A maximum entropy approach to OFDM channel estimation

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 this work, a Bayesian framework for OFDM channel estimation is proposed. Using the maximum entropy principle to exploit prior system information at the receiver, we successively derive channel estimates in situations when (i) the channel delay spread and (ii) the channel time correlation statistics are *a priori* unknown. More generally, this framework allows to derive MMSE channel estimates under any state of knowledge at the receiver. Simulations are provided that confirm the theoretical claims and show the novel results to perform as good or better than classical estimators.**

I. Introduction

Modern high rate wireless communication systems, such as 3GPP-Long Term Evolution (LTE) [1], usually come along with large bandwidths. In multipath fading channels, this entails high frequency selectivity, which theoretically is beneficial for it provides channel diversity. But in practice, this constitutes a strong challenge for equalization. Orthogonal frequency division multiplexing (OFDM) modulation [2] allows for simplified equalization and channel estimation based on pilot symbols scattered in the time-frequency grid and possibly over the space dimension when multiple antennas are used.

One challenge in channel estimation with a limited number of pilots is to correctly exploit the prior information at the receiver. Classically only data received from pilot positions are considered informative. As a consequence, between pilot positions, the estimated channel must be reconstructed using satisfactory interpolation techniques, e.g. [3]. A Bayesian minimum mean square error (MMSE) [4] estimator can be derived when not only the pilot sequences but also the channel covariance matrix are known [5]. However, when the latter is unknown, only *ad-hoc* techniques have been proposed to cope with the lack of information, e.g. by making a definite choice of a prior channel correlation matrix (identity matrix, exponentially decaying matrix [6] etc.). However, all those techniques are only justified by good performance arising in selected field simulations and do not provide any proof as for their overall performance.

In the following work, we tackle channel estimation for OFDM as a problem of inductive reasoning based both on received pilots and on the available prior information at the receiver. To cope with missing information, we extensively use the maximum entropy principle, shown by Jaynes [7] to be the desirable mathematical tool to deal with limited information. Some of the aforementioned classical results will be found anew and proven optimal in our framework, while new results will be provided which show to perform better than classical approaches. The remainder of this paper is structured as follows: in Section II, we introduce the channel pilot-aided OFDM system, then in Section III, we carry out the Bayesian channel estimation study based on different levels of knowledge. Simulations are then proposed in Section IV. We finally give our conclusions in Section V.

Remark 1: In the remainder of this document, due to page limitation, complete derivations are not provided. Those are fully developed in an extended version of this work [13].

Notations: Boldface lower case symbols represent vectors, capital boldface characters denote matrices ($\mathbf{I}_N$ is the $N \times N$ identity matrix). The transpose and Hermitian transpose are denoted $(\cdot)^\mathsf{T}$ and $(\cdot)^\mathsf{H}$. The operator $\mathrm{diag}(\mathbf{x})$ turns the vector $\mathbf{x}$ into a diagonal matrix. The symbol $\det(\mathbf{X})$ is the determinant of $\mathbf{X}$. The symbol $\mathrm{E}[\cdot]$ denotes expectation. The Kronecker function δ_x equals 1 if $x = 0$ and equals 0 otherwise.

II. System Model

Consider a single cell OFDM system with N subcarriers. The cyclic prefix (CP) length is N_{CP} samples. In the time-frequency OFDM symbol grid, pilots are found in the symbol positions indexed by the function $\phi_t(n) \in \{0, 1\}$ which equals 1 if a pilot symbol is present at subcarrier n, at symbol time index t, and 0 otherwise. The time-frequency grid is depicted in Figure 1. Both data and pilots at time t are modeled by the frequency-domain vector $\mathbf{s}_t \in \mathbb{C}^N$ with pilot entries of amplitude $|s_{t,k}|^2 = 1$. The transmission channel is denoted $\mathbf{h}_t \in \mathbb{C}^N$ in the frequency-domain with entries of variance $\mathrm{E}[|h_{t,k}|^2] = 1$. The additive noise is denoted $\mathbf{n}_t \in \mathbb{C}^N$ with entries of variance $\mathrm{E}[|n_{t,k}|^2] = \sigma^2$. Since this variance is the only available information on $\mathbf{n}_t$, the maximum entropy principle [10] requires that the noise process be assigned a Gaussian independent and identically distributed (i.i.d.) density[1], $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_N)$. The time-domain representation of $\mathbf{h}_t$ is denoted $\boldsymbol{\nu}_t \in \mathbb{C}^L$ with L the channel length, i.e. the

[1] the reason is that the distribution which maximizes entropy under variance constraint is Gaussian i.i.d., see e.g. [12].

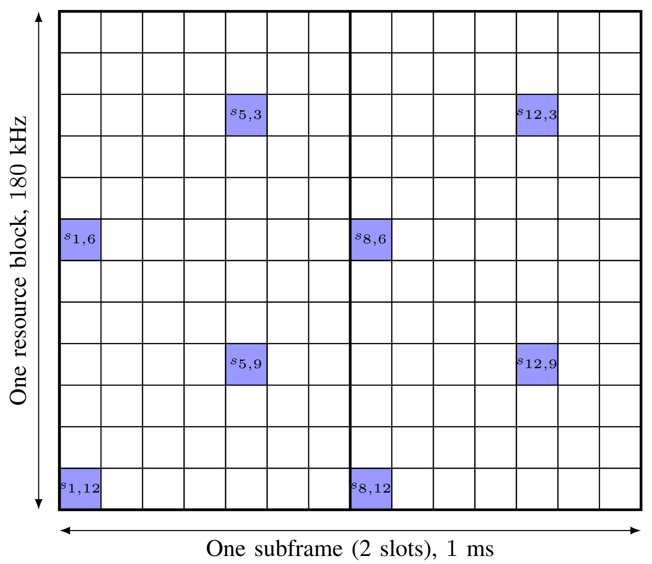

Fig. 1. Time-frequency OFDM grid with pilot positions enhanced

channel delay spread expressed in OFDM-sample unit. The frequency-domain received signal $\mathbf{y}_t \in \mathbb{C}^N$ is then

$$\mathbf{y}_t = \operatorname{diag}(\mathbf{h}_t)\mathbf{s}_t + \mathbf{n}_t \tag{1}$$

We will also denote, $\forall k \in \{1, \ldots, N\}$, $h'_{t,k} = y_{t,k}/s_{t,k} = h_{t,k} + n_{t,k}/s_{t,k}$ and $\mathbf{h}'_t = (h'_{t,1}, \ldots, h'_{t,N})^\mathsf{T}$.

This work aims at providing MMSE estimates $\hat{\mathbf{h}}_t \in \mathbb{C}^N$ of the vector $\mathbf{h}_t$ for different states of knowledge at the receiver: (i) the channel length L is either known or unknown, (ii) at discrete time t_0, the pilots received at time $t < t_0$ as well as the channel time-correlation, are either known or unknown. The total prior information in either case will be denoted I and the amount of information that can be inferred on the statement E from I will be denoted $(E|I)$.

III. Channel Estimation

A. The channel length is known

First consider that the channel power delay profile, i.e. $\mathrm{E}[\boldsymbol{\nu}_t \boldsymbol{\nu}^\mathsf{H}]$, is unknown. Only the channel length L and the received pilot sequence at discrete time t are known to the receiver[2]. For ease of reading, we remove the index t in the notations when unnecessary.

From the channel model (1), the multipath channel of length L is only known to be of unit mean power. The maximum entropy principle therefore demands $\boldsymbol{\nu} \sim \mathcal{CN}(0, \frac{1}{L}\mathbf{I}_L)$. Since $\boldsymbol{\nu}$ is Gaussian with i.i.d. entries, its discrete Fourier transform is a correlated Gaussian vector $\mathbf{h} \sim \mathcal{CN}(0, \mathbf{Q})$ with, for $(n, m) \in \{1, \ldots, N\}^2$,

$$Q_{nm} = \mathrm{E}\left[\sum_{k=0}^{L-1}\sum_{l=0}^{L-1} \nu_k \nu_l^* e^{-2\pi i \frac{kn-lm}{N}}\right] = \frac{1}{L}\sum_{k=0}^{L-1} e^{-2\pi i k \frac{n-m}{N}} \tag{2}$$

[2] we assume then first that the receiver is not able to remember either past received signals, nor past estimates of the channel.

The MMSE channel estimator $\hat{\mathbf{h}}$ knowing $\mathbf{y}$ and I reads [4]

$$\begin{aligned}
\hat{\mathbf{h}} &= \mathrm{E}[\mathbf{h}|\mathbf{y}] && (3)\\
&= \int_{\mathbb{C}^N} \mathbf{h} \frac{P(\mathbf{h})P(\mathbf{y}|\mathbf{h})}{P(\mathbf{y})} d\mathbf{h} && (4)\\
&= \int_{\mathbb{C}^N} \mathbf{h} \frac{P(\mathbf{h})P(\mathbf{y}|\mathbf{h})}{\left(\int_{\mathbb{C}^N} P(\mathbf{h})P(\mathbf{y}|\mathbf{h})d\mathbf{h}\right)} d\mathbf{h} && (5)\\
&= \lim_{\tilde{\mathbf{Q}} \to \mathbf{Q}} \int_{\mathbb{C}^N} \frac{\mathbf{h} \cdot e^{-\mathbf{h}^\mathsf{H}\tilde{\mathbf{Q}}^{-1}\mathbf{h}} \cdot e^{-\frac{1}{\sigma^2}(\mathbf{h}-\mathbf{h}')^\mathsf{H}\mathbf{P}(\mathbf{h}-\mathbf{h}')} d\mathbf{h}}{\left(\int_{\mathbb{C}^N} e^{-\mathbf{h}^\mathsf{H}\tilde{\mathbf{Q}}^{-1}\mathbf{h}} \cdot e^{-\frac{1}{\sigma^2}(\mathbf{h}-\mathbf{h}')^\mathsf{H}\mathbf{P}(\mathbf{h}-\mathbf{h}')} d\mathbf{h}\right)} && (6)
\end{aligned}$$

in which the limit is taken over a set of invertible matrices $\tilde{\mathbf{Q}}$ which tends to $\mathbf{Q}$ (which is by definition of rank $L < N$), and $\mathbf{P}$ is a projection matrix over the set of pilot frequency carriers, i.e. $P_{ij} = \delta_{i-j}\delta_{\phi(i)}$.

The product of the exponential terms in (6) can be written, after expansion and identification,

$$\mathbf{h}^\mathsf{H}\tilde{\mathbf{Q}}^{-1}\mathbf{h} + \frac{(\mathbf{h}-\mathbf{h}')^\mathsf{H}\mathbf{P}(\mathbf{h}-\mathbf{h}')}{\sigma^2} = (\mathbf{h}-\tilde{\mathbf{k}})^\mathsf{H}\tilde{\mathbf{M}}(\mathbf{h}-\tilde{\mathbf{k}}) + \tilde{C} \tag{7}$$

with

$$\begin{cases} \tilde{\mathbf{M}} &= \tilde{\mathbf{Q}}^{-1} + \frac{1}{\sigma^2}\mathbf{P} \\ \tilde{\mathbf{k}} &= \frac{1}{\sigma^2}\tilde{\mathbf{M}}^{-1}\mathbf{P}\mathbf{h}' \\ \tilde{C} &= \frac{1}{\sigma^2}\mathbf{h}'^\mathsf{H}\mathbf{P}\mathbf{h}' - \tilde{\mathbf{k}}^\mathsf{H}\tilde{\mathbf{M}}\tilde{\mathbf{k}} \end{cases} \tag{8}$$

This allows to isolate the dummy variable $\mathbf{h}$ in the integrals and leads then to compute the first order moment of a multivariate Gaussian distribution,

$$\hat{\mathbf{h}} = \lim_{\tilde{\mathbf{Q}} \to \mathbf{Q}} \int_{\mathbb{C}^N} \frac{\mathbf{h} \cdot e^{-(\mathbf{h}-\tilde{\mathbf{k}})^\mathsf{H}\tilde{\mathbf{M}}(\mathbf{h}-\tilde{\mathbf{k}})} d\mathbf{h}}{\int_{\mathbb{C}^N} e^{-(\mathbf{h}-\tilde{\mathbf{k}})^\mathsf{H}\tilde{\mathbf{M}}(\mathbf{h}-\tilde{\mathbf{k}})} d\mathbf{h}} = \lim_{\tilde{\mathbf{Q}} \to \mathbf{Q}} \tilde{\mathbf{k}} \tag{9}$$

As $\tilde{\mathbf{M}}^{-1} = (\tilde{\mathbf{Q}}^{-1} + \frac{1}{\sigma^2}\mathbf{P})^{-1} = (\mathbf{I}_N + \frac{1}{\sigma^2}\tilde{\mathbf{Q}}\mathbf{P})^{-1}\tilde{\mathbf{Q}}$,

$$\hat{\mathbf{h}} = \lim_{\tilde{\mathbf{Q}} \to \mathbf{Q}} (\sigma^2\mathbf{I}_N + \tilde{\mathbf{Q}}\mathbf{P})^{-1}\tilde{\mathbf{Q}}\mathbf{P}\mathbf{h}' = (\sigma^2\mathbf{I}_N + \mathbf{Q}\mathbf{P})^{-1}\mathbf{Q}\mathbf{P}\mathbf{h}' \tag{10}$$

which is classical LMMSE solution [4], when the covariance matrix of the channel is known to be (or assumed to be) $\mathbf{Q}$.

B. Unknown channel length

If L is only known to be in an interval $\{L_{\min}, \ldots, L_{\max}\}$, the maximum entropy principle assigns a uniform prior distribution for L [10]; otherwise one would add non desirable explicit information on a particular value for L. And then,

$$\begin{aligned}
\hat{\mathbf{h}} &= \mathrm{E}[\mathbf{h}|\mathbf{y}] && (11)\\
&= \int_{\mathbb{C}^N} \mathbf{h} \frac{\left(\sum_L P(\mathbf{h}|L)P(L)\right) P(\mathbf{y}|\mathbf{h})}{\left(\int_{\mathbb{C}^N} \left(\sum_L P(\mathbf{h}|L)P(L)\right) P(\mathbf{y}|\mathbf{h}) d\mathbf{h}\right)} d\mathbf{h} && (12)
\end{aligned}$$

which, from similar derivations as in Section III-A, is

$$\begin{aligned}
\hat{\mathbf{h}} = &\lim_{\substack{\tilde{\mathbf{Q}}_k \to \mathbf{Q}_k \\ L_{\min} \leq k \leq L_{\max}}} \sum_{L=L_{min}}^{L_{max}} \frac{1}{\det \tilde{\mathbf{Q}}_L} \\
&\times \int_{\mathbb{C}^N} \frac{\mathbf{h} \cdot e^{-\mathbf{h}^\mathsf{H}\tilde{\mathbf{Q}}_L^{-1}\mathbf{h}} \cdot e^{-\frac{1}{\sigma^2}(\mathbf{h}-\mathbf{h}')^\mathsf{H}\mathbf{P}(\mathbf{h}-\mathbf{h}')} d\mathbf{h}}{\sum_L \frac{1}{\det \tilde{\mathbf{Q}}_L} \int e^{-\mathbf{h}^\mathsf{H}\tilde{\mathbf{Q}}_L^{-1}\mathbf{h}} \cdot e^{-\frac{1}{\sigma^2}(\mathbf{h}-\mathbf{h}')^\mathsf{H}\mathbf{P}(\mathbf{h}-\mathbf{h}')} d\mathbf{h}}
\end{aligned} \tag{13}$$

where $\mathbf{Q}_k$ is the channel covariance matrix for a channel length $k \in \{L_{\min}, \ldots, L_{\max}\}$ and $\tilde{\mathbf{Q}}_k$ are taken in a set of invertible matrices in the neighborhood of $\mathbf{Q}_k$.

Using the same transformations as in (7) , we end up with

$$\hat{\mathbf{h}} = \lim_{\substack{\tilde{\mathbf{Q}}_k \to \mathbf{Q}_k \\ L_{\min} \leq k \leq L_{\max}}} \frac{\sum_{L=L_{\min}}^{L_{\max}} \det(\tilde{\mathbf{M}}^{(L)} \tilde{\mathbf{Q}}_L)^{-1} e^{-\tilde{C}^{(L)}} \tilde{\mathbf{k}}^{(L)}}{\sum_{L=L_{\min}}^{L_{\max}} \det(\tilde{\mathbf{M}}^{(L)} \tilde{\mathbf{Q}}_L)^{-1} e^{-\tilde{C}^{(L)}}} \tag{14}$$

with

$$\begin{cases} \tilde{\mathbf{M}}^{(L)} & = \tilde{\mathbf{Q}}_L^{-1} + \frac{1}{\sigma^2}\mathbf{P} \\ & = \tilde{\mathbf{Q}}_L^{-1}(\mathbf{I}_N + \frac{1}{\sigma^2}\tilde{\mathbf{Q}}_L\mathbf{P}) \\ \tilde{\mathbf{k}}^{(L)} & = \frac{1}{\sigma^2}(\mathbf{I}_N + \frac{1}{\sigma^2}\tilde{\mathbf{Q}}_L\mathbf{P})^{-1}\tilde{\mathbf{Q}}_L\mathbf{P}\mathbf{h}' \\ \tilde{C}^{(L)} & = \mathbf{h}'^{\mathsf{H}}\left(\mathbf{I}_N + \frac{1}{\sigma^2}\tilde{\mathbf{Q}}_L\mathbf{P}\right)^{-1}\frac{\mathbf{P}}{\sigma^2}\mathbf{h}' \end{cases} \tag{15}$$

The scalar $\det(\tilde{\mathbf{M}}^{(L)})$ can be further developed to obtain

$$\det(\tilde{\mathbf{M}}^{(L)}\tilde{\mathbf{Q}}_L) = \det(\mathbf{I}_N + \frac{1}{\sigma^2}\tilde{\mathbf{Q}}_L\mathbf{P}) \tag{16}$$

No inversion of $\tilde{\mathbf{Q}}_k$ matrices is then necessary so that the limit is well-defined and

$$\hat{\mathbf{h}} = \frac{\sum_{L=L_{\min}}^{L_{\max}} \det\left((\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{Q}_L\mathbf{P})^{-1}\right) e^{-C^{(L)}}\mathbf{k}^{(L)}}{\sum_{L=L_{\min}}^{L_{\max}} \det\left((\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{Q}_L\mathbf{P})^{-1}\right) e^{-C^{(L)}}} \tag{17}$$

in which $\mathbf{k}^{(L)}$ and $C^{(L)}$ are the limits of $\tilde{\mathbf{k}}^{(L)}$ and $\tilde{C}^{(L)}$ respectively,

$$\begin{cases} \mathbf{k}^{(L)} & = \left(\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{Q}_L\mathbf{P}^{\mathsf{H}}\mathbf{P}\right)^{-1}\frac{1}{\sigma^2}\mathbf{Q}_L\mathbf{P}\mathbf{h}' \\ C^{(L)} & = \mathbf{h}'^{\mathsf{H}}\left(\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{Q}_L\mathbf{P}\right)^{-1}\frac{\mathbf{P}}{\sigma^2}\mathbf{h}' \end{cases} \tag{18}$$

Since $C^{(L)}$ comprises the quadratic term $\mathbf{h}'^{\mathsf{H}}\mathbf{P}\mathbf{h}'$, the MMSE estimation of $\mathbf{h}$ is not linear in $\mathbf{h}'$. Therefore, the LMMSE estimate in the scenario when L is unknown is not an MMSE estimate. We also note that Equation (17) is a weighted sum of the individual LMMSE estimates for different hypothetical values of L. The weighting coefficients allow to enhance the estimates that rather fit the correct L hypothesis and to discard the others, based solely on the $\mathbf{h}'$ observation.

C. *Using time correlation*

When the channel coherence time, i.e. the time over which the channel realizations are correlated [8], is of the same order or larger than a few OFDM symbols, then past and future received data carry important information on the present channel realization. This information must be taken into account.

Classically, channel time correlation is described through Jakes' model [9]. For a Doppler spread f_d (proportional to the vehicular speed), the correlation figure is modeled as

$$\mathrm{E}[\nu_{t+T,p}\nu_{t,p}^*] = \frac{1}{L} \cdot J_0(2\pi f_d T) \tag{20}$$

in which J_0 is the zero-order Bessel function of the first kind. Note that Jakes' model (20) is based on the maximum entropy principle based on the only assumption that $\mathrm{E}[|\nu_{t,p}|^2] = \frac{1}{L}$.

Suppose now that only the present and previous past pilot symbols are considered by the terminal. Those correspond to two time instants t_1 and t_2, respectively. Consider first that L is known to the receiver. For notational simplicity, we denote $\mathbf{h}_k = \mathbf{h}_{t_k}$. The MMSE estimator for $\mathbf{h}_2$ under this state of knowledge is then

$$\hat{\mathbf{h}}_2 = \mathrm{E}[\mathbf{h}_2|\mathbf{h}_1'\mathbf{h}_2'] \tag{21}$$

$$= \int_{\mathbf{h}_2} \mathbf{h}_2 \frac{P(\mathbf{h}_2)P(\mathbf{h}_2'|\mathbf{h}_2)P(\mathbf{h}_1'|\mathbf{h}_2)}{P(\mathbf{h}_1'\mathbf{h}_2')} d\mathbf{h}_2 \tag{22}$$

$$= \int_{\mathbf{h}_2} \mathbf{h}_2 \frac{P(\mathbf{h}_2)P(\mathbf{h}_2'|\mathbf{h}_2)\int_{\mathbf{h}_1} P(\mathbf{h}_1'|\mathbf{h}_1)P(\mathbf{h}_1|\mathbf{h}_2)d\mathbf{h}_1}{P(\mathbf{h}_1'\mathbf{h}_2')} d\mathbf{h}_2 \tag{23}$$

in which we implicitly stated that $\mathbf{h}_1$ and $\mathbf{h}_2$ do not bring any information to, i.e. are independent of, $(\mathbf{h}_2'|\mathbf{h}_2)$ and $(\mathbf{h}_1'|\mathbf{h}_1)$ respectively.

Note that, apart from the new term $P(\mathbf{h}_1|\mathbf{h}_2)$, all probabilities to be derived here have already been evaluated in the previous section. Our knowledge on $(\mathbf{h}_1|\mathbf{h}_2)$ is limited to Equation (20). Burg's theorem [11] states then that the maximum entropy distribution for $(\boldsymbol{\nu}_1|\boldsymbol{\nu}_2)$ is an L-multivariate Gaussian distribution of mean $\lambda\boldsymbol{\nu}_2$ and variance $\frac{1}{L}(1-\lambda^2)\mathbf{I}_L$ with $\lambda = J_0(2\pi f_d[t_2 - t_1])$. Thanks to the same linearity argument as above, the distribution of $(\mathbf{h}_1|\mathbf{h}_2)$ is

$$P(\mathbf{h}_1|\mathbf{h}_2) = \lim_{\tilde{\mathbf{\Phi}} \to \mathbf{\Phi}} \frac{1}{\pi^N \det(\tilde{\mathbf{\Phi}})} e^{-(\mathbf{h}_1 - \lambda\mathbf{h}_2)^{\mathsf{H}}\tilde{\mathbf{\Phi}}^{-1}(\mathbf{h}_1 - \lambda\mathbf{h}_2)} \tag{24}$$

with $\mathbf{\Phi}(T) = (1-\lambda^2)\mathbf{Q}$.

Proceeding as previously, and denoting $\mathbf{P}_k$ the projection matrix of the pilots at time t_k and $\mathbf{M}_2$ such that

$$\mathbf{Q}\mathbf{M}_2 = \frac{\mathbf{I}_N}{1-\lambda^2} - \frac{\lambda^2}{1-\lambda^2}(\mathbf{I}_N + \frac{1-\lambda^2}{\sigma^2}\mathbf{Q}\mathbf{P}_1)^{-1} + \frac{\mathbf{Q}\mathbf{P}_2}{\sigma^2} \tag{25}$$

we finally have

$$\hat{\mathbf{h}}_2 = \mathbf{M}_2^{-1}\left(\frac{\mathbf{P}_2\mathbf{h}_2'}{\sigma^2} + (\mathbf{I}_N + \frac{1-\lambda^2}{\sigma^2}\mathbf{P}_1\mathbf{Q})^{-1}\frac{\lambda}{\sigma^2}\mathbf{P}_1\mathbf{h}_1'\right) \tag{26}$$

This formula stands only when the channel length L is known. Then, if L is only known to belong to an interval $\{L_{\min}, \ldots, L_{\max}\}$,

$$\hat{\mathbf{h}}_2 = \frac{\sum_{L_{\min}}^{L_{\max}} \det(\mathbf{A}_L\mathbf{B}_L)^{-1} e^{-C_2{}^{(L)}}\mathbf{k}_2^{(L)}}{\sum_{L=L_{\min}}^{L_{\max}} \det(\mathbf{A}_L\mathbf{B}_L)^{-1} e^{-C_2{}^{(L)}}} \tag{27}$$

with

$$\begin{cases} \mathbf{A}_L & = \mathbf{I}_N + \frac{1-\lambda^2}{\sigma^2}\mathbf{Q}_L\mathbf{P}_1 \\ \mathbf{B}_L & = \left(1 + \frac{\lambda^2}{1-\lambda^2}\right)\mathbf{I}_N - \frac{\lambda^2}{1-\lambda^2}\left(\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{Q}\mathbf{P}_1\right) \\ C_2 & = \frac{1}{\sigma^2}\mathbf{h}_2'^{\mathsf{H}}\mathbf{P}_2\mathbf{h}_2' - \mathbf{k}_2^{\mathsf{H}}\mathbf{M}_2\mathbf{k}_2 + \frac{1}{\sigma^2}\mathbf{h}_1'^{\mathsf{H}}\mathbf{P}_1\mathbf{h}_1' \\ & \quad - \mathbf{h}_1'^{\mathsf{H}}\left(\mathbf{I}_N + \frac{1-\lambda^2}{\sigma^2}\mathbf{Q}\mathbf{P}_1\right)^{-1}\frac{1}{\sigma^2}\mathbf{P}_1\mathbf{h}_1' \end{cases}$$

and $\mathbf{k}_2$ is given by the right hand side of Equation (26).

The final formulas (26) and (27) are interesting in the sense that they do not carry intuitive properties; if we were to find an *ad-hoc* technique to assess the relative importance

$$\hat{\mathbf{h}} = \left(\left(\left(1+\sum_{k=1}^{K}\frac{\lambda_k^2}{1-\lambda_k^2}\right)\mathbf{I}_N - \sum_{k=1}^{K}\frac{\lambda_k^2}{1-\lambda_k^2}\left(\mathbf{I}_N+\frac{1-\lambda_k^2}{\sigma^2}\mathbf{Q}\mathbf{P}_k\right)^{-1}\right)^{-1}\mathbf{Q}\left(\sum_{k=1}^{K}\lambda_k\left(\mathbf{I}_N+\frac{1-\lambda_k^2}{\sigma^2}\mathbf{P}_k\mathbf{Q}\right)^{-1}\frac{1}{\sigma^2}\mathbf{P}_k\mathbf{h}'_k\right)\tag{19}$$

of the prior information I, the pilot data $\mathbf{h}'_2$ and the past or future pilot data $\mathbf{h}'_1$, we would suggest a linear combination of those constraints. Our result is not linear in those constraints. However it carries the expected intuition in the limits,

- when $\lambda = 0$, past and present channels are uncorrelated so that no information carried by the past pilots is of any use. Equation (26) is consistent in this sense since it reduces to the LMMSE solution (10).
- $\lambda \to 1$ leads to the same Equation as (26) but with past and present pilots $\mathbf{h}'_1$ and $\mathbf{h}'_2$ gathered into a single pilot sequence $\mathbf{h}' = \mathbf{h}'_1 + \mathbf{h}'_2$ with projector $\mathbf{P} = \mathbf{P}_1 + \mathbf{P}_2$,

$$\hat{h}_2 = \frac{1}{\sigma^2}\left(\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{Q}\mathbf{P}\right)^{-1}\mathbf{Q}\mathbf{P}\mathbf{h}' \tag{28}$$

Note also that (26) is linear in $\mathbf{h}'_1$ and $\mathbf{h}'_2$, so the MMSE solution is also the LMMSE solution when L is known.

The previous derivation (26) is further generalized for a number K of pilot sequences $\mathbf{h}'_k$, $k \in \{1,\dots,K\}$, sent through channel $\mathbf{h}_k$ at time t_k, and a channel $\mathbf{h}$ at time t which satisfy, for $(i,k) \in \{1,\dots,L\} \times \{1,\dots,K\}$,

$$\mathrm{E}[\nu_{i,t}\nu^*_{i,t+t_k}] = \lambda_k / L \tag{29}$$

The maximum entropy principle and the same derivations as in previous calculus give in this situation the MMSE estimator $\hat{\mathbf{h}}$ of Equation (19).

D. Unknown correlation factor λ

Similarly to what we did previously with the possibly unknown parameter L, we can equally integrate out the parameter λ from our formulas, which is in practice difficult to estimate. For $K = 2$, in Equation (23), $P(\mathbf{h}_2)$ equals $\int_\lambda P(\mathbf{h}_2|\lambda)P(\lambda)d\lambda$ and $\hat{\mathbf{h}}_2$ then reads,

$$\hat{\mathbf{h}}_2 = \mathrm{E}[\mathbf{h}_2|\mathbf{h}'_1,\mathbf{h}'_2] \tag{30}$$

$$= \int_\lambda P(\lambda)\mathrm{E}[\mathbf{h}_2|\mathbf{h}'_1,\mathbf{h}'_2,\lambda]d\lambda \tag{31}$$

in which $P(\lambda)$ is the probability assigned to $(\lambda|I)$. This last integral is however rather involved[3]. It can then be approximated by the discrete summation,

$$\hat{\mathbf{h}}_2 \simeq \sum_{\lambda\in\mathcal{S}(\lambda_{\min},\lambda_{\max})} P(\lambda)\mathrm{E}[\mathbf{h}_2|\mathbf{h}'_1,\mathbf{h}'_2,\lambda] \tag{32}$$

if λ is known to belong to some discrete set $\mathcal{S}(\lambda_{\min},\lambda_{\max})$ such that $[\lambda_{\min},\lambda_{\max}] \subset \mathcal{S}(\lambda_{\min},\lambda_{\max})$.

Remark 2: In previous sections, we considered $\mathrm{E}[\mathbf{n}\mathbf{n}^{\mathsf{H}}] = \sigma^2\mathbf{I}_N$. In the presence of interference, the noise correlation matrix $\mathbf{C_n} = \mathrm{E}[\mathbf{n}\mathbf{n}^{\mathsf{H}}]$ is colored. If the information about $\mathbf{C_n}$ is considered, all the previous equations must then be updated by replacing all terms $\frac{1}{\sigma^2}\mathbf{P}$ by $\mathbf{P}^{\mathsf{H}}\mathbf{C_n}\mathbf{P}$.

[3]note that in the previous sections, we implicitly took a Dirac in the known value for λ as $P(\lambda)$.

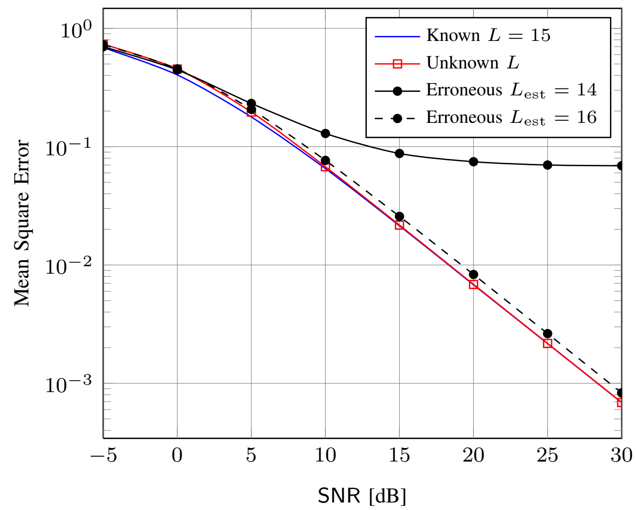

Fig. 2. Mean square error of channel estimate when $L = 15$ is known, known to be in $\{1,\dots,32\}$, or erroneously estimated to $L_{\text{est}} \in \{14, 16\}$, SNR = 20 dB, $N = 128$.

IV. Simulations and Results

In this section, we provide Monte Carlo simulations of some of the previously derived algorithms. We consider an OFDM system provided with $N = 128$ subcarriers, and a CP size $N_{\text{CP}} = 32$. In a first simulation, we take a channel of average length $L = 15$. The pilot symbols are spaced every 6 subcarriers in frequency. This situation is one of the configurations of the 3GPP-LTE standard [1], which is depicted in Figure 1. Figure 2 provides the mean square error, averaged over 10,000 channel realizations, of a channel estimator induced by Equation (17). Each channel realization is drawn from an L-multipath model with i.i.d. Gaussian entries. The channel length L is either known to the receiver (we use then Equation (9)), known to be such that $L \in \{1,\dots,32\}$ (we use then Equation (17)) or erroneously estimated to $L_{\text{est}} = 14$ (we use then Equation (9) with $L_{\text{est}} = 14$). Interestingly, while erroneous approximations for L may lead to dramatic loss in performance in the high SNR region, almost no performance decay is observed when L is either known or unknown. This means that the Bayesian framework, when inferring on $(\mathbf{h}|\mathbf{y})$, indirectly performs inference on $(L|\mathbf{y})$. Indeed,

$$P(L|\mathbf{y}) = \frac{P(L)P(\mathbf{y}|L)}{P(\mathbf{y})} = \frac{P(L)}{P(\mathbf{y})}\int_{\mathbf{h}} P(\mathbf{y}|\mathbf{h})P(\mathbf{h}|L)\mathrm{d}\mathbf{h} \tag{33}$$

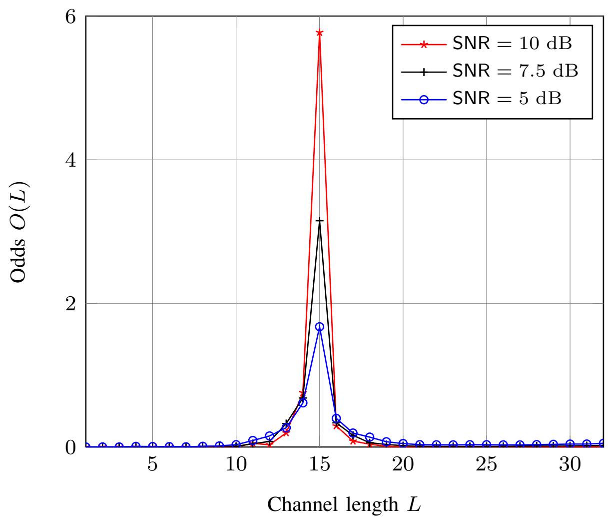

Fig. 3. Inference on channel length $O(L)$ for different SNR, $N = 128$, $L = 15$ (true length), $L_{\min} = 32$, $L_{\max} = 32$

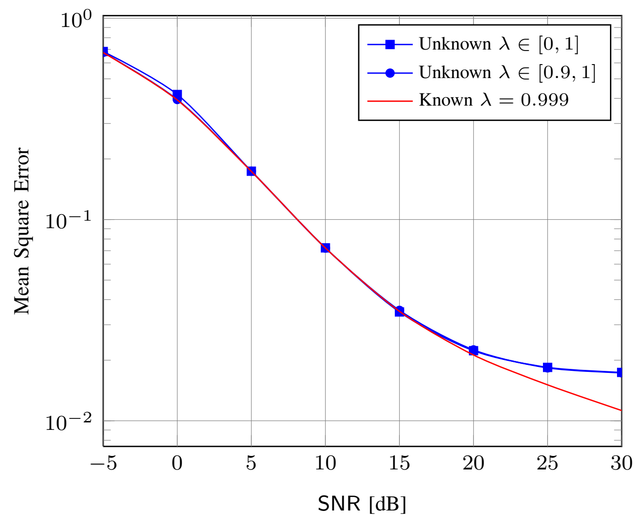

Fig. 4. Channel estimate mean square error, with unknown time correlation $\lambda = 0.999$, $N = 128$, $L = 25$

in which the integral is the same as in Section III-A and $P(L)$ is the uniform prior (maximum entropic) distribution for L.

Figure 3 provides the *odds* $O(L|\mathbf{y})$ [10] associated to $(L|\mathbf{y})$, averaged on $10,000$ channel realizations, for $N = 128$, $L = 15$, $L_{\min} = 1$, $L_{\max} = 32$ and different values of SNR. The odds function $O(L)$ is defined as

$$O(L|\mathbf{y}) = \frac{P(L|\mathbf{y})}{\sum_{l \neq L} P(l|\mathbf{y})} \tag{34}$$

As predicted, $O(L = 15) > O(L \neq 15)$ and this behaviour is enhanced for high SNR. Therefore, the posterior distribution $P(L|\mathbf{y})$ almost discards all hypothesis but $L = 15$.

Note also that simulations with 3GPP-standardized multi-path channels (instead of Gaussian i.i.d. channel models) were performed, which did not show dramatic performance loss due to the erroneous maximum entropy assumption on the channel. The complete results are provided in [13].

In Figure 4, channel estimation is performed using two pilot sequences correlated in time. The receiver either knows the exact correlation coefficient $\lambda = 0.999$ or only knows that $\lambda \in [0, 1]$ or $\lambda \in [0.9, 1]$. The channel length $L = 25$ is known (performance gain in this case is only significant for long channels). We observe that a limited knowledge on λ does not strongly impact the performance of the estimation in the low-to-medium SNR region. When the receiver knows $\lambda > 0.9$, the mean square error is similar to that when the receiver knows only $\lambda > 0$. This suggests again that the Bayesian machinery is able to implicitly infer on λ, which is more efficient than using empirical values for λ.

V. Conclusion

In this work, a Bayesian framework for OFDM channel estimation is proposed. Under different levels of knowledge on the relevant system parameters at the receiver, MMSE estimates are derived that re-demonstrate known classical solutions while new formulas are also proposed. The whole work can be synthesized in a simple Bayesian methodology that allows to optimally use the information at the receiver. Simulations are also provided that confirm the performance superiority of the derived formulas over classical methods.

VI. Acknowledgement

This work was partially supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless Communications NEWCOM++.

References

[1] S. Sesia, I. Toufik and M. Baker, "LTE, The UMTS Long Term Evolution: From Theory to Practice", Wiley & Sons, 2009.

[2] J.A. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," IEEE Commun. Mag., vol. 28, pp. 5-14, 1990.

[3] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems", IEEE Trans. on Broadcasting, vol. 48, pp. 223-229, 2002.

[4] S. M. Kay, "Fundamentals of statistical signal processing", PTR Prentice-Hall Englewood Cliffs, NJ, 1993.

[5] M. Morelli and U. Mengali, "A comparison of pilot-aided channel estimation methods for OFDM systems", IEEE Trans. on Signal Processing, vol. 49, no. 12, pp. 3065-3073, 2001.

[6] S. Omar, A. Ancora and D. Slock, "Performance analysis of general pilot-aided linear channel estimation in LTE OFDMA systems with apllication to simplified MMSE schemes", PIMRC'08 Conference on Wireless Communications, 2008.

[7] E. T. Jaynes, "Where do we stand on maximum entropy?", *The Maximum Entropy Formalism*, M. I. T. Press, Cambridge, MA, 1979.

[8] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," IEEE Trans. on Inf. Theory, vol. 44, no. 6, pp. 2619-2692, 1998.

[9] P. Dent, G. E. Bottomley and T. Croft, "Jakes fading model revisited", Electonics Letters, vol. 29, no. 13, 1993.

[10] E. T. Jaynes, "Probability Theory: The Logic of Science", Cambridge University Press, 2003.

[11] T. M. Cover and J. A. Thomas, "Elements of Information Theory", Wiley New York, pp. 274-277, 1991.

[12] M. Debbah and R. Müller, "MIMO channel modeling and the principle of maximum entropy," IEEE Trans. on IT, vol. 51, no. 5, 2005.

[13] R. Couillet, M. Debbah, "A maximum entropy approach to OFDM channel estimation", arxiv preprint http://arxiv.org/pdf/0811.0778