Improved Iterative Detection and Achieved Throughputs of Multiband OFDM Systems Under Imperfect Channel Estimation¹

S. M. Sajad Sadough *student member, IEEE*, and Pierre Duhamel *Fellow, IEEE*

Abstract

Assuming imperfect channel estimation, we propose an improved detector for ultra wideband (UWB) systems employing multiband orthogonal frequency-division multiplexing (MB-OFDM). By adopting a Bayesian approach involving the statistics of the channel estimation errors, we formulate an improved maximum-likelihood (ML) detection metric taking into account the characteristics of the channel estimates. As a first step, we propose a modified iterative detector based on maximum *a posteriori* which reduces the impact of channel uncertainty on the decoder performance, by an appropriate use of this metric. The results are compared to those obtained by using a detector based on a mismatched ML metric, which uses the channel estimate as if it was the perfect channel. In a second step, we calculate the information rates achieved by both the improved and mismatched ML detectors, in terms of achievable outage rates. These outage rates are compared to those provided by a theoretical (not practical) decoder defined as the best decoder in the presence of channel estimation errors. Numerical results over both uncorrelated Rayleigh fading channels and realistic UWB channels show that the improved detector provides significant gains in terms of bit error rate and achievable outage rates, compared to the classically-used mismatched approach, without any increase in the receiver complexity.

I. INTRODUCTION

Multiband orthogonal frequency-division multiplexing (MB-OFDM) is a spectrally efficient technique proposed for high data-rate, short-range ultra wideband (UWB) applications operating in the range from 3.1 to 10.6 GHz [1]. In this scheme, the available 7.5 GHz of UWB spectrum is divided into several subbands of at least 500 MHz bandwidth, with a conventional OFDM modulation within each subband.

¹The material in this paper was published in part at the *40-th Asilomar Conference on Signals, Systems, and Computers*.

Several standard proposals such as IEEE802.15.3a that proposed MB-OFDM, employ bit-interleaving combined with convolutional channel coding. This scheme, referred in the literature as bit-interleaved coded modulation (BICM) [2], can provide high diversity order for transmission over Rayleigh fading channels with coherent detection and *perfect* knowledge of the channel.

It is well known that reliable coherent data detection is not possible unless an accurate channel state information is available at the receiver. A typical scenario for wireless communication systems assumes the channel changes so slowly that it can be considered time invariant during the transmission of an entire frame. In such situations, pilot symbol assisted modulation (PSAM) [3] has been shown to be an effective solution for obtaining channel state information at the receiver (CSIR). It consists of sending some known training (also called pilot) symbols, based on which the receiver estimates the channel before proceeding to the detection of data symbols. However, in practical systems, due to the finite number of pilot symbols and to noise, the receiver can only obtain an *imperfect* (and possibly very poor) estimate of the channel.

The performance of BICM over fading channels was studied, for instance, in [2], [4], [5] under perfect channel knowledge at the receiver. A rich literature exists on the impact of imperfect channel estimation. In [6], the authors showed that in low signal-to-noise ratio (SNR) situations, the poor quality of channel estimates may prevent iterative decoding of BICM to be used at the decoder. In [7], Garg *et al.* considered a training-based MIMO system and showed that for compensating the performance degradation due to imperfect channel estimation, the number of receive antennas should be increased. Obviously, this may not be always possible in mobile applications.

In order to deal with imperfect channel estimation, the classical technique, referred to as *mismatched* detection, consists in replacing the exact channel by its estimate in the receiver metric. It is shown, for instance, in [8], [9], [10] that this scheme greatly degrades the detection performance in the presence of channel estimation errors for a multi-carrier system. Furthermore in [11], authors show that under imperfect channel estimation, the rates achieved by the mismatched detector are significantly lower than the channel capacity.

As an alternative to the aforementioned mismatched approach, Tarokh *et al.* in [12] and recently Taricco and Biglieri in [13], proposed an *improved* maximum likelihood (ML) detection metric that mitigates the impact of imperfect CSIR and applied it to a space-time coded MIMO system. A similar investigation was carried out in [14] for trellis-coded modulation in scalar

channels. Very recently, in [15], the authors derived the maximum possible outage capacity induced by the channel estimation errors, in the context of Rayleigh channels and a theoretical decoder. This was further developed in [16], [17] to address the MIMO situation.

A practical concern for the design of a communication system is to guarantee a prescribed target rate with small error probability, irrespective of the accuracy achieved during channel estimation. However, adopting a mismatched approach in the described scenario raises two important questions : **i)** What are the maximal achievable information rates of mismatched decoding and how close does it perform with respect to the rates provided by the best possible decoder in the presence of channel estimation errors ? **ii)** What type of practical decoder can improve the overall system performance under imperfect CSIR ?

The objective of this paper is to address the above questions for MB-OFDM channels, estimated by a small number of training sequences. To this end, we propose a Bayesian framework based on the *a posteriori* probability density function (pdf) of the perfect channel conditioned on its estimate. This general framework enables us to formulate any detection problem by considering the average over the channel estimations errors, of the detector's cost function, that would be used if the channel was perfectly known. Using this approach, we can recover the ML metric of [13], as a special case of our general framework. This ML metric is first used for deriving an improved maximum *a posteriori* (MAP) iterative detector, a technique which is known to be very efficient when channel coding and bit interleaving are jointly used. By modifying properly the soft-values at the output of the MAP detector, we reduce the impact of channel uncertainty on the decoder performance. In a second step, by using some tools from information theory, we provide the expression of the achievable information rates associated to the improved and mismatched ML metric. Then, we characterize the limits of reliable information rates achieved by the mismatched and improved detectors in terms of outage rates, which is more appropriate for quasi-static channels [18].

Our results may serve in the evaluation of the trade-off between the required quality of service (in terms of BER and achieved throughputs) and system parameters (e.g., training power, transmission power, period of training, outage probability, etc.) in the presence of channel estimation errors.

The remainder of this paper is organized as follows. Section II describes the model for MB-OFDM system and our main assumptions concerning data transmission and channel estimation.

Section III introduces a general Bayesian framework for improved detection in the presence of channel estimation errors and then formulates the improved ML detection metric. In section IV, we propose the improved MAP detector in the case of imperfect channel estimation. In section V, we calculate the achievable outage rates of receivers based on the improved and mismatched ML metric. Section VI illustrates via simulations, the performance of the proposed receiver over both realistic UWB and Rayleigh fading channel environments. Finally, section VII concludes the paper.

Notational conventions are as follows. $\mathbb{I}_N$ represents an $(N \times N)$ identity matrix; $\mathcal{CN}$ denotes a complex Gaussian distribution; $\mathbb{E}_{\mathbf{x}}[\cdot]$ refers to expectation with respect to the random vector $\mathbf{x}$; $\det\{\cdot\}$, $\text{tr}(\cdot)$, $|\cdot|$ and $\|\cdot\|_F$ denote matrix determinant, matrix trace, absolute value and Frobenius norm, respectively; $(\cdot)^T$, $(\cdot)^\dagger$ and $(\cdot)^*$ denote vector transpose, Hermitian transpose and conjugation, respectively.

II. TRANSMISSION MODEL AND CHANNEL ESTIMATION

A. Multiband OFDM System

We consider the MB-OFDM scenario proposed in the IEEE 802.15.3a WPAN proposal [19]. In this scheme, the frequency spectrum between 3.1 to 10.6 GHz is divided into several non-overlapping subbands, each one occupying approximately a 500 MHz bandwidth [1]. As depicted in Fig. 1, the information is transmitted using OFDM modulation over one of the subbands in a particular time-slot. The binary sequence is encoded by a non-recursive non-systematic convolutional (NRNSC) code, before being interleaved by a pseudo-random interleaver. The interleaved bits are gathered in subsequences of B bits $d_k^1, \dots, d_k^B$ and mapped to complex M_c -QAM ($M_c = 2^B$) symbols s_k . At each time-slot, a time-frequency code (TFC) selects the center frequency of the subband over which the OFDM symbol is sent. The TFC is used not only to provide frequency diversity, but also to distinguish between multiple users. We assume single-user OFDM transmission with M subcarriers per subband, through a frequency selective multipath fading channel, described in discrete-time baseband equivalent form by the taps $\{h_l\}_{l=0}^{L-1}$. At the receiver, after removing the cyclic prefix (CP) and performing fast Fourier transform (FFT), the received OFDM symbol over a given subband can be written as

$$\mathbf{y} = \mathbf{H}_d \mathbf{s} + \mathbf{z}, \quad (1)$$

where the $(M \times 1)$ vectors $\mathbf{y}$ and $\mathbf{s}$ respectively denote the received and transmitted symbols; the noise vector $\mathbf{z}$ is assumed to be zero-mean circularly symmetric complex Gaussian (ZMCSCG) with distribution $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \sigma_z^2 \mathbf{I}_M)$; and $\mathbf{H}_d = \text{diag}(\mathbf{H})$ is the $(M \times M)$ diagonal channel matrix with diagonal elements given by vector $\mathbf{H} = [H_0, \dots, H_{M-1}]^T$, where $H_k = \sum_{l=0}^{L-1} h_l e^{-j2\pi kl/M}$. The channel coefficients are assumed to be time invariant over the transmission of one frame and change to new independent values at each frame.

For convenience, we also define the diagonal matrix $\mathbf{S}_d = \text{diag}(\mathbf{s})$ with $\mathbf{s} = [s_0, \dots, s_{M-1}]^T$. In the following, when there is no confusion, we use $\mathbf{H}$ instead of $\mathbf{H}_d$.

B. Pilot Based Channel Estimation

Consider the estimation of the channel vector $\mathbf{H}$ with N training symbols $\mathbf{s}_{T,i}$, ($i = 1, \dots, N$). According to the observation model (1), during a given channel training interval, we receive

$$\mathbf{Y}_T = \mathbf{H}_d \mathbf{S}_T + \mathbf{Z}_T \quad (2)$$

where each column of the $(M \times N)$ matrix $\mathbf{S}_T = [\mathbf{s}_{T,1} | \dots | \mathbf{s}_{T,N}]$ contains one OFDM pilot symbol. The entries of the noise matrix $\mathbf{Z}_T$ have the same distribution as those of $\mathbf{z}$. Moreover, the definition of $\mathbf{Y}_T$ and $\mathbf{Z}_T$ follow that of $\mathbf{S}_T$. The average energy E_T of the training symbols is

$$E_T \triangleq \frac{1}{NM} \|\mathbf{S}_T\|_F^2 = \frac{1}{NM} \text{tr}(\mathbf{S}_T \mathbf{S}_T^\dagger). \quad (3)$$

The least-squares estimate of $\mathbf{H}_d$ is obtained by minimizing $\|\mathbf{Y}_T - \mathbf{H}_d \mathbf{S}_T\|_F^2$ with respect to $\mathbf{H}_d$ which coincides here with the ML estimate. This yields

$$\begin{aligned} \hat{\mathbf{H}}_d^{\text{ML}} &= \mathbf{Y}_T \mathbf{S}_T^\dagger (\mathbf{S}_T \mathbf{S}_T^\dagger)^{-1} \\ &= \mathbf{H}_d + \mathcal{E}_d \end{aligned} \quad (4)$$

where $\mathcal{E}_d = \mathbf{Z}_T \mathbf{S}_T^\dagger (\mathbf{S}_T \mathbf{S}_T^\dagger)^{-1}$ is the channel estimation error term. We assume that the training matrix $\mathbf{S}_T$ has orthogonal rows such that $\mathbf{S}_T \mathbf{S}_T^\dagger = NE_T \mathbb{I}_M$, so that the estimation error is uncorrelated. Under these assumptions, equation (4) can be rewritten as

$$\hat{\mathbf{H}}^{\text{ML}} = \mathbf{H} + \mathcal{E} \quad (5)$$

where $\mathcal{E}$ is the diagonal of the error matrix $\mathcal{E}_d$. The covariance matrix of $\mathcal{E}$ is

$$\Sigma_{\mathcal{E}} = \mathbb{E}[\mathcal{E} \mathcal{E}^\dagger] = \sigma_{\mathcal{E}}^2 \mathbb{I}_M, \quad \text{where } \sigma_{\mathcal{E}}^2 = \frac{\sigma_z^2}{NE_T}. \quad (6)$$

From the latter equation it is clear that the choice of an orthogonal training sequence matrix leads to a white error vector, i.e., the entries of $\mathcal{E}$ are independent identically distributed (i.i.d.) ZMCSCG random variables with variance $\sigma_{\mathcal{E}}^2$.

Thus, from (5), the conditional pdf of $\hat{\mathbf{H}}^{\text{ML}}$ given $\mathbf{H}$ reads

$$p(\hat{\mathbf{H}}^{\text{ML}}|\mathbf{H}) = \mathcal{CN}(\mathbf{H}, \Sigma_{\mathcal{E}}). \quad (7)$$

Consider an uncorrelated i.i.d. Rayleigh channel with a *prior* distribution $\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \Sigma_H)$, where $\Sigma_H = \sigma_h^2 \mathbb{I}_M$. By using the latter pdf and the pdf of $(\hat{\mathbf{H}}^{\text{ML}}|\mathbf{H})$ from (7), we can derive the *posterior* distribution of the perfect channel, conditioned on its ML estimate as follows (see Appendix I.A).

$$p(\mathbf{H}|\hat{\mathbf{H}}^{\text{ML}}) = \mathcal{CN}(\Sigma_{\Delta}\hat{\mathbf{H}}^{\text{ML}}, \Sigma_{\Delta}\Sigma_{\mathcal{E}}), \quad (8)$$

where

$$\Sigma_{\Delta} = \Sigma_H(\Sigma_{\mathcal{E}} + \Sigma_H)^{-1} = \delta \mathbb{I}_M \quad \text{and} \quad \delta = \frac{\sigma_h^2}{(\sigma_h^2 + \sigma_{\mathcal{E}}^2)}. \quad (9)$$

The availability of the estimation error distribution constitutes an interesting feature of pilot assisted channel estimation which is exploited in the next section, resulting in an improved detector under imperfect channel estimation.

For the sake of simplicity, we will not specify hereafter the superscript ML for $\hat{\mathbf{H}}$.

III. DETECTOR DESIGN IN THE PRESENCE OF CHANNEL ESTIMATION ERRORS

We now provide the formulation of a detection rule that takes into account the available imperfect CSIR and refer to it as the *improved* detector. To this end, we consider the model (1) and denote by $f(\mathbf{y}, \mathbf{s}, \mathbf{H})$ the quantity (cost function) that would let us to decide in favor of a particular $\mathbf{s}$ at the receiver if the channel was *perfectly* known. Note that depending on the detection criteria, the quantity $f(\mathbf{y}, \mathbf{s}, \mathbf{H})$ can be the posterior pdf $p(\mathbf{s}|\mathbf{y}, \mathbf{H})$, the logarithm of the likelihood function $W(\mathbf{y}|\mathbf{H}, \mathbf{s})$, the mean square error, etc. Under a pilot-based channel estimation characterized by the posterior pdf of the channel (8), we propose a detector based on the minimization of a new cost function defined as

$$\tilde{f}(\mathbf{y}, \mathbf{s}, \hat{\mathbf{H}}) = \int_{\mathbf{H}} f(\mathbf{y}, \mathbf{s}, \mathbf{H}) p(\mathbf{H}|\hat{\mathbf{H}}) d\mathbf{H} = \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} [f(\mathbf{y}, \mathbf{s}, \mathbf{H})|\hat{\mathbf{H}}] \quad (10)$$

where we have averaged the cost function f over all realizations of the unknown channel $\mathbf{H}$ conditioned on its available estimate $\hat{\mathbf{H}}$ by using the distribution (8). Note that this approach

is an alternative to the sub-optimal *mismatched* detector, which is based on the minimization of the cost function $f(\mathbf{y}, \mathbf{s}, \hat{\mathbf{H}})$. The latter is obtained by using the estimated channel $\hat{\mathbf{H}}$ in the same metric that would be used if the channel was perfectly known, i.e., $f(\mathbf{y}, \mathbf{s}, \mathbf{H})$. The approach in (10) differs from the mismatched detection on the conditional expectation $\mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}}[\cdot]$ which provides a robust design by averaging the cost function $f(\mathbf{y}, \mathbf{s}, \mathbf{H})$ over all (true) channel realizations which could correspond to the available estimate.

Consider the problem of detecting symbol $\mathbf{s}$ from the observation model (1) in the ML sense, i.e., so as to maximize the likelihood function $W(\mathbf{y}|\mathbf{H}, \mathbf{s})$. The connection of the considered detection problem with the MAP detection of MB-OFDM is discussed in the next section. It is well known that under *perfect* channel knowledge and i.i.d. Gaussian noise, detecting $\mathbf{s}$ by maximizing the likelihood $W(\mathbf{y}|\mathbf{s}, \mathbf{H})$ is equivalent to minimizing the Euclidean distance $\mathcal{D}_{\text{ML}}$ as

$$\hat{\mathbf{s}}_{\text{ML}}(\mathbf{H}) = \arg \min_{\mathbf{s} \in \mathbb{C}^{M \times 1}} \{ \mathcal{D}_{\text{ML}}(\mathbf{s}, \mathbf{y}, \mathbf{H}) \}, \quad (11)$$

with $\mathcal{D}_{\text{ML}}(\mathbf{s}, \mathbf{y}, \mathbf{H}) \triangleq -\log W(\mathbf{y}|\mathbf{H}, \mathbf{s}) \propto \|\mathbf{y} - \mathbf{H}_d \mathbf{s}\|_F^2$, where $\propto$ means “is proportional to” and $\mathbb{C}$ denotes the set of complex numbers.

The detection rule (11) requires the knowledge of the *perfect* channel vector $\mathbf{H}$. The sub-optimal mismatched ML detector consists in replacing the exact channel by its estimate $\hat{\mathbf{H}}$ in the receiver metric as

$$\hat{\mathbf{s}}_{\text{MM}}(\hat{\mathbf{H}}) = \arg \min_{\mathbf{s} \in \mathbb{C}^{M \times 1}} \{ \mathcal{D}_{\text{MM}}(\mathbf{s}, \mathbf{y}, \hat{\mathbf{H}}) \} = \arg \min_{\mathbf{s} \in \mathbb{C}^{M \times 1}} \{ \|\mathbf{y} - \hat{\mathbf{H}}_d \mathbf{s}\|_F^2 \}, \quad (12)$$

where

$$\mathcal{D}_{\text{MM}}(\mathbf{s}, \mathbf{y}, \mathbf{H}) \triangleq \mathcal{D}_{\text{ML}}(\mathbf{s}, \mathbf{y}, \mathbf{H}) \Big|_{\mathbf{H}=\hat{\mathbf{H}}}. \quad (13)$$

Obviously, the sub-optimality of this detection technique is due to the mismatch introduced by the channel estimation errors; while the decision metric is derived from the likelihood function $W(\mathbf{y}|\mathbf{H}, \mathbf{s})$ conditioned on the perfect channel $\mathbf{H}$, the receiver uses an estimate $\hat{\mathbf{H}}$ different from $\mathbf{H}$ in the detection process.

As an alternative to this mismatched detection, an improved ML detection metric is proposed in [13], [12]. Actually, by using the formulation in (10) and the pdf (8), the latter metric can be derived as a special case of our general framework, as follows.

$$\widetilde{W}(\mathbf{y}|\hat{\mathbf{H}}, \mathbf{s}) = \int_{\mathbf{H} \in \mathbb{C}^{M \times 1}} W(\mathbf{y}|\mathbf{H}, \mathbf{s}) p(\mathbf{H}|\hat{\mathbf{H}}) \, d\mathbf{H} = \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}} \left[W(\mathbf{y}|\mathbf{H}, \mathbf{s}) \Big| \hat{\mathbf{H}} \right], \quad (14)$$

where $p(\mathbf{H}|\hat{\mathbf{H}})$ is the channel posterior distribution of equation (8).

By applying Theorem 1.2 of Appendix I.B, the averaged likelihood in (14) is shown to be complex Gaussian distributed vector given by

$$\widetilde{W}(\mathbf{y}|\hat{\mathbf{H}}, \mathbf{s}) = \mathcal{CN}(\mathbf{m}_{\mathcal{M}}, \Sigma_{\mathcal{M}}), \quad (15)$$

where $\mathbf{m}_{\mathcal{M}} = \delta \hat{\mathbf{H}}_d \mathbf{s}$, and $\Sigma_{\mathcal{M}} = \Sigma_z + \delta \sigma_{\mathcal{E}}^2 \mathbf{S}_d \mathbf{S}_d^\dagger$.

Finally, the estimate of the symbol $\mathbf{s}$ is

$$\hat{\mathbf{s}}_{\mathcal{M}}(\hat{\mathbf{H}}) = \arg \min_{\mathbf{s} \in \mathbb{C}^{M \times 1}} \{ \mathcal{D}_{\mathcal{M}}(\mathbf{s}, \mathbf{y}, \hat{\mathbf{H}}) \}, \quad (16)$$

where

$$\mathcal{D}_{\mathcal{M}}(\mathbf{s}, \mathbf{y}, \hat{\mathbf{H}}) \triangleq -\log \widetilde{W}(\mathbf{y}|\mathbf{s}, \hat{\mathbf{H}}) = \sum_{k=0}^{M-1} \log \pi(\sigma_z^2 + \delta \sigma_{\mathcal{E}}^2 |s_k|^2) + \frac{|y_k - \delta \hat{H}_k s_k|^2}{\sigma_z^2 + \delta \sigma_{\mathcal{E}}^2 |s_k|^2} \quad (17)$$

is referred to as the *improved* ML decision metric under imperfect CSIR.

Note that when CSIR tends to the exact value, which is obtained when the number of pilot symbols tends to infinity, we have $\delta \rightarrow 1$, $\sigma_{\mathcal{E}}^2 \rightarrow 0$ and thus

$$\lim_{N \rightarrow \infty} \frac{\mathcal{D}_{\mathcal{M}}(\mathbf{s}, \mathbf{y}, \hat{\mathbf{H}})}{\mathcal{D}_{\text{MM}}(\mathbf{s}, \mathbf{y}, \hat{\mathbf{H}})} = 1. \quad (18)$$

Consequently, the improved metric of (17) tends to the mismatched metric, when the estimation errors tend to zero.

IV. ITERATIVE DETECTION OF MB-OFDM UNDER IMPERFECT CHANNEL ESTIMATION

At the receiver, we perform MAP symbol detection and channel decoding in an iterative manner. As shown in Fig. 2, the receiver mainly consists in the combination of two sub-blocks that exchange soft informations with each other. The first sub-block, referred to as soft detector (also called demapper), produces bit metrics (probabilities and so called “extrinsic informations”) from the input symbols and the second one is a soft-input soft-output (SISO) decoder. Each sub-block can take advantage of the quantities provided by the other sub-block as an *a priori* information. Here, SISO decoding is performed using the well known forward-backward algorithm [20]. In the following, we make use of the improved ML metric derived in the previous section to modify the MAP detector part for the case of imperfect CSIR.

Let $d_k^{j,m}$ be the m -th ($m = 1, \dots, B$) coded and interleaved bit corresponding to the constellation symbol s_k transmitted at the j -th time slot over the k -th subcarrier. We denote by $L(d_k^{j,m})$ the

coded log-likelihood ratio (LLR) of the bit $d_k^{j,m}$ at the output of the detector. Conditioned on the imperfect CSIR $\hat{H}_k$, $L(d_k^{j,m})$ is given by

$$L(d_k^{j,m}) = \log \frac{P_{\text{dem}}(d_k^{j,m} = 1 | y_k, \hat{H}_k)}{P_{\text{dem}}(d_k^{j,m} = 0 | y_k, \hat{H}_k)}, \quad (19)$$

where $P_{\text{dem}}(d_k^{j,m} | y_k, \hat{H}_k)$ is the probability of transmission of $d_k^{j,m}$ at the detector output. Let $\underline{\mathcal{S}}$ be the set of all possibly-transmitted symbol s_k . We partition $\underline{\mathcal{S}}$ into two sets $\underline{\mathcal{S}}_0^m$ and $\underline{\mathcal{S}}_1^m$, for which the m -th bit of s_k equals “0” or “1”, respectively. We have

$$L(d_k^{j,m}) = \log \frac{\sum_{s_k \in \underline{\mathcal{S}}_1^m} e^{-\mathcal{D}_{\mathcal{M}}(s_k, y_k, \hat{H}_k)} \prod_{\substack{n=1 \\ n \neq m}}^B P_{\text{dec}}^1(d_k^{j,n})}{\sum_{s_k \in \underline{\mathcal{S}}_0^m} e^{-\mathcal{D}_{\mathcal{M}}(s_k, y_k, \hat{H}_k)} \prod_{\substack{n=1 \\ n \neq m}}^B P_{\text{dec}}^0(d_k^{j,n})}, \quad (20)$$

where $P_{\text{dec}}^1(d_k^{j,n})$ and $P_{\text{dec}}^0(d_k^{j,n})$ are *prior* probabilities coming from the SISO decoder.

Note that using the metric $\mathcal{D}_{\mathcal{M}}(s_k, y_k, \hat{H}_k)$ for the evaluation of the LLRs in (20) is an alternative to using the mismatched ML metric $\mathcal{D}_{\text{MM}}(s_k, y_k, \hat{H}_k)$ which replaces at each iteration, the exact channel H_k by its estimate $\hat{H}_k$ in $\mathcal{D}_{\text{ML}}(s_k, y_k, H_k)$. By doing so, the LLRs are adapted to the imperfect channel knowledge available at the receiver and consequently the impact of channel uncertainty on the SISO decoder performance is reduced. We refer to the latter approach as *improved* MAP detector.

The summations in (20) are taken over the product of the likelihood $\widetilde{W}(y_k | s_k, \hat{H}_k) = e^{-\mathcal{D}_{\mathcal{M}}(s_k, y_k, \hat{H}_k)}$ given a symbol s_k and the estimated channel coefficient $\hat{H}_k$, and of the *a priori* probability on s_k (the term $\prod P_{\text{dec}}$), fed back from the SISO decoder at the previous iteration. In this latter term, the *a priori* probability of the bit $d_k^{j,m}$ itself has been excluded, so as to let the exchange of *extrinsic* informations between the channel decoder and the soft detector. Also, note that this term assumes independent coded bits $d_k^{j,n}$, which is a valid approximation for random interleaving of large size. At the first iteration, no *a priori* information is available on bits $d_k^{j,n}$, therefore the probabilities $P_{\text{dec}}^0(d_k^{j,n})$ and $P_{\text{dec}}^1(d_k^{j,n})$ are set to 1/2. The decoder accepts the LLRs of all coded bits and computes the LLRs of information bits, which are used for decision, at the last iteration.

V. ACHIEVABLE INFORMATION RATES ASSOCIATED TO IMPROVED AND MISMATCHED DETECTION METRICS

A. Instantaneous Achievable Information Rates Of MB-OFDM

In this section, we derive the instantaneous achievable information rates $C_{\mathcal{M}}$ and C_{MM} associated to a receiver based on the improved and mismatched detection rules given by (17) and (13), respectively. To this end, we apply the following theorem provided in [21], to a transmission characterized by the likelihood $W(\mathbf{y}|\mathbf{s}, \mathbf{H}) = \prod_{k=0}^{M-1} \mathcal{CN}(H_k s_k, \sigma_z^2)$ which is a mapping from the input symbols $s \in \mathcal{S}$ with distribution $P_{\mathbf{S}}$, to the set of probability measures on the output alphabet $\mathbf{y} \in \mathcal{Y}$. Furthermore, for any input distribution $P_{\mathbf{S}}$, we define the output distribution as $P_{\mathbf{Y}} = \int W(\mathbf{y}|\mathbf{s}, \mathbf{H}) dP_{\mathbf{S}}(\mathbf{s})$.

Theorem 1: Given a general detection metric $\mathcal{D}_{\mathcal{G}}(\mathbf{s}, \mathbf{y}, \hat{\mathbf{H}})$ a channel and its estimate $(\mathbf{H}, \hat{\mathbf{H}})$, a detector using the metric $\mathcal{D}_{\mathcal{G}}$ achieves the following maximal information rate

$$C_{\mathcal{D}_{\mathcal{G}}} = \max_{P_{\mathbf{S}}} \min_{V \in \mathcal{F}} I(S; Y), \quad (21)$$

where the maximization is performed over all probability distributions $P_{\mathbf{S}}$ on $\mathcal{S}$, and

$$I(S; Y) = D(V || P_{\mathbf{S}} P_{\mathbf{Y}}), \quad (22)$$

where $D(\cdot || \cdot)$ denotes the relative entropy functional [22]. In (21), the set $\mathcal{F}$ denotes the set of all likelihood functions $V(\mathbf{y}|\mathbf{s}, \Upsilon) = \prod_{k=0}^{M-1} \mathcal{CN}(v_k s_k, \sigma_k^2)$ on $\mathcal{S} \times \mathcal{Y}$ that satisfy the following constraints

$$(c_1) : \mathbb{E}_{P_{\mathbf{S}}} [V(\mathbf{y}|\mathbf{s}, \Upsilon)] = \mathbb{E}_{P_{\mathbf{S}}} [W(\mathbf{y}|\mathbf{s}, \mathbf{H})], \quad (23)$$

$$(c_{2,k}) : \mathbb{E}_{P_{\mathbf{S}}} \left[\mathbb{E}_V [\mathcal{D}_{\mathcal{G}}(s_k, y_k, \hat{H}_k)] \right] \leq \mathbb{E}_{P_{\mathbf{S}}} \left[\mathbb{E}_W [\mathcal{D}_{\mathcal{G}}(s_k, y_k, \hat{H}_k)] \right], \quad (24)$$

for every $k = 0, \dots, M-1$, where $\Upsilon = [v_0, \dots, v_{M-1}]^T$.

1) Case of Improved ML Metric: In the following, we aim at solving the above constrained minimization problem for our specific MB-OFDM channel and metric $\mathcal{D}_{\mathcal{G}} = \mathcal{D}_{\mathcal{M}}$ of (17). To this end, we assume that the transmitter has no information about the channel estimate and consequently uniform power allocation is done across subcarriers. Furthermore, we assume a Gaussian i.i.d. input distribution $P_{\mathbf{S}} = \mathcal{CN}(\mathbf{0}, \bar{P} \mathbb{I}_M)$. Under these conditions, the mutual

information (22), written for one subband of the MB-OFDM transmission and averaged over all subcarriers is given by [23]

$$I(S; Y | \Upsilon) = \frac{1}{M} \log_2 \det \left\{ \mathbb{I}_M + \bar{P} \Upsilon_d \Upsilon_d^\dagger \Sigma^{-1} \right\} = \frac{1}{M} \sum_{k=0}^{M-1} \log_2 \left(1 + \frac{\bar{P} |v_k|^2}{\sigma_k^2} \right), \quad (25)$$

where $\Sigma = \text{diag}([\sigma_0^2, \dots, \sigma_{M-1}^2])$ and $\Upsilon_d = \text{diag}(\Upsilon)$. According to Theorem 1, we have to find the covariance matrix Σ and the optimal channel vector Υ_{opt} (these two characterize the pdf $V(\mathbf{y} | \mathbf{s}, \Upsilon)$), so as to minimize the mutual information (25) and to satisfy constraints (23) and (24).

The unknown diagonal covariance matrix Σ is obtained from the constraint (c_1) of (23). It is easily seen that this constraint leads to the equality $V(\mathbf{y} | \Upsilon) = W(\mathbf{y} | \mathbf{H})$. Moreover, from the likelihoods $V(\mathbf{y} | \mathbf{s}, \Upsilon)$ and $W(\mathbf{y} | \mathbf{s}, \mathbf{H})$, it is clear that the conditional random vectors $(\mathbf{y} | \Upsilon)$ and $(\mathbf{y} | \mathbf{H})$ are distributed as $\mathcal{CN}(\mathbf{0}, \Sigma + \bar{P} \Upsilon_d \Upsilon_d^\dagger)$ and $\mathcal{CN}(\mathbf{0}, \Sigma_z + \bar{P} \mathbf{H}_d \mathbf{H}_d^\dagger)$, respectively. Thus the diagonal covariance matrix Σ is obtained as

$$\Sigma = \bar{P}(\mathbf{H}_d \mathbf{H}_d^\dagger - \Upsilon_d \Upsilon_d^\dagger) + \Sigma_z, \quad (26)$$

with its k -th ($k = 0, \dots, M-1$) diagonal element equal to

$$\sigma_k^2 = \bar{P}(|H_k|^2 - |v_k|^2) + \sigma_z^2. \quad (27)$$

Now, in order to find the optimal vector Υ_{opt} , we have to specify the set of inequality constraints $(c_{2,k})$ associated to the metric $\mathcal{D}_{\mathcal{M}}(s_k, y_k, \hat{H}_k)$ of (17). As shown in Appendix II, the k -th inequality constraint set in (24) is equivalent to the set

$$\mathcal{V}_{\mathcal{M},k} = \left\{ v_k : |v_k - a \hat{H}_k|^2 \geq |H_k - a \hat{H}_k|^2 \right\}, \quad (28)$$

where $a = \delta(\lambda \sigma_z^2 - \bar{P} \delta \sigma_{\mathcal{E}}^2) / (\lambda \sigma_z^2 - \bar{P} \delta \sigma_{\mathcal{E}}^2 (1 - \lambda))$; $\lambda = \exp \left\{ \frac{\sigma_z^2}{\bar{P} \delta \sigma_{\mathcal{E}}^2} \right\} E_1 \left(\frac{\sigma_z^2}{\bar{P} \delta \sigma_{\mathcal{E}}^2} \right)$ and $E_1(x) \triangleq \int_x^{+\infty} \frac{\exp\{-u\}}{u} du$ is defined as the exponential integral.

By using (25), (27) and (28) in Theorem 1, the achievable rates $C_{\mathcal{M}}(\mathbf{H}, \hat{\mathbf{H}})$ associated to the improved metric $\mathcal{D}_{\mathcal{M}}$, can be obtained by solving the following minimization, which is equivalent to the initial formulation (21)

$$C_{\mathcal{M}}(\mathbf{H}, \hat{\mathbf{H}}) = \begin{cases} \min_{\Upsilon} \frac{1}{M} \sum_{k=0}^{M-1} \log_2 \left(1 + \frac{\bar{P} |v_k|^2}{\sigma_k^2(|v_k|)} \right) \\ \text{subject to } |v_k - a \hat{H}_k|^2 \geq |H_k - a \hat{H}_k|^2 \quad \text{for } k = 0, \dots, M-1. \end{cases} \quad (29)$$

Note that the k -th term in (29) is an increasing function of $|v_k|^2$ and the constraints are convex. Consequently, the minimum in (29) is obtained at the extremal of the sets [24]. Thus, the problem simplifies to the search of the optimal channel coefficients $v_k = v_{\text{opt},k}$ from the minimization of $\|\Upsilon\|_F^2$ under the constraint sets resulting from the equality in (28). This becomes a classical convex minimization problem that can be solved by using Lagrange multipliers. After simple algebra we get $\Upsilon_{\text{opt}} = [v_{\text{opt},0}, \dots, v_{\text{opt},M-1}]^T$, where

$$v_{\text{opt},k} = \eta_{\mathcal{M},k} \hat{H}_k, \quad k = 0, \dots, M-1, \quad (30)$$

and

$$\eta_{\mathcal{M},k} = \begin{cases} a - \frac{|H_k - a \hat{H}_k|}{|\hat{H}_k|}, & \text{if } a \geq 0 \\ a + \frac{|H_k - a \hat{H}_k|}{|\hat{H}_k|}, & \text{if } a < 0. \end{cases} \quad (31)$$

Finally, the achievable rates associated to metric $\mathcal{D}_{\mathcal{M}}$ are given as follows.

$$C_{\mathcal{M}}(\mathbf{H}, \hat{\mathbf{H}}) = \frac{1}{M} \sum_{k=0}^{M-1} \log_2 \left(1 + \frac{\bar{P} \eta_{\mathcal{M},k}^2 |\hat{H}_k|^2}{\sigma_z^2 + \bar{P} (|H_k|^2 - \eta_{\mathcal{M},k}^2 |\hat{H}_k|^2)} \right). \quad (32)$$

2) *Case of Mismatched ML Metric:* Under the same assumptions as above, we can compute the achievable rates C_{MM} associated to the mismatched metric $\mathcal{D}_{\text{MM}}$ of (13) in the same way. The equality constraint (23) leads to the same result as equation (27). By following similar steps as in Appendix II, we can easily derive the k -th inequality constraint set associated to the metric $\mathcal{D}_{\text{MM}}$ as

$$\mathcal{V}_{\text{MM},k} = \left\{ v_k : \Re(v_k^* \hat{H}_k) \geq \Re(H_k^* \hat{H}_k) \right\} \quad \text{for } k = 0, \dots, M-1, \quad (33)$$

where $\Re(\cdot)$ takes the real part of its argument.

Now, by considering the minimization problem of equation (29) under the constraints of (33), and using the Lagrange method, we obtain

$$C_{\text{MM}}(\mathbf{H}, \hat{\mathbf{H}}) = \frac{1}{M} \sum_{k=0}^{M-1} \log_2 \left(1 + \frac{\bar{P} \eta_{\text{MM},k}^2 |\hat{H}_k|^2}{\sigma_z^2 + \bar{P} (|H_k|^2 - \eta_{\text{MM},k}^2 |\hat{H}_k|^2)} \right), \quad (34)$$

where $\eta_{\text{MM},k}$ is given by

$$\eta_{\text{MM},k} = \frac{\Re(H_k^* \hat{H}_k)}{|\hat{H}_k|^2}. \quad (35)$$

In order to get insight about the expressions (32) and (34), assume that perfect CSIR is available at the receiver, i.e., $H_k = \hat{H}_k$ for $k = 0, \dots, M-1$. In this situation, $\sigma_{\mathcal{E}}^2 = 0$, $\delta = 1$, and we get $\eta_{\text{MM},k} = \eta_{\mathcal{M},k} = 1$ for all k . Consequently, both detectors provide the instantaneous rate

$$C(\mathbf{H}) = \frac{1}{M} \sum_{k=0}^{M-1} \log_2 \left(1 + \frac{\bar{P} |H_k|^2}{\sigma_z^2} \right). \quad (36)$$

In this case, assuming that there is no delay constraint and the transmission time is longer than the channel coherence time, both of the detectors achieve the ergodic capacity [25] which is defined as

$$C_{\text{erg}} = \mathbb{E}_{\mathbf{H}}[C(\mathbf{H})]. \quad (37)$$

B. Evaluation Of Outage Rates Under Imperfect Channel Estimation

In several applications in wireless communications of which MB-OFDM is a special case, the channel is chosen randomly at the beginning of the transmission and is held fixed during the whole transmission session. In fact, in this case the transmission time is not longer than the channel coherence time and consequently the mean of the maximum mutual information in (37) is in general not equal to the channel capacity. On the other hand, the capacity in the Shannon sense does not exist since there is a non-zero probability that the realized $\mathbf{H}$ and its estimate $\hat{\mathbf{H}}$ are not capable of supporting even a very small rate (it is a matter of luck...). In this case, the capacity is a random entity, as it depends on the instantaneous channel $\mathbf{H}$. Thus, the concept of *capacity-versus-outage* [18] has to be invoked. That is, with any given rate R and channel realization $\mathbf{H}$, we associate a set of channels $\Lambda(R, \mathbf{H})$. This set is the largest possible set for which $C(\mathbf{H})$, i.e., the achieved rate for a realization of $\mathbf{H} \in \Lambda(R, \mathbf{H})$ satisfies $C(\mathbf{H}) < R$. Formally, we can write $\Lambda(R, \mathbf{H}) \triangleq \{\mathbf{H} : C(\mathbf{H}) < R\}$. The outage (or failure) probability P^{out} is then determined by $P^{\text{out}} = \text{Prob}(\mathbf{H} \in \Lambda(R, H))$. This is the definition of the classical notion of outage capacity, which implicitly assumes the true channel to be available at the decoder. In contrast, when this knowledge is not available at the receiver, but only the channel estimate, we have to invoke the notion of *estimation induced outage capacity*, which has different properties, as explained in [15].

Following this approach, we make use of the posterior channel distribution of (8) that characterizes our channel estimation process, in order to define the outage probability $P_{\mathcal{M}}^{\text{out}}$ (associated

to the metric $\mathcal{D}_{\mathcal{M}}$, for an outage rate $R \geq 0$ and an estimated channel $\hat{\mathbf{H}}$ as [15]

$$P_{\mathcal{M}}^{\text{out}}(R, \hat{\mathbf{H}}) = \int_{\{\mathbf{H} \in \Lambda_{\mathcal{M}}(R, \hat{\mathbf{H}})\}} d p(\mathbf{H} | \hat{\mathbf{H}}), \quad (38)$$

where $\Lambda_{\mathcal{M}}(R, \hat{\mathbf{H}}) \triangleq \{\mathbf{H} : C_{\mathcal{M}}(\mathbf{H}, \hat{\mathbf{H}}) < R\}$. Using this, the outage rate for an outage probability equal to γ is given by

$$C_{\mathcal{M}}^{\text{out}}(\gamma, \hat{\mathbf{H}}) = \sup_R \{R \geq 0 : P_{\mathcal{M}}^{\text{out}}(R, \hat{\mathbf{H}}) \leq \gamma\}. \quad (39)$$

The outage rates (39) depend on the random channel estimate $\hat{\mathbf{H}}$. Thus, we consider as performance measure, the averaged outage rates over all channel estimates as follows.

$$\overline{C}_{\mathcal{M}}^{\text{out}}(\gamma) = \mathbb{E}_{\hat{\mathbf{H}}} [C_{\mathcal{M}}^{\text{out}}(\gamma, \hat{\mathbf{H}})]. \quad (40)$$

By using (34) and similar steps as above, we can derive the averaged outage rates $\overline{C}_{\text{MM}}^{\text{out}}(\gamma)$ associated to the mismatched receiver. The outage rates associated to the best possible (but not practical) decoder under imperfect channel estimation is derived in [15]. The next section evaluates the outage rates of the improved and mismatched receivers compared to those provided by this decoder, to which we refer as *theoretical* decoder. In our case, the averaged outage rate of the theoretical decoder is given by

$$\overline{C}_{\mathcal{T}}^{\text{out}}(\gamma) = \mathbb{E}_{\hat{\mathbf{H}}} [C_{\mathcal{T}}^{\text{out}}(\gamma, \hat{\mathbf{H}})]. \quad (41)$$

Here, the outage rates $C_{\mathcal{T}}^{\text{out}}(\gamma, \hat{\mathbf{H}})$ are computed according to (38) and (39) by using

$$C_{\mathcal{T}}(\mathbf{H}, \hat{\mathbf{H}}) = \frac{1}{M} \sum_{k=0}^{M-1} \log_2 \left(1 + \frac{\bar{P} |H_k|^2}{\sigma_z^2} \right), \quad (42)$$

where H_k are random channel coefficients drawn from the posterior distribution (8).

VI. SIMULATION RESULTS

In this section, we provide numerical results to evaluate the performance provided by the proposed detector in the presence of channel estimation errors, in comparison with more classical approaches. We focus on the impact of imperfect channel estimation on receiver bit error rate (BER) and the achievable outage rates associated to improved and mismatched detectors. The situation of interest is a BICM scheme, involving the rate 1/2 NRNSC code of constraint length 3 defined in octal form by $(5, 7)_8$. Data symbols belong to 16-QAM constellation with Gray

or set-partition (SP) labeling, and the impact of this labeling is discussed below, since it may enlarge or reduce the improvement brought by the improved metric.

Throughout the simulations, one OFDM symbol is composed of $M = 100$ data subcarriers and the subband channel bandwidth is 528 MHz according to the parameters defined in [19]. An MB-OFDM system employing only the first three subbands starting at 3.1 GHz has been considered. The pseudo-random interleaver operates over the entire frame. The performance evaluation is performed over two channels: (i) the realistic UWB channel model CM1 specified in [26] and (ii) the uncorrelated Rayleigh fading channel. For each frame that contains 100 OFDM symbols, a different realization of the channel has been drawn and remains constant during the whole frame. We use mutually orthogonal pilot sequences for channel estimation and the average pilot-symbol power is equal to the average data-symbol power. Moreover, the number of decoding iterations is set to 4.

A. Bit Error Rate Analysis

First, the BER performance of the improved and mismatched detectors are compared. Let us first address the case of BICM iterative decoding with 16-QAM and Gray labeling. It can be observed from Fig. 3 that for $N = 1$ (the shortest possible training sequence), the improvement in terms of required E_b/N_0 in order to attain a given BER is about 2 dB, compared to the mismatched situation, while the perfect channel knowledge would even be 3 dB better. Obviously, it is also observed that these quantities are reduced (the performance of both the mismatched decoder and the improved one get closer to the perfect channel one) when increasing the length of the training sequence. Note that the performance of the improved receiver with 1 pilot is very close to that of the mismatched receiver with 2 pilots. For comparison, results obtained with uncorrelated Rayleigh block fading channel are illustrated in Fig. 4. It can be observed that for $N = 2$, the SNR to obtain a BER of 10^{-4} is reduced by about 1 dB if the improved MAP detector is used instead of the mismatched detector. The conclusions are otherwise quite similar on the Rayleigh channel.

Similar plots are shown in Fig. 5 for the case of 16-QAM and SP labeling on the CM1 channel. This shows the behavior of the improved detector with respect to the type of labeling. A first observation is that the distance between the mismatched detector and the perfect channel knowledge is even larger (about 6 dB for $N = 1$ and a BER of 7×10^{-4}). Another observation

is that, even if the global performance is largely improved by using the SP labeling when perfect channel knowledge is available, the difference between SP and Gray labeling is not very large with the mismatched decoder, i.e., the sensitivity of the iterative decoder to the channel knowledge seems to be larger for SP labeling. However, the use of the improved metric allows to recover most of the improvement, since, even with a single training sample ($N = 1$), the BER is improved with SP compared to Gray labeling by about 0.9 dB for a BER of 5×10^{-4} . In other words, iterative decoding with SP labeling benefits more from the improved metric than the one with Gray labeling. Otherwise, similar conclusions hold between the SP-labeling curves.

Figures 6-7 show the BER performance versus the number of pilot symbols N at a fixed E_b/N_0 of 12 dB for 16-QAM with Gray and set-partition labeling, respectively. This allows to evaluate the length of training sequence necessary to achieve a certain BER. From Fig. 6 we observe that the improved detector requires 10 pilot symbols per frame to achieve a BER of 10^{-4} at $E_b/N_0 = 12$ dB while the mismatched detector attains this BER for 12 training symbols. From Fig. 7 we notice that the performance loss due to the mismatched receiver with respect to the improved receiver becomes insignificant for $N \geq 12$ (about 11 % of the overall frame of pilot and data symbols). Actually, our results illustrate that the improved detector outperforms the mismatched detector especially when a small numbers of pilot symbols is dedicated to channel estimation, which is in perfect agreement with equation (18). This makes the improved metric particularly useful, since training sequences in multicarrier systems are usually of length 1 or 2. Note also that this improvement is obtained almost at no computational cost.

B. Achievable Outage Rates

We now analyze the achievable outage rates provided by a receiver based on the improved and mismatched detection technique for MB-OFDM under imperfect CSIR. The channel is an uncorrelated i.i.d. Rayleigh fading and the data symbols are assumed to be distributed as $\mathcal{CN}(0, 1)$. Figure 8 shows the expected outage rates (in bits per channel-use) versus the SNR (in dB), obtained by adopting mismatched and improved detection approaches. The outage probability has been fixed to $\gamma = 0.01$ and the channel is estimated by sending $N = 1$ pilot per frame. For comparison, we also display the upper bound on the expected outage rates provided by the theoretical decoder and also the ergodic capacity, given by equations (41) and (37), respectively. The figure clearly shows the sub-optimality of mismatched detection in terms of

expected outage rates compared to the rates provided by the theoretical decoder. It can be observed that the mismatched outage rate is about 4.8 dB (at a mean outage rates of 5 bits) of SNR far from the rates achieved by the theoretical decoder. We note that by adopting the improved receiver, the above SNR gap is reduced by about 1.8 dB.

Similarly, Fig. 9 shows the outage rates obtained by the improved and mismatched receivers when the number of channel-uses for pilot transmission is $N = 2$ per frame. As observed, due to a more accurate channel estimation, both the detectors achieve a higher outage rate as compared to that obtained with $N = 1$. However, we note that at a mean outage rate of 5 bits, the rate of the mismatched receiver is still about 2.8 dB of SNR far from the rate of the theoretical receiver. The figure shows that the improved detector achieves higher rates and decreases the required SNR by about 1 dB at a mean outage rate of 5 bits. An open question is still how the gap to the optimal estimation induced outage capacities could be filled.

Finally, Fig. 10 compares the achievable rates of the improved and mismatched receivers with respect to the number of training symbols at fixed SNR values of 10 and 15 dB. In each case, we also display the corresponding upper bound on the achievable outage rates provided by the theoretical decoder and the ergodic capacity. As can be seen, the improved detector requires fewer pilot symbols in order to provide a prescribed mean outage rate. For instance, the gain in the number of channel-uses for pilot transmission is 2 pilots at a mean outage rates of 3.5 bits. These simulation results also confirm the obvious expectation that under near perfect channel estimation (here for $N \geq 16$), the performance of mismatched and improved receivers is almost the same.

VII. CONCLUSION

The problem of detection in a practical communication system where the receiver has only access to a noisy estimate of the channel provided by pilot symbols was investigated. Based on a characterization of the channel estimation process, we proposed a general detector design that takes into account the imperfect channel available at the receiver. In the case of ML detection, this approach led to an improved ML metric that we used in the derivation of a modified iterative MAP detector. We also derived the expression of the achievable outage rates associated to the improved and mismatched ML metric. Our numerical results indicated that mismatched detection is sub-optimal in terms of BER and achievable outage rates, especially for short training sequence lengths, and confirmed the adequacy of the improved detector under imperfect channel estimation.

The impact of the improved metric was shown to be dependent on the type of labeling used in the construction of the symbols, and it was shown that this new metric could save the improvements brought by set-partitioning labeling in the case of poor channel estimates (obtained by very short training sequences). This performance improvement was obtained without requiring additional complexity in the receiver. Although the improved detector was able to reduce the SNR gap between the mismatched and the theoretical receiver at fixed outage rates, the derivation of other practical detectors providing closer rates to the theoretical capacity limit is still an open problem.

APPENDIX I

A. Derivation of the A Posteriori Probability (8)

The following theorem is derived in [27].

Theorem 1.1: Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be circularly symmetric complex Gaussian random vectors with zero means and full-rank covariance matrices $\Sigma_{ij} = \mathbb{E}\{\mathbf{x}_i \mathbf{x}_j^\dagger\}$. Then the conditional random vector $\mathbf{x}_1 | \mathbf{x}_2 \sim \mathcal{CN}(\boldsymbol{\mu}, \Sigma)$ is a circularly symmetric complex Gaussian random vector with mean $\boldsymbol{\mu} = \Sigma_{12} \Sigma_{22}^{-1} \mathbf{x}_2$ and covariance matrix $\Sigma = \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$.

We set $\mathbf{x}_1 = \mathbf{H}$ and $\mathbf{x}_2 = \hat{\mathbf{H}}$. From equation (7) and the assumption $\mathbf{H} \sim \mathcal{CN}(0, \Sigma_H)$, we have $\Sigma_{11} = \Sigma_{12} = \Sigma_H$ and $\Sigma_{22} = \Sigma_H + \Sigma_\varepsilon$ in Theorem 1.1. According to this theorem, the conditional random vector $\mathbf{H} | \hat{\mathbf{H}}$ has a circularly symmetric complex Gaussian distribution with

$$\text{mean} = \Sigma_\Delta \hat{\mathbf{H}} \quad \text{where} \quad \Sigma_\Delta \triangleq \Sigma_H (\Sigma_H + \Sigma_\varepsilon)^{-1}, \text{ and} \quad (43)$$

$$\text{covariance matrix} = \Sigma_H - \Sigma_H (\Sigma_H + \Sigma_\varepsilon)^{-1} \Sigma_H^\dagger = \Sigma_\Delta \Sigma_\varepsilon. \quad (44)$$

The equivalence in (44) can be seen by left multiplying both sides of $(\Sigma_\varepsilon + \Sigma_H) - \Sigma_H^\dagger = \Sigma_\varepsilon$ by $\Sigma_H (\Sigma_\varepsilon + \Sigma_H)^{-1}$.

As a result, we obtain the a posteriori pdf (8).

B. Evaluation of the Likelihood Function (14)

In order to evaluate the expectation in (14), we use the following theorem derived in [28].

Theorem 1.2: For a circularly symmetric complex random vector $\mathbf{u} \sim \mathcal{CN}(\mathbf{m}, \Sigma)$ with mean $\mathbf{m} = \mathbb{E}[\mathbf{u}]$ and covariance matrix $\Sigma = \mathbb{E}[\mathbf{u} \mathbf{u}^\dagger] - \mathbf{m} \mathbf{m}^\dagger$, and a Hermitian matrix $\mathbf{A}$ such that $\mathbb{I} + \Sigma \mathbf{A} > 0$, we have

$$\mathbb{E}_{\mathbf{u}} \left[\exp \left\{ -\mathbf{u}^\dagger \mathbf{A} \mathbf{u} \right\} \right] = \frac{\exp \left\{ -\mathbf{m}^\dagger \mathbf{A} (\mathbb{I} + \Sigma \mathbf{A})^{-1} \mathbf{m} \right\}}{\det \{ \mathbb{I} + \Sigma \mathbf{A} \}}. \quad (45)$$

Let us define $\mathbf{u} = \mathbf{y} - \mathbf{H}_d \mathbf{s}$. Using the *a posteriori* distribution of (8) and after some algebra, we can derive the conditional pdf of $\mathbf{u}$ given $\mathbf{s}$ and $\hat{\mathbf{H}}_d$ as $\mathbf{u} | (\mathbf{s}, \hat{\mathbf{H}}_d) \sim \mathcal{CN}(\mathbf{m}_u, \Sigma_u)$, where $\mathbf{m}_u = \mathbf{y} - \Sigma_\Delta \hat{\mathbf{H}}_d \mathbf{s}$ and $\Sigma_u = \Sigma_\Delta \Sigma_\varepsilon \mathbf{S}_d \mathbf{S}_d^\dagger$. We further let $\mathbf{A} = \Sigma_z^{-1}$. By applying Theorem 1.2, we get (14) as

$$\begin{aligned} \widetilde{W}(\mathbf{y} | \hat{\mathbf{H}}, \mathbf{s}) &= \mathbb{E}_{\mathbf{H}_d | \hat{\mathbf{H}}_d} \left[\frac{\exp \left\{ -(\mathbf{y} - \mathbf{H}_d \mathbf{s})^\dagger \Sigma_z^{-1} (\mathbf{y} - \mathbf{H}_d \mathbf{s}) \right\}}{\det \{ \pi \Sigma_z \}} \right] \\ &= \frac{\exp \left\{ -(\mathbf{y} - \Sigma_\Delta \hat{\mathbf{H}}_d \mathbf{s})^\dagger \Sigma_z^{-1} (\mathbb{I} + \Sigma_\Delta \Sigma_\varepsilon \mathbf{S}_d \mathbf{S}_d^\dagger \Sigma_z^{-1})^{-1} (\mathbf{y} - \Sigma_\Delta \hat{\mathbf{H}}_d \mathbf{s}) \right\}}{\det \{ \pi \Sigma_z (\mathbb{I} + \Sigma_\Delta \Sigma_\varepsilon \mathbf{S}_d \mathbf{S}_d^\dagger \Sigma_z^{-1}) \}} \end{aligned} \quad (46)$$

Since Σ_z , Σ_Δ and Σ_ε are diagonal matrices, the latter equation is rewritten as

$$\begin{aligned} \widetilde{W}(\mathbf{y} | \hat{\mathbf{H}}, \mathbf{s}) &= \frac{\exp \left\{ -(\mathbf{y} - \delta \hat{\mathbf{H}}_d \mathbf{s})^\dagger (\Sigma_z + \delta \sigma_\varepsilon^2 \mathbf{S}_d \mathbf{S}_d^\dagger)^{-1} (\mathbf{y} - \delta \hat{\mathbf{H}}_d \mathbf{s}) \right\}}{\det \{ \pi (\Sigma_z + \delta \sigma_\varepsilon^2 \mathbf{S}_d \mathbf{S}_d^\dagger) \}} \\ &= \mathcal{CN} \left(\delta \hat{\mathbf{H}}_d \mathbf{s}, \Sigma_z + \delta \sigma_\varepsilon^2 \mathbf{S}_d \mathbf{S}_d^\dagger \right). \end{aligned} \quad (47)$$

APPENDIX II

Details On The Derivation Of The Inequality Constraint (28)

Using the expression of the modified metric $\mathcal{D}_M$ from (17), the left-hand side of the k -th constraint in (24) can be expanded as follows (the index k is omitted for notational brevity).

$$\begin{aligned} &\mathbb{E}_{P_S} \left[\mathbb{E}_{V_Y | S} [\mathcal{D}_M(s, y, \hat{H})] \right] \\ &= \underbrace{\mathbb{E}_{P_S} [\log \pi (\sigma_z^2 + \delta \sigma_\varepsilon^2 |s|^2)]}_K + \mathbb{E}_{P_S} \mathbb{E}_{V_Y | S} \left[\frac{|y|^2 - 2\operatorname{Re}(y^* \delta \hat{H} s) + \delta^2 |\hat{H}|^2 |s|^2}{\sigma_z^2 + \delta \sigma_\varepsilon^2 |s|^2} \right] \\ &= K + \mathbb{E}_{P_S} \left[\frac{\mathbb{E}_{V_Y | S} [|y|^2] - 2\operatorname{Re}(\mathbb{E}_{V_Y | S} [y^*] \delta \hat{H} s) + \delta^2 |\hat{H}|^2 |s|^2}{\sigma_z^2 + \delta \sigma_\varepsilon^2 |s|^2} \right] \\ &= K + \mathbb{E}_{P_S} \left[\frac{\sigma^2 + |v|^2 |s|^2 - 2\delta \operatorname{Re}(v^* \hat{H}) |s|^2 + \delta^2 |\hat{H}|^2 |s|^2}{\sigma_z^2 + \delta \sigma_\varepsilon^2 |s|^2} \right] \\ &= K + \mathbb{E}_{P_S} \left[\frac{\sigma^2 + |v - \delta \hat{H}|^2 |s|^2}{\sigma_z^2 + \delta \sigma_\varepsilon^2 |s|^2} \right]. \end{aligned} \quad (48)$$

Similarly, the right-hand side of (24) can be obtained as

$$\mathbb{E}_{P_S} \left[\mathbb{E}_{W_Y | S} [\mathcal{D}_M(s, y, \hat{H})] \right] = K + \mathbb{E}_{P_S} \left[\frac{\sigma_z^2 + |H - \delta \hat{H}|^2 |s|^2}{\sigma_z^2 + \delta \sigma_\varepsilon^2 |s|^2} \right]. \quad (49)$$

Using (48) and (49) in (24), the inequality constraint is written as

$$\mathbb{E}_{P_S} \left[\frac{\sigma^2 + |v - \delta \hat{H}|^2 |s|^2}{\sigma_z^2 + \delta \sigma_{\hat{E}}^2 |s|^2} \right] \leq \mathbb{E}_{P_S} \left[\frac{\sigma_z^2 + |H - \delta \hat{H}|^2 |s|^2}{\sigma_z^2 + \delta \sigma_{\hat{E}}^2 |s|^2} \right]. \quad (50)$$

Now, in order to evaluate the expectation in (50), we introduce the following lemma.

Lemma 1: Assume $s \sim \mathcal{CN}(0, \beta)$ and $x = |s|^2$ be a centered Chi-squared random variable with two degrees of freedom with pdf $p(x) = \frac{1}{\beta} e^{-x/\beta}$, and let a, b, c and d be real and positive scalars. We have

$$\mathbb{E}_X \left[\frac{a + bx}{c + dx} \right] = \frac{b}{d} + \left(\frac{a}{d\beta} - \frac{bc}{d^2\beta} \right) \exp \left\{ \frac{c}{d\beta} \right\} E_1 \left(\frac{c}{d\beta} \right), \quad (51)$$

where $E_1(\alpha) \triangleq \int_{\alpha}^{\infty} \frac{\exp\{-u\}}{u} du$ is the exponential integral.

Proof: The proof is easy and thus omitted for brevity.

By using *Lemma 1* and replacing a, b, c and d by their respective values from (50), we get

$$\frac{|v - \delta \hat{H}|^2}{\delta \sigma_{\hat{E}}^2} + \left[\frac{\sigma^2}{(\delta \sigma_{\hat{E}}^2) \bar{P}} - \frac{|v - \delta \hat{H}|^2 \sigma_z^2}{(1 - \delta^2) \bar{P}} \right] \lambda \leq \frac{|H - \delta \hat{H}|^2}{\delta \sigma_{\hat{E}}^2} + \left[\frac{\sigma_z^2}{\delta \sigma_{\hat{E}}^2 \bar{P}} - \frac{|H - \delta \hat{H}|^2 \sigma_z^2}{(1 - \delta^2) \bar{P}} \right] \lambda \quad (52)$$

where $\lambda \triangleq \exp \left\{ \frac{\sigma_z^2}{\delta \sigma_{\hat{E}}^2 \bar{P}} \right\} E_1 \left(\frac{\sigma_z^2}{\delta \sigma_{\hat{E}}^2 \bar{P}} \right)$.

After some algebraic manipulations, (28) can be derived from (52).

ACKNOWLEDGMENT

The authors would like to acknowledge many useful discussions with P. Piantanida in which he provided insightful and accurate comments.

REFERENCES

- [1] A. Batra, J. Balakrishnan, G. R. Aiello, J. R. Foerster, and A. Dabak, "Design of multiband OFDM system for realistic UWB channel environments," *IEEE Trans. Microwave Theory and Techniques*, vol. 52, pp. 2123–2138, Sep. 2004.
- [2] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 927–945, May 1998.
- [3] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 686–693, Nov. 1991.
- [4] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, pp. 873–887, May 1992.
- [5] X. Li, A. Chindapol, and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding and 8PSK modulation," *IEEE Trans. Commun.*, vol. 50, pp. 1250–1257, Aug. 2002.
- [6] Y. Huang and J.A. Ritcey, "16-QAM BICM-ID in fading channels with imperfect channel state information," *IEEE Trans. Commun.*, vol. 2, pp. 1000–1007, Sept. 2003.
- [7] P. Garg, R. K. Mallik, and H. M. Gupta, "Performance analysis of space-time coding with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 257–265, Jan. 2005.

- [8] M. Speth, S. A. Fechel, G. Fock, and H. Meyr, "Optimum receiver design for wireless broadband systems using OFDM-part I," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1668–1677, Nov. 1999.
- [9] A. Leke and J. M. Cioffi, "Impact of imperfect channel knowledge on the performance of multicarrier systems," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, pp. 951–955, Nov. 1998.
- [10] K. Ahmed, C. Tepedelenhoglu, and A. Spanias, "Effect of channel estimation on pair-wise error probability in OFDM," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, pp. 745–748, May 2004.
- [11] A. Lapidoth and S. Shamai, "Fading channels: How perfect need 'perfect side information' be?," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1118–1134, May 2002.
- [12] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criteria in the presence of channel estimation errors, mobility, and multiple paths," *IEEE Trans. Commun.*, vol. 47, pp. 199–207, Feb. 1999.
- [13] G. Taricco and E. Biglieri, "Space-time decoding with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1874, 1888 2005.
- [14] J. K. Cavers and P. Ho, "Analysis of the error performance of trellis-coded modulations in Rayleigh-fading channels," *IEEE Trans. Commun.*, vol. 40, no. 1, pp. 74–83, Jan. 1992.
- [15] P. Piantanida, G. Matz, and P. Duhamel, "Estimation-induced outage capacity of Ricean channels," in *Proc. Signal Process. Advances Wireless Commun. (SPAWC)*, July 2006.
- [16] S. Sadough, P. Piantanida, and P. Duhamel, "MIMO-OFDM optimal decoding and achievable information rates under imperfect channel estimation," in *Proc. Signal Process. Advances Wireless Commun. (SPAWC)*, June 2007.
- [17] P. Piantanida, S. Sadough, and P. Duhamel, "On the outage capacity of a practical decoder using channel estimation accuracy," in *Proc. IEEE International Symp. Inform. Theory (ISIT)*, Jun. 2007.
- [18] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
- [19] A. Batra, J. Balakrishnan, and A. Dabak, "Multiband OFDM physical layer proposal for IEEE 802.15 task group 3a," Tech. Rep., IEEE, July 2003.
- [20] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, pp. 284–287, March 1974.
- [21] Neri Merhav, Gideon Kaplan, Amos Lapidoth, and Sholmo Shamai, "On information rates for mismatched decoders," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1953–1967, Nov. 1994.
- [22] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications. Wiley & Sons New York, 1991.
- [23] I. E. Telatar, "Capacity of multi-antenna gaussian channels," Tech. Rep., AT&T Bell Labs Tech. Memo., 1995.
- [24] J. Hirriart-Urrty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I.*, Springer-Verlag, 1993.
- [25] L. Ozarow, S. Shamai, and A. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Inform. Theory*, vol. 43, pp. 359–378, May 1994.
- [26] J. Foerster, "Channel modeling sub-committee report final," Tech. Rep., IEEE802.15-02/490, 2003.
- [27] M. Bilodeau and D. Brenner, *Theory of Multivariate Statistics*, New York: Springer, 1999.
- [28] M. Schwartz, W. R. Bennett, and S. Stein, *Communications Systems and Techniques*, McGraw-Hill, 1966.

Fig. 1. Transmitter architecture of the multiband OFDM system.

Fig. 2. Block diagram of the improved receiver under imperfect channel estimation.

Fig. 3. BER performance of improved and mismatched MAP detector over the CM1 channel, training sequence lengths $N \in \{1, 2, 8\}$, 16-QAM modulation with Gray labeling.

Fig. 4. BER performance of the improved and mismatched MAP detector over i.i.d. Rayleigh fading channel, training sequence lengths $N \in \{1, 2, 8\}$, 16-QAM modulation with Gray labeling.

Fig. 5. BER performance of the improved and mismatched MAP detector over the CM1 channel, training sequence lengths $N \in \{1, 2, 8\}$, 16-QAM modulation with set-partition labeling.

Fig. 6. Reduction of the number of training sequence at $E_b/N_0 = 12 \text{ dB}$ over the CM1 channel, 16-QAM modulation with Gray labeling .

Fig. 7. Reduction of the number of training sequence at $E_b/N_0 = 12$ dB over the CM1 channel, 16-QAM modulation with set-partition labeling.

Fig. 8. Expected outage rates of MB-OFDM transmission versus SNR for $N = 1$ pilot per frame and different detection approaches, outage probability $\gamma = 0.01$, i.i.d. Rayleigh fading channel.

Fig. 9. Expected outage rates of MB-OFDM transmission versus SNR for $N = 2$ pilots per frame and different detection approaches, outage probability $\gamma = 0.01$, i.i.d. Rayleigh fading channel.

Fig. 10. Expected outage rates of MB-OFDM transmission versus the number of training sequence for different detection approaches at SNR of 10 and 15 dB, outage probability $\gamma = 0.01$, i.i.d. Rayleigh fading channel.