



**HAL**  
open science

# Information theoretic approach to synchronization: the OFDM carrier frequency offset example

Romain Couillet, Merouane Debbah

► **To cite this version:**

Romain Couillet, Merouane Debbah. Information theoretic approach to synchronization: the OFDM carrier frequency offset example. AICT, 2010, Barcelona, Spain. pp.AICT - Couillet - 2010. hal-00553675

**HAL Id: hal-00553675**

**<https://hal.science/hal-00553675>**

Submitted on 7 Jan 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Information theoretic approach to synchronization: the OFDM carrier frequency offset example

Romain Couillet  
ST-NXP Wireless, Supélec  
505 Route des Lucioles  
06560 Sophia Antipolis, France  
Email: romain.couillet@supelec.fr

Mérouane Debbah  
Alcatel-Lucent Chair, Supélec  
Plateau de Moulon, 3 rue Joliot-Curie  
91192 Gif sur Yvette, France  
Email: merouane.debbah@supelec.fr

**Abstract**—In practical mobile communication systems, data-aided synchronization is performed before actual data exchanges, and synchronization methods are tailored to the pilot sequence. In this paper, we propose a framework for initial synchronization that works independently of the pilot sequence. We show that classical data-aided and blind techniques are particular cases of this general framework. We thoroughly study the specific problem of OFDM data-aided carrier frequency offset estimation. For the latter, we provide theoretical expressions of performance upper-bounds. Also, a practical thin CFO estimator for OFDM is provided under the form of a novel algorithm which is shown by simulations to perform better than classical pilot based methods.

## I. INTRODUCTION

Information theory, which originates from the pioneering work of Shannon [1], provides a theoretical expression of the optimum rate to which data can be reliably exchanged between a transmitting entity  $A$  and a receiving entity  $B$ , that is called channel capacity. This capacity is linked to the mutual information between  $A$  and  $B$ , i.e. the amount of respective information  $X_A$  and  $X_B$  that both entities share. If  $X_A$  and  $X_B$  both contain all relevant synchronization information but the transmitted data themselves, i.e. central frequency, symbol timing, transmission channel are known by both  $A$  and  $B$ , then the capacity is known and many practical solutions exist which can achieve transmission data rates close to the theoretical limit. When the synchronization parameters are unknown though, especially to the receiver, it is difficult to derive the capacity. This led to consider auxiliary methods such as channel estimation, frequency offset tracking etc. which do not follow the information theoretic framework proposed by Shannon. In the following, we cast our attention to initial synchronization and particularly to initial carrier frequency offset (CFO) estimation in orthogonal frequency division multiplexing (OFDM) systems; CFO is defined as the frequency mismatch between the transmitter and the receiver.

In the *initial* synchronization phase, both transmitter and receiver know little about the surrounding environment. To cope with this problem, *ad-hoc* methods have been designed that are insensitive to the unknown parameters. In the specific example of CFO estimation for OFDM, the historical method, due to Moose [2], is based on a pilot sequence designed to this purpose. This pilot is handily made such that Moose's

estimator is independent of the channel realization, to the expense of a short bandwidth acquisition range. Those pilot-based methods are often called data-aided (DA) algorithms. Non data-aided (NDA) algorithms were also proposed which exploit the cyclic prefix redundancy [5], [6]. However, all those solutions originate from *ad-hoc* ideas and do not rely on theoretical foundations. Advanced methods based on Bayesian probabilities or orthodox parameter estimation already exist [7]-[9]. In [7], synchronization is performed using the decision errors in decoded symbols. In [8], theoretical bounds on joint channel estimation and sampling clock offset are derived. However, those schemes are often referred to as thin synchronization methods as they assume the *a priori* knowledge of many system parameters; the problem of initial synchronization is therefore not addressed.

Since the essential issue in initial synchronization is a lack of information, we propose to derive an information theoretic framework for synchronization, through the example of CFO estimation. To cope with the problem of missing information, we extensively refer to the information theoretic tools that are Cox's probability theory as extended logic [10] and Jaynes' maximum entropy principle [11]. Given a prior system information  $I$ , Cox's theory describes in probability assignments the degree of confidence on any system parameter  $\theta$ ; the inferred information about  $\theta$  is denoted  $(\theta|I)$  and the assigned probability is denoted  $p(\theta|I)$ . If  $I$  only contains deterministic information,  $p(\theta|I)$  is computed from Bayes' rule, while if  $I$  contains stochastic information about  $\theta$ ,  $p(\theta|I)$  is determined by the maximum entropy principle.

This paper is organized as follows: In Section II, we provide the system model with CFO mismatch from which the general Bayesian CFO estimation framework unfolds. Particular attention is then cast on the data-aided CFO estimation for OFDM systems. From this study, a practical CFO estimation algorithm is provided. Simulations are then proposed in Section III which show an important gain of the Bayesian techniques over classical DA methods. The results are then discussed in Section IV before we provide our conclusions in Section V.

*Notation:* In the following, boldface lower-case symbols represent vectors, capital boldface characters denote matrices ( $\mathbf{I}_N$  is the  $N \times N$  identity matrix). The Hermitian transpose is denoted  $(\cdot)^H$ . The operators  $\det(\mathbf{X})$  and  $\text{tr}(\mathbf{X})$  represent

the determinant and the trace of matrix  $\mathbf{X}$ , respectively. The symbol  $E[\cdot]$  denotes expectation.

## II. FREQUENCY OFFSET ESTIMATION

Consider a pair of transmitter and receiver communicating through a noisy channel. The transmitter sends a data sequence  $\mathbf{x} \in \mathbb{C}^N$  of length  $N$  which the receiver captures as a sequence  $\mathbf{y} \in \mathbb{C}^N$ . The transmission memory channel of length  $L$  time symbols is denoted  $\mathbf{h} \in \mathbb{C}^L$ . The noise process is an additive white Gaussian (AWGN) sequence  $\mathbf{w} \in \mathbb{C}^N$ . The knowledge of the receiver, prior to data transmission, is denoted  $I$ . The receiver frequency reference is not perfectly aligned to that of the transmitter: this introduces a frequency offset  $\theta$  whose knowledge to the receiver is summarized into the density function  $p(\theta|I)$ . By inductive reasoning, we provide in the following an expression of the optimal inference the receiver can make on  $(\theta|\mathbf{y}, I)$  which we apply to the example of DA CFO estimation in OFDM.

### A. General communication model

From Bayes' rule, the probability distribution  $p(\theta|\mathbf{y}, I)$  can be broken into

$$p(\theta|\mathbf{y}, I) = p(\mathbf{y}|\theta, I) \frac{p(\theta|I)}{p(\mathbf{y}|I)} \quad (1)$$

in which  $p(\mathbf{y}|I)$  is independent of  $\theta$ .

If  $\theta$  is known to be bounded or  $\theta$  is known to enjoy some stochastic properties, the maximum entropy principle provides a unique prior distribution  $p(\theta|I)$ .

Then, by marginalization over  $\mathbf{h}$  and  $\mathbf{x}$ ,  $p(\mathbf{y}|\theta, I)$  further reads

$$p(\mathbf{y}|\theta, I) = \int_{\mathbf{h}} \int_{\mathbf{x}} p(\mathbf{y}|\theta, \mathbf{h}, \mathbf{x}, I) p(\mathbf{h}, I) p(\mathbf{x}, I) d\mathbf{x} d\mathbf{h} \quad (2)$$

This is the most general formulation to evaluate  $p(\mathbf{y}|\theta, I)$  and then  $p(\theta|\mathbf{y}, I)$ . When both priors on  $\mathbf{h}$  and  $\mathbf{x}$  are rather uniform on their definition space, i.e. not peaky around a particular value, the resulting distribution for  $(\theta|\mathbf{y}, I)$  is also broad and does not lead to efficient estimators. Much more experience (through additional data transmission) is required for such schemes to be valuable. This explains why NDA methods, which rely on little knowledge, require lots of data to converge to a satisfying solution.

If  $\mathbf{x}$  is *a priori* known to the receiver, then  $I$  contains all information about  $\mathbf{x}$  and  $p(\mathbf{x}|I)$  is a single mass in the position of  $\mathbf{x}$ . In this case, Equation (2) reduces to

$$p(\mathbf{y}|\theta, I) = \int_{\mathbf{h}} p(\mathbf{y}|\theta, \mathbf{h}, I) p(\mathbf{h}, I) d\mathbf{h} \quad (3)$$

This situation is the general framework of DA methods. The remaining problem is to assign a prior density  $p(\mathbf{h}|I)$  to the parameter  $\mathbf{h}$ . It is desirable to attach to  $\mathbf{h}$  the *most non-committal* probability distribution which fulfills the constraints gathered in the information  $(\mathbf{h}|I)$ , in order to avoid subjective assumptions. This distribution is shown by Jaynes to be the distribution which maximizes the entropy of  $\mathbf{h}$  and is consistent with  $I$  [11].

In particular, consider  $\mathbf{h}$  is a multi-path channel. If the receiver knows  $\mathbf{h}$  is composed of  $L$  paths and is of average variance  $E[\mathbf{h}\mathbf{h}^H] = 1$ , then [14] demonstrates that the maximum entropy distribution for  $(\mathbf{h}|I)$  is

$$p(\mathbf{h}|I) = \frac{1}{\pi^L} e^{-L\mathbf{h}^H\mathbf{h}} \quad (4)$$

which is the classical Gaussian i.i.d. (independent and identically distributed) channel. Interesting considerations on channel modeling through the maximum entropy principle are discussed in [14]. In particular, if the covariance matrix  $\mathbf{Q} = E[\mathbf{h}\mathbf{h}^H]$  is *a priori* known, then the maximum entropy principle associates to  $(\mathbf{h}|I)$  the more general Gaussian distribution,

$$p(\mathbf{h}|I) = \frac{1}{\pi^L \det(\mathbf{Q})} e^{-\mathbf{h}^H \mathbf{Q}^{-1} \mathbf{h}} \quad (5)$$

From Equation (4), the optimal inference on  $\theta$  posterior to data reception can then be computed. The corresponding minimum mean square error (MMSE) frequency offset estimator then reads

$$E[\theta|\mathbf{y}, I] = \int_{\mathbb{R}} \theta \cdot p(\theta|\mathbf{y}, I) d\theta \quad (6)$$

### B. DA CFO estimation for OFDM

1) *Inductive reasoning on  $(\theta|\mathbf{y}, I)$* : Consider an OFDM system with  $N$  subcarriers. The transmitter sends a time-domain pilot sequence  $\mathbf{x} = (x_0, \dots, x_{N-1})^T$  (cyclic prefix excluded), received as a sequence  $\mathbf{y} = (y_0, \dots, y_{N-1})^T$  (cyclic prefix discarded). The transmission channel is discretized in  $L$  taps  $\mathbf{h} = (h_0, \dots, h_{L-1})^T$  and the AWGN noise  $\mathbf{w} = (w_0, \dots, w_{N-1})^T$  has entries of variance  $E[|w_k|^2] = \sigma^2$ . For simplicity reasons, we will not consider the information contained in the cyclic prefixes in the rest of this study. Let  $\theta$  represent the CFO to be estimated at the receiver, normalized to the subcarrier spacing, i.e.  $\theta = 1$  is a frequency mismatch of one subcarrier spacing.

CFO in OFDM engenders a rotation of the time-domain symbols  $x_k$  of an angle  $2\pi k\theta/N$ . Denote  $\mathbf{D}_\theta$  the diagonal matrix of main diagonal  $\mathbf{d}_\theta = (1, e^{2\pi \frac{i\theta}{N}}, \dots, e^{2\pi \frac{(N-1)i\theta}{N}})^T$  and  $\mathbf{X}$  the matrix

$$\mathbf{X} = \begin{pmatrix} x_0 & x_{N-1} & \cdots & x_{N-L-1} \\ x_1 & x_0 & \cdots & x_{N-L-2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{L-2} & x_{L-3} & \cdots & x_{N-1} \\ x_{L-1} & x_{L-2} & \cdots & x_0 \\ \vdots & \vdots & \vdots & \vdots \\ x_{N-1} & x_{N-2} & \cdots & x_{N-L} \end{pmatrix} \quad (7)$$

Then at the receiver,

$$\mathbf{y} = \mathbf{D}_\theta \mathbf{X} \mathbf{h} + \mathbf{w} \quad (8)$$

Assume the receiver knows the covariance matrix  $\mathbf{Q}$  of  $\mathbf{h}$ .

Then, from (5),

$$p(\mathbf{y}|\theta, I) = \frac{1}{\pi^{N+L}\sigma^{2N} \det(\mathbf{Q})} \times \int e^{-\frac{1}{\sigma^2}(\mathbf{D}_\theta \mathbf{y} - \mathbf{X}\mathbf{h})^H(\mathbf{D}_\theta \mathbf{y} - \mathbf{X}\mathbf{h})} e^{-\mathbf{h}^H \mathbf{Q}^{-1} \mathbf{h}} d\mathbf{h} \quad (9)$$

$$= \frac{1}{\pi^{N+L}\sigma^{2N} \det(\mathbf{Q})} \times \int e^{-(\mathbf{h} - \tilde{\mathbf{h}})^H \mathbf{M}(\mathbf{h} - \tilde{\mathbf{h}}) - C(\theta)} d\mathbf{h} \quad (10)$$

with

$$\begin{cases} \mathbf{M} &= \frac{1}{\sigma^2} \mathbf{X}^H \mathbf{X} + \mathbf{Q}^{-1} \\ \tilde{\mathbf{h}} &= \frac{1}{\sigma^2} \mathbf{M}^{-1} \mathbf{X}^H \mathbf{D}_\theta \mathbf{y} \\ C(\theta) &= \mathbf{y}^H \underbrace{\mathbf{D}_\theta^H \mathbf{D}_\theta}_{=\mathbf{I}_N} \mathbf{y} - \tilde{\mathbf{h}}^H \mathbf{M} \tilde{\mathbf{h}} \end{cases} \quad (11)$$

Note that if, as is almost always the case, the correlation matrix  $\mathbf{Q}$  is *a priori* unknown to the receiver but only  $L$  is known, the maximum entropy principle states (see Equation (4)) that  $\mathbf{h}$  is Gaussian i.i.d. The previous study is then still valid with  $\mathbf{Q} = \frac{1}{L} \mathbf{I}_L$ . But again, it is rare that in the initial synchronization step an exact value for  $L$  is known. A marginalization of Equation (9) over all possible values for the number of channel taps is necessary. This is further discussed in Section IV. In the following we keep the assumption that  $L$  (or  $\mathbf{Q}$ ) is known.

This leads to

$$p(\mathbf{y}|\theta, I) = \frac{1}{\pi^N \sigma^{2N} \det(\mathbf{QM})} e^{-C(\theta)} \quad (12)$$

in which only  $C$  depends on  $\theta$ . Expanding  $C(\theta)$  we have,

$$C(\theta) = \mathbf{y}^H \left[ \mathbf{I} - \mathbf{D}_\theta^H \mathbf{X} (\mathbf{X}^H \mathbf{X} + \sigma^2 \mathbf{Q}^{-1})^{-1} \frac{1}{\sigma^2} \mathbf{X}^H \mathbf{D}_\theta \right] \mathbf{y} \quad (13)$$

from which we isolate the part dependent on  $\theta$  in the function  $C'$ ,

$$C'(\theta) = -\mathbf{y}^H \mathbf{D}_\theta^H \mathbf{X} (\mathbf{X}^H \mathbf{X} + \sigma^2 \mathbf{Q}^{-1})^{-1} \frac{1}{\sigma^2} \mathbf{X}^H \mathbf{D}_\theta \mathbf{y} \quad (14)$$

which can be rewritten, with  $\mathbf{Y} = \text{diag}(\mathbf{y})$ , and  $\mathbf{A} = \mathbf{Y}^H \mathbf{X} (\mathbf{X}^H \mathbf{X} + \sigma^2 \mathbf{Q}^{-1})^{-1} \frac{1}{\sigma^2} \mathbf{X}^H \mathbf{Y}$ ,

$$C'(\theta) = -\mathbf{d}_\theta^H \mathbf{Y}^H \mathbf{X} (\mathbf{X}^H \mathbf{X} + \sigma^2 \mathbf{Q}^{-1})^{-1} \frac{1}{\sigma^2} \mathbf{X}^H \mathbf{Y} \mathbf{d}_\theta \quad (15)$$

$$= -\sum_{n=1}^N \sum_{m=1}^N e^{-2\pi i \theta \frac{n-m}{N}} A_{n,m} \quad (16)$$

$$= -2\Re \left[ \sum_{k=1}^{N-1} e^{-2\pi i \theta \frac{k}{N}} \tilde{a}_k \right] - 2 \sum_{n=1}^N A_{n,n} \quad (17)$$

in which  $\tilde{a}_k = \sum_{m=1}^{N-k} A_{k+m,m}$  for  $1 \leq k \leq N-1$ ; Equation (17) is obtained from the Hermitian property of  $\mathbf{A}$ .

Denote  $f(\theta) = \Re \left[ \sum_{k=1}^{N-1} e^{-2\pi i \theta \frac{k}{N}} \tilde{A}_k \right]$ , we finally have

$$\log(p(\theta|\mathbf{y}, I)) = \alpha + \log(p(\theta|I)) + f(\theta) \quad (18)$$

for some constant  $\alpha$  independent of  $\theta$ .

The Bayesian inference on  $(\theta|\mathbf{y}, I)$  is then completely determined by the prior density function  $p(\theta|I)$  and  $f$ .

2) *CFO estimators*: The estimation of  $\theta$  consists in electing, from the inferred information  $p(\theta|\mathbf{y}, I)$ , the value  $\hat{\theta}$  that optimally approximates  $\theta$ . This optimal approximation requires to minimize some cost function. Usually, MMSE or minimum absolute error estimators are chosen but in the context of synchronization parameter estimation, it is difficult to decide which cost function would better suit our initial desire: maximize the channel capacity or minimize decoding errors given  $\hat{\theta}$ . In the following, we consider two classical approaches: MMSE and maximum *a posteriori* (MAP) estimations.

The MMSE estimate  $\hat{\theta}_{\text{MMSE}}$  minimizes the mean quadratic error  $E[(\theta - \hat{\theta})^2]$  in which the expectation is taken over the possible received signals  $\mathbf{y}$  and over the possible CFO  $\theta$ . It is given by [15]

$$\hat{\theta}_{\text{MMSE}} = E[\theta|\mathbf{y}, I] = e^\alpha \int_{-\infty}^{+\infty} \theta \cdot p(\theta|I) \cdot e^{f(\theta)} d\theta \quad (19)$$

Unfortunately, this integral is mathematically involved and will stay under this analytic form.

For practical (less mathematically demanding) usage, it seems *desirable* to consider the MAP estimator, which provides the *most likely* value for  $\theta$  given  $\mathbf{y}$ . However, particular care is demanded to handle this estimator. As proven in [13]-Chapter 13, the cost function that this estimator minimizes is the limit function  $\lim_{x \rightarrow 0} E[|\theta - \hat{\theta}|^x]$ . This suggests that large errors on the estimation of  $\theta$  are as bad as small errors. In particular, if  $p(\theta|\mathbf{y}, I)$  is not very peaky in the vicinity of the true CFO, this estimator is expected to be very inefficient.

In our situation, the MAP estimator  $\hat{\theta}_{\text{MAP}}$  reads

$$\hat{\theta}_{\text{MAP}} = \max_{\theta} [\log(p(\theta|I)) + f(\theta)] \quad (20)$$

In most situations, the prior knowledge upon  $(\theta|I)$  is difficult to describe in either deterministic or stochastic terms. We consider only here the knowledge of bounds  $\theta_{\min}$  and  $\theta_{\max}$  on the values for  $\theta$ , such that  $(\theta|I)$  is uniform on the set  $[\theta_{\min}, \theta_{\max}]$  and equation (20) can be rewritten

$$\hat{\theta}_{\text{MAP}} = \max_{\theta_{\min} \leq \theta \leq \theta_{\max}} f(\theta) \quad (21)$$

From extensive simulations, we observe without proving it that  $f$  is concave in the domain  $[-\frac{1}{2}, \frac{1}{2}]$ , i.e. one subcarrier spacing, while it is not concave on the total bandwidth  $[-\frac{N}{2}, \frac{N}{2}]$ . This suggests that a thin estimation for  $\theta$  can be found by gradient descent methods in the interval  $[-\frac{1}{2}, \frac{1}{2}]$  while only coarse estimation can be handled on the larger interval  $[-\frac{N}{2}, \frac{N}{2}]$ . An algorithm is proposed in the following to tackle the thin estimation problem.

### C. Algorithm

Since it appears that  $f$  is concave on the space  $[-1/2, 1/2]$ , we consider in the following a steepest descent method to perform CFO estimation. Note first that  $f$  can be rewritten as the sum of two products,

$$f(\theta) = \Re \left[ \sum_{k=1}^{N-1} e^{-2\pi i \theta \frac{k}{N}} \tilde{a}_k \right] = \tilde{\mathbf{a}}_{\Re}^T \cdot \mathbf{c}_\theta + \tilde{\mathbf{a}}_{\Im}^T \cdot \mathbf{s}_\theta \quad (22)$$

with

$$\begin{cases} \tilde{\mathbf{a}}_{\Re}^{\top} &= \Re[\tilde{a}_1, \dots, \tilde{a}_{N-1}] \\ \tilde{\mathbf{a}}_{\Im}^{\top} &= \Im[\tilde{a}_1, \dots, \tilde{a}_{N-1}] \\ \mathbf{c}_{\theta} &= [\cos(2\pi\theta/N), \dots, \cos(2\pi(N-1)\theta/N)] \\ \mathbf{s}_{\theta} &= [\sin(2\pi\theta/N), \dots, \sin(2\pi(N-1)\theta/N)] \end{cases} \quad (23)$$

From (22), we deduce the first derivative  $f'$  of  $f$  as

$$f'(\theta) = \frac{2\pi}{N} \left( -\tilde{\mathbf{b}}_{\Re}^{\top} \cdot \mathbf{s} + \tilde{\mathbf{b}}_{\Im}^{\top} \cdot \mathbf{c} \right) \quad (24)$$

where,  $\forall k \in \{1, \dots, N-1\}$ ,  $\tilde{b}_{\Re,k} = k \cdot \tilde{a}_{\Re,k}$  and  $\tilde{b}_{\Im,k} = k \cdot \tilde{a}_{\Im,k}$ .

Now the steepest descent algorithm can be described. Assume first that  $p(\theta|I)$  is uniform on  $[-1/2, 1/2]$ . Given some termination constraint, the recursive algorithm unfolds as follows,

- 1) at signal reception, compute the vectors  $\tilde{\mathbf{b}}_{\Re}$  and  $\tilde{\mathbf{b}}_{\Im}$ .
- 2) *initialization*: set the variables  $\theta_{\min} = -\frac{1}{2}$  and  $\theta_{\max} = \frac{1}{2}$ .
- 3) *loop*: while the termination constraint is not reached, compute  $d = f'([\theta_{\min} + \theta_{\max}]/2)$ .
  - if  $d > 0$ , set  $\theta_{\min} = [\theta_{\min} + \theta_{\max}]/2$ .
  - if  $d < 0$ , set  $\theta_{\max} = [\theta_{\min} + \theta_{\max}]/2$ .
- 4) at termination, set  $\hat{\theta} = [\theta_{\min} + \theta_{\max}]/2$ .

The algorithm converges to the MAP value for the CFO. If a non-uniform prior is chosen for  $\theta$ , one has to ensure that the *a posteriori* (18) is concave on the set  $[-1/2, 1/2]$ , in which case the function to be maximized is  $f(\theta) + \log(p(\theta|I))$ . Note that, for “smooth” priors  $p(\theta|I)$ , even if  $f(\theta) + \log(p(\theta|I))$  is not truly concave, it is expected to behave nicely around the exact value of  $\theta$  and then the aforementioned algorithm can be further refined.

### III. SIMULATION AND RESULTS

In the following we consider an OFDM transmission with  $N = 128$  subcarriers. We assume perfect timing offset alignment between the base station and the receiving terminal. A CFO mismatch  $\theta$  is introduced. The receiver only knows that  $\theta \in [-1/2, 1/2]$ . In Figure 1, we use as a pilot the double-half sequence suggested by Moose [2]. The Bayesian MAP method is compared against the Moose’s correlation algorithm on 20,000 channels and CFO realizations ( $\theta$  is uniformly distributed in  $[-1/2, 1/2]$ ). The channel length is set to  $L = 3$ , while the *a priori* on the channel length is either considered known, i.e.  $L_{\text{assumed}} = 3$ , or wrongly estimated, here  $L_{\text{assumed}} = 9$ . The respective performances are analyzed in terms of average quadratic error  $E[(\hat{\theta} - \theta)^2]$ . We observe a significant performance gain provided by the Bayesian method, especially in low SNR regime. The Bayesian estimator is indeed more able to cope with the noise impairment which is more thoroughly modelled than in Moose’s algorithm. Note also that a wrongly assigned prior  $p(\mathbf{h}|I)$  on the channel realization does not lead to critical performance decay; in the high SNR region, this is almost unimportant.

In Figure 2, we analyze the performance of the steepest descent algorithm proposed in Section II-C. The system parameters are the same as in the previous simulation, with a

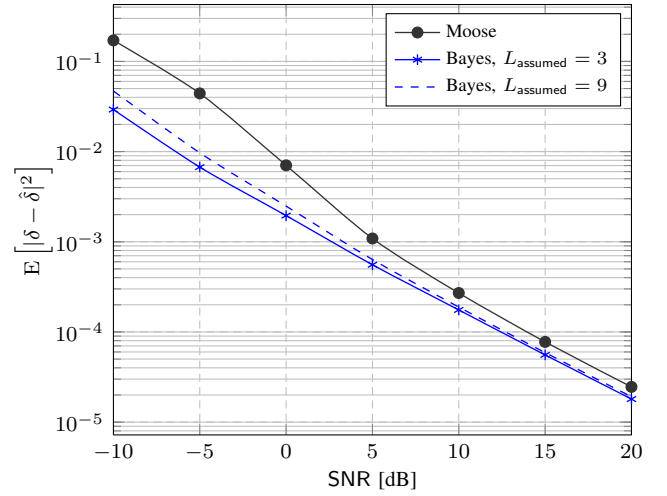


Fig. 1. Moose and Bayesian CFO estimates, MSE comparison -  $N = 128$ ,  $L = 3$ ,  $L_{\text{assumed}} \in \{3, 9\}$

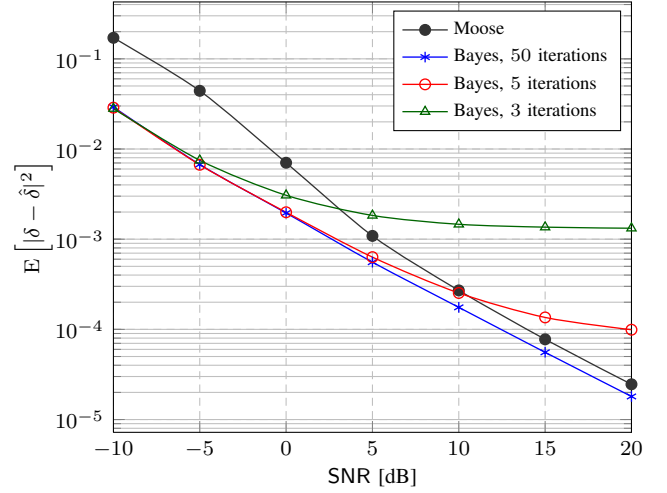


Fig. 2. Bayesian CFO estimation, Steepest descent algorithm -  $N = 128$ ,  $L = 3$

correct prior  $L_{\text{assumed}} = 3$  on the channel length at the receiver. The termination constraint is simply the number of iterations  $k$  of the inner loop, which we limit to  $k = 3$ ,  $k = 5$ ,  $k = 10$  and  $k = 50$ . It is observed that saturations appear for small  $k$ , which are explained by the systematic error introduced by the minimal step size  $2^{-k}$  in the iteration loop. For  $k \geq 10$ , the performance plots fit the plot  $k = 50$  in the  $-15$  dB to  $10$  dB SNR range. Note also that the saturated standard deviation for  $k = 5$  is around 1% of the subcarrier spacing, which corresponds to the maximum allowable CFO mismatch in most OFDM systems. Therefore, 5 iterations might be sufficient to ensure a reliable estimation of the CFO.

### IV. DISCUSSION

As previously mentioned, the expression  $p(\mathbf{y}|\theta, I)$  given by equation (9) is only valid when the channel length  $L$  is known to the receiver, i.e. it is part of the prior information  $I$ . This

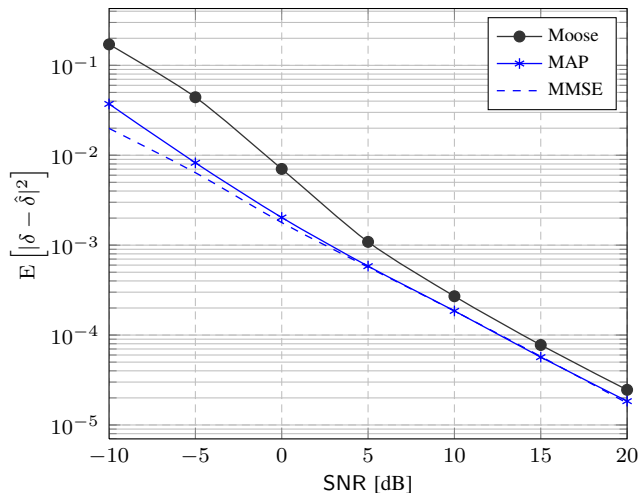


Fig. 3. Bayesian CFO estimation, MMSE against MAP estimation -  $N = 128$ ,  $L = 3$

would rarely be the case in the initial synchronization step of classical OFDM systems. Instead, one should consider the prior  $I_0$  which contains some statistical information about  $L$ , and therefore one has,

$$p(\mathbf{y}|\theta, I_0) = \int_L p(\mathbf{y}|\theta, L, I_0) \cdot p(L|I_0) dL \quad (25)$$

In the OFDM case,  $L$  is usually taken as an integer, and therefore one can write

$$p(\mathbf{y}|\theta, I_0) = \sum_{L=1}^{N_{CP}} p(\mathbf{y}|\theta, L, I_0) \cdot p(L|I_0) \quad (26)$$

where  $N_{CP}$  is the cyclic prefix length.

The computation is no more involved but is computationally demanding in practice when  $N_{CP}$  is large. As we observed previously that an error in the *a priori* for  $L$  does not lead to a significant CFO estimation error,  $p(\mathbf{y}|\theta, I_0)$  can be in general well approximated by Equation (9). If not, the sum (26) can be restricted to a subset of  $\{1, \dots, N_{CP}\}$ .

Concerning the choice of the MAP estimator, we mentioned earlier that this estimator might turn out hazardous as large errors in the sense of absolute difference are as good as small errors. However, when the posterior distribution  $p(\theta|\mathbf{y}, I)$  is very peaky in the vicinity of the true value for  $\theta$ , any classical estimator is alike. In particular, the only relevant integrands of the MMSE estimator (19) are those located around the distribution peak, which tends to a single mass in the high SNR region; as a consequence, the MMSE estimate asymptotically corresponds to the MAP estimator. Figure 3 provides a comparison between MMSE and MAP which shows equal performance for  $\text{SNR} > 5$  dB; the MMSE estimator was computed from formula (19) in which the integrals were approximated by discrete sums with step  $1/256$ .

Note also that we provided a general framework for initial synchronization which considers fixed channels and a limited amount of input data  $\mathbf{x}$ . However, dynamics in the

channel as well as updated prior information are not taken into account. Therefore, successive parameter estimates enjoy the same performance. Ideally, the estimate of a parameter  $\theta_{t_0}$  at time  $t_0$  should take into account all the additional information provided at time  $t < t_0$ . Updating Bayesian probabilities is treated in [12], which allows to infer on  $\theta_{t_0}$  given the prior information  $I$ , the previous estimate  $\theta_{t_0-1}$  and some information on the system dynamics. These probability updates are performed thanks to the so-called ME method.

## V. CONCLUSION

In this work, an information theoretic framework for initial synchronization is provided. The case study of carrier frequency offset estimation in OFDM systems is developed and expressions for performance upper-bounds are provided. Novel data-aided estimators are also proposed, which are shown to perform better than classical techniques. This gives birth to a practical thin CFO algorithm. Simulations are carried out which analyze the behaviour and confirm the gain in performance of this new algorithm.

## REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communications", Bell System Technical Journal, vol. 27, no. 7, pp. 379-423, 1948.
- [2] P. H. Moose, "A technique for orthogonal frequency-division multiplexing frequency offset correction", IEEE Trans. on Communications, vol. 42, no. 10, pp. 2908-2914, Oct. 1994.
- [3] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM", IEEE Trans. on Communications, vol. 45, no.12, pp. 1613-1621, Dec. 1997.
- [4] M. Morelli and U. Mengali, "An improved frequency offset estimator for OFDM applications", IEEE Comm. Letters, vol. 3, no. 3, pp. 75-77, Mars 1999.
- [5] J. J. Van de Beek, M. Sandelland and P. O. Börjesson, "ML estimation of time and frequency offset correction in OFDM systems", IEEE Trans. on Signal Processing, vol. 45, no. 7, pp. 1800-1805, July 1997.
- [6] B. Ai, J. Ge, Y. Wang, S. Yang and P. Liu, "Decimal frequency offset estimation in COFDM wireless communications", IEEE Trans. on Broadcasting, vol. 50, no. 2, June 2004.
- [7] K. Shi, E. Serpedin and P. Ciblat, "Decision-Directed Fine synchronization in OFDM systems", IEEE Transactions on Communications, vol. 53, no. 3, pp. 408-412, Mars 2005.
- [8] S. Gault, W. Hachem and P. Ciblat, "Joint Sampling Clock Offset and channel estimation for OFDM signals : Cramer-Rao bound and Algorithms", IEEE Transactions on Signal Processing, vol. 54, no. 5, pp. 1875-1885, Mai 2006.
- [9] N. Noels, C. Herzet, A. Dejonghe, V. Lottici, H. Steendam, M. Moeneclaey, M. Luise and L. Vandendorpe, "Turbo synchronization: an EM algorithm interpretation", IEEE Int. Conf. on Communications, vol. 4, pp. 11-15, 2003.
- [10] R. T. Cox, "Probability, frequency and reasonable expectation", American journal of physics, vol. 14 (1946), pp. 1-13.
- [11] E. T. Jaynes, "Information Theory and Statistical Mechanics", Physical Review, APS, vol. 106, no. 4, pp. 620-630, 1957.
- [12] A. Caticha, "Lectures on Probability, Entropy and Statistical Physics", arXiv:0808.0012v1 [physics.data-an], 2008.
- [13] E. T. Jaynes, "Probability Theory: The Logic of Science", Cambridge University Press, June 2003.
- [14] M. Guillaud, M. Debbah, A. L. Moustakas, "Maximum Entropy MIMO Wireless Channel Models", Submitted to IEEE Trans. Information Theory Dec. 2006, <http://arxiv.org/abs/cs.IT/0612101>
- [15] S. M. Kay, "Fundamentals of statistical signal processing: estimation theory", Prentice Hall, New Jersey, 1993.