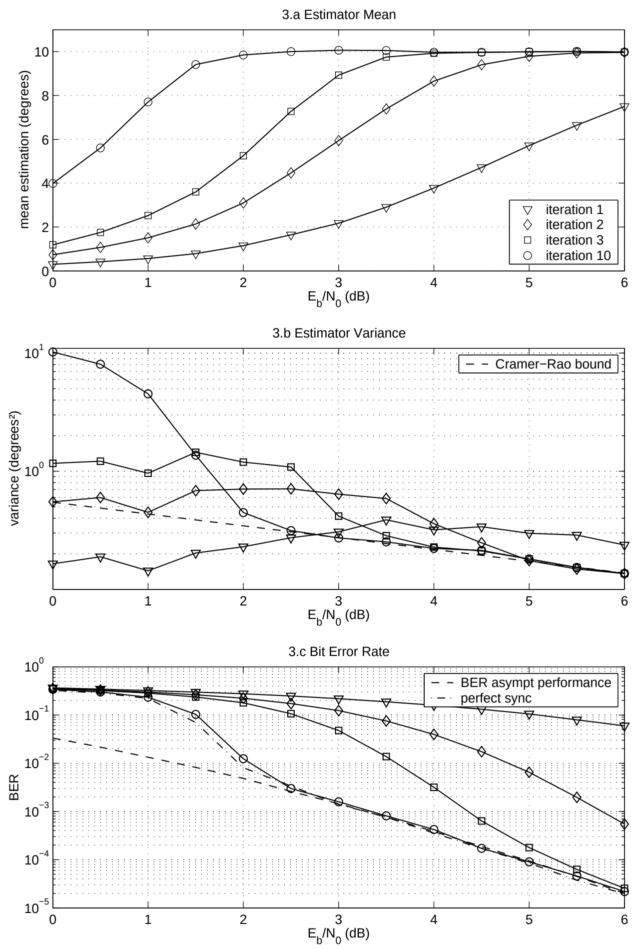

Fig. 3. Algorithm 2 - Estimator mean, estimator variance and BER for $\theta = 10$ degrees versus E_b/N_0-ratio

Fig. 2: Algorithm 1 and Algorithm 2 manage to remove the phase ambiguity (45^o for 8-PSK) inherent to a classical NDA estimator. Even if algorithms take the code structure into account, their acquisition region in the considered system is not greater than the maximum acquisition range achievable by a NDA estimator (45 degrees for 8-PSK). As shown by simulations results (not reported here), such a feature seems actually to be (essentially) a function of the used mapping. Gray mapping e.g. enables to reach an acquisition range of about 80 degrees (at the cost of a higher BER after convergence).

Fig. 3.a, 3.b and 3.c represent the mean, the variance and the BER obtained with Algorithm 2 for a phase offset $\theta = 10$ degrees and for E_b/N_0-ratios ranging from 0dB to 6dB. The setup is the same as for the previous simulations. The curve referred to as "perfect sync" in Fig. 3.c represent the performance reached by a perfectly synchronized system at iteration 10. Note that both Algorithm 2 and perfectly synchronized system reach asymptotic performance at iteration 10 for $E_b/N_0 \geq 2.5$dB.

We see from Fig. 3.b that the variance curves have a convergence threshold, that is to say an E_b/N_0-value (around 1.5dB here) below which variance no longer converges towards the Cramer-Rao bound whatever the number of iterations is. For E_b/N_0-ratios greater than this limit value, the estimator variance always reaches the Cramer-Rao bound provided that the number of iterations is sufficient. The convergence requires all the less iterations as the E_b/N_0-ratio increases. A similar behavior may be observed for the estimator mean in Fig. 3.a. This feature can be actually related to the BER curves in Fig. 3.c. Indeed, (asymptotic) unbiasedness and efficiency are performance of Data-Aided (DA) ML estimators. Now if the turbo detector is able to decrease the bit error rate, the behavior of our estimator will be closer and closer of a DA one. This is the reason why the convergence threshold of the mean and variance curves is located in the "water-fall" region in Fig. 3.c.

VI. Conclusion

This paper provides a general framework for turbo-synchronization (i.e. soft decision-directed synchronization within a turbo receiver). A general formulation of iterative ML estimation of synchronization parameters has first been derived by means of the EM algorithm. This implementation has then been extended to the turbo context by showing that turbo synchronization is actually an approximation of an EM algorithm. Simulation results (in terms of estimator mean, estimator variance and BER) have eventually been provided for the particular case of carrier phase recovery in a BICM transmission system with iterative demodulation and decoding at the receiver. These simulations show that the system reaches global (i.e. detection/synchronization) ML performance after convergence.

References

[1] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.*, Ser. B, vol. 39, no. 1, pp. 1-38, Jan. 1977.

[2] V. Lottici and M. Luise, "Carrier phase recovery for turbo-coded linear modulations," in *Proc. IEEE ICC'2002*, New-York, USA, Apr.-May 2002.

[3] S. ten Brink, J. Speidel and R.-H. Yan, "Iterative demapping and decoding for multilevel modulation," in *Proc IEEE Globecom'98*, pp. 579–584, Sydney, Australia, Nov. 1998.

[4] J. Hagenauer, "The turbo principle: tutorial introduction and state of the art," in *Proc. Int. Symposium on Turbo Codes & Related Topics*, pp. 1–11, Brest, France, Sep. 1997.

[5] A. Burr and L. Zhang, "A novel carrier phase recovery method for turbo-coded QPSK system," in *Proc. EW'2002*, Florence, Italy, Feb. 2002.

[6] W. Oh and K. Cheun, "Joint decoding and carrier phase recovery algorithm for turbo codes," *IEEE Comm. Letters*, vol. 5, no. 9, pp. 375-377, Sep. 2001.

[7] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Trans. Comm.*, vol. 42, no. 7, pp. 2406–2413, Jul. 1994.

[8] M. K. Kobayashi, J. Boutros and G. Caire, "Successive interference cancellation with SISO decoding and EM channel estimation," *IEEE J. Select. Areas Commun.*, vol. 19, no. 8, pp. 1450-1460, Aug. 2001.

[9] C. Lamy, F. Boixadera and J. Boutros, "Iterative APP decoding and channel estimation for multiple-input multiple-output channels," *submitted to IEEE Trans Comm.*, Jan. 2000.

[10] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Statistics*, vol. 11, no. 1, pp. 95–103, 1983.