Near Optimal Iterative Channel Estimation for KSP-OFDM

Dieter Van Welden and Heidi Steendam

Abstract—In this correspondence, we propose an iterative "turbo" channel estimation algorithm for known symbol padding (KSP) orthogonal frequency-division multiplexing (OFDM), where the guard interval is filled with pilot symbols. Additional pilot symbols are transmitted on some of the OFDM carriers. The channel estimation algorithm is based on the expectation-maximization (EM) algorithm. In the initialization phase of this iterative algorithm, the received time-domain samples are first converted to the frequency domain, and the initial channel estimate is based on the observation of the pilot carriers only. Then the EM-algorithm switches back to the time-domain and updates the channel estimates until convergence is reached. The proposed estimator performs very good: the mean square error (MSE) performance of the proposed estimator is close to the Cramér–Rao lower bound (CRB) corresponding to the all pilots case, for the SNR region of practical interest, and the resulting bit error rate essentially coincides with the case of the perfectly known channel.

Index Terms— Channel estimation, iterative methods, OFDM.

I. INTRODUCTION

Multicarrier transmission (MC) [1] is widely accepted as the transmission technique most suited for high data rate transmission over dispersive channels. This is demonstrated by the various standards based on the MC-technique: xDSL [2], digital audio and video broadcasting [3], [4], wireless LAN [5], etc. To avoid intersymbol interference between successively transmitted MC blocks, a guard interval is inserted between the MC blocks. Among the most popular guard interval techniques, we find the cyclic prefix (CP) technique and the zero-padding (ZP) technique [6], [7]. In the CP technique [6], the last samples of each MC block are copied and added in front of the MC block, whereas in the ZP technique [6], no signal is transmitted during the guard interval. In these two guard interval techniques, maximum-likelihood (ML) channel estimation from pilot carriers is trivial and equalization can be performed in the frequency domain with low complexity [8]. However, for these two guard interval techniques have no control about the content of the guard interval, which makes the guard interval not very useful for channel estimation and sometimes insufficient for synchronization tasks [9], and therefore a waste of resources. Indeed, the main disadvantage of these two guard interval techniques is the ambiguity problem that occurs in low-complexity timing synchronizers [8], [10]. This unability to find the boundaries of the transmitted OFDM blocks will cause residual intersymbol interference, resulting in an overall performance degradation.

In this contribution, we consider another guard interval technique, i.e., known symbol padding (KSP) [9], where the guard interval consists of known pilot samples. As the guard interval does not contain any data symbols, the pilot symbols in the guard interval can be optimized to obtain desirable properties for synchronization tasks. In [11], it is shown that this technique, unlike CP-OFDM and ZP-OFDM, does not suffer from this ambiguity in timing synchronization [9]. As the samples of the guard interval are known, they can be used for data-aided channel estimation. However, usually the length of the guard interval is only slightly larger than the duration of the channel impulse response, so the number of pilot symbols in the guard interval is not sufficient to perform accurate channel estimation. Therefore, extra pilots are inserted in the transmitted MC signal to improve the channel estimation accuracy. In the literature, several channel estimation techniques have been proposed [12]–[16]. It turns out that channel estimation in KSP-OFDM is harder than in CP-OFDM and ZP-OFDM, because of the combination of frequency-domain pilots and time-domain pilots. ML channel estimation in KSP-OFDM is very complex and suboptimal estimation techniques must be used. The algorithms from [12]–[14] assume that the unknown data symbols are Gaussian distributed. In [14], a suboptimal ML-based channel estimation algorithm is proposed. However, [14] assumes that the autocorrelation matrix of the disturbance (containing contributions from the noise, the data symbols and the channel) is known. Hence, before this channel estimator can be used, first the autocorrelation matrix must be estimated from the received signal. Further, even if the autocorrelation matrix is perfectly known, the resulting mean square error (MSE) shows an error floor at high SNR, indicating that the presence of the unknown data symbols disturbs the channel estimation. In [13] channel estimation for KSP transmission over stationary frequency-selective channels is considered, which is a special case of the situation considered in [14]. In [15], the signals transmitted on the pilot carriers are selected such that the last IFFT outputs correspond to the pilot symbols from the KSP sequence. Before the OFDM block, the same samples are transmitted, so we can consider this situation as a special case of CP-OFDM. An equalizer is then trained based on the received KSP sequences of several consecutive blocks. The transmitted signals on the pilot carriers also contain a contribution of the unknown data symbols, so only the KSP sequence in the time domain can be used to perform synchronization. In [16], the effect of the data symbols on the channel estimation is reduced by applying a linear transform to the observed time-domain samples; the resulting observations can be split into a part that depends on the data symbols and a part that is (nearly) independent of the data symbols. In the subset estimator from [16], only the latter subset of observations is used to estimate the channel. Although performing better than the estimator from [14], the subset estimator also shows an error floor at high SNR because of the residual interference from the data symbols.

To deal with the unknown data symbols, we can apply the expectation-maximization (EM) algorithm [17], which is an iterative algorithm that converges to the ML estimate. In [18], EM-based channel estimation algorithms operating in the frequency domain are proposed for CP-OFDM. However, when this algorithm is applied to KSP-OFDM, the pilot symbols from the guard interval cannot optimally be used for the channel estimation, as in the algorithm from [18], the samples from the guard interval are thrown away. The extension of this algorithm to take also into account the guard interval samples to estimate the channel in the frequency domain is not straightforward. Therefore, we propose in this correspondence an iterative channel estimator for KSP-OFDM operating in the time domain, using both the samples from the data part and the samples from the guard interval. To initialize the algorithm, first the channel is estimated using a data-aided estimator that operates in the frequency domain. Because of the orthogonality of the carriers, the pilot carriers can easily be separated from the data carriers.
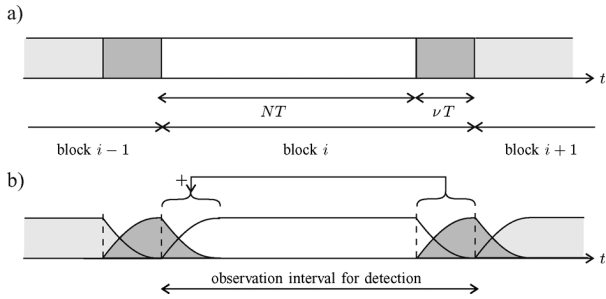
a)

b)

Fig. 1. Time-domain signal of KSP-OFDM a) transmitted signal b) received signal and observation interval.

As the observations used for the channel estimation are the FFT outputs corresponding to the pilot carriers, the observations are independent of the unknown data symbols, and therefore the data symbols will not disturb the channel estimation. Therefore, unlike the algorithms in [14] and [16], this initial frequency domain channel estimate causes no error floor at high SNR. After the initialization phase, the estimator switches back to the time domain, allowing all available information to be used to iteratively update the channel estimate. The performance of the proposed algorithm is compared with the estimator from [14], the subset estimator from [16] and the Cramér–Rao lower bound (CRB) corresponding to the all pilots case. It turns out that the proposed estimator has nearly optimal performance. For the SNR region of practical interest, the MSE performance of the proposed estimator is close to the former CRB. Further, the resulting bit error rate essentially coincides with the case of the perfectly known channel.

II. SYSTEM DESCRIPTION

We consider a KSP-OFDM system with N carriers and a guard interval length ν, as shown in Fig. 1. The guard interval (dark gray area in Fig. 1) consists of ν known pilot samples $\mathbf{b}_g = (b_g(0) \ldots b_g(\nu-1))^T$. The N symbols transmitted on the carriers during the ith OFDM block are denoted $\mathbf{a}_i = (a_i(0), \ldots, a_i(N-1))^T$. This vector $\mathbf{a}_i$ consists of $M - \nu$ pilot symbols and $N - M + \nu$ information-carrying data symbols, denoted as $\mathbf{b}_c^{(i)} = (b_c^{(i)}(0) \ldots b_c^{(i)(M-\nu-1)})^T$ and $\mathbf{a}_d^{(i)} = (a_d^{(i)}(0) \ldots a_d^{(i)}(N-M+\nu-1))^T$, respectively. Hence, the total number of pilots transmitted on the carriers and in the guard interval equals M. We define the sets I_p and I_d as the sets of carriers modulated by pilots and data, respectively, where $I_p \cup I_d = \{0, \ldots, N-1\}$. The $N + \nu$ time-domain samples during block i are then given by

$$\mathbf{s}_i = \sqrt{\frac{N}{N+\nu}} \begin{pmatrix} \mathbf{F}^+ \mathbf{a}_i \\ \mathbf{b}_g \end{pmatrix}. \tag{1}$$

In (1), $\mathbf{F}$ is the $N \times N$ matrix corresponding to the fast Fourier transform (FFT) operation, with $\mathbf{F}_{k,\ell} = (1/\sqrt{N})e^{-j2\pi(k\ell/N)}$, and the superscript + denotes the Hermitian transpose. Hence, the matrix $\mathbf{F}^+$ corresponds to the inverse FFT (IFFT) operation. We assume that the data symbols are independently and identically distributed (i.i.d.) and have energy per symbol $E[|a_d(n)|^2] = E_s$. Further, we assume that $E[|b_c(n)|^2] = E[|b_g(m)|^2] = E_s$. The normalization factor $\sqrt{N/(N+\nu)}$ in (1) implies that $E[|\mathbf{s}_i|^2] = NE_s$.

The KSP-OFDM signal is transmitted over a dispersive channel with L taps; the channel impulse response is given by the vector $\mathbf{h} = (h(0), \ldots, h(L-1))^T$. We select $\nu \geq L-1$ in order to avoid interference between successively transmitted OFDM symbols. The signal is disturbed by additive white Gaussian noise $\mathbf{w}$, of which the statistically independent components $w(k)$ have zero mean and variance N_0. Without loss of generality, we restrict our attention to the

detection of the OFDM block with index $i = 0$, and we drop the block index for notational convenience. Considering the observation interval shown in Fig. 1(b), we can write the received $N + \nu$ time-domain samples as

$$\mathbf{r} = \mathbf{H}_{ch} \mathbf{s} + \mathbf{w}. \tag{2}$$

The $(N+\nu) \times (N+\nu)$ channel matrix $\mathbf{H}_{ch}$ is given by

$$(\mathbf{H}_{ch})_{k,k'} = h\left(|k-k'|_{N+\nu}\right) \tag{3}$$

where $|x|_K$ is the modulo-K reduction of x yielding a result in the interval $[0, K[$.

For data detection, the contribution from the guard interval pilots must first be subtracted from the received signal. Then, the last ν samples from the observation interval, which now contain only a data component as the contribution from the guard interval pilots is removed, are added to the first ν samples of the OFDM symbol, and the resulting samples are transformed to the frequency domain by applying an FFT. Note however that the guard interval pilots are affected by the unknown channel; hence, before their contribution can be subtracted from the received signal, the channel has to be estimated first.

III. CHANNEL ESTIMATION

For channel estimation, we rewrite the observation model (2) as

$$\mathbf{r} = \mathbf{Bh} + \mathbf{Ah} + \mathbf{w}. \tag{4}$$

In (4), the $(N+\nu) \times L$ matrix $\mathbf{B}$ contains the contributions from the pilot symbols. It can be split into the contribution from the guard interval pilots and the pilot carriers: $\mathbf{B} = \mathbf{B}_g + \mathbf{B}_c$. The matrix $\mathbf{B}_c$ contains the contributions from the pilot carriers and is given by

$$(\mathbf{B}_c)_{k,\ell} = \sqrt{\frac{N}{N+\nu}} s_p(k-\ell). \tag{5}$$

The vector $\mathbf{s}_p$ equals the N-point IFFT of the pilot carriers only, i.e., $\mathbf{s}_p = \mathbf{F}_p \mathbf{b}_c$, where the $N \times (M-\nu)$ matrix $\mathbf{F}_p$ consists of a subset of columns of the IFFT matrix $\mathbf{F}^+$ corresponding to the set I_p of pilot carriers. Note that $s_p(k) = 0$ for $k < 0$ or $k \geq N$. The contribution $\mathbf{B}_g$ from the guard interval pilots is given by

$$(\mathbf{B}_g)_{k,\ell} = \sqrt{\frac{N}{N+\nu}} b_g\left(|k-\ell+\nu|_{N+\nu}\right) \tag{6}$$

with $b_g(k) = 0$ for $k \geq \nu$. The contributions from the unknown data symbols in (4) are collected in the $(N+\nu) \times L$ matrix $\mathbf{A}$:

$$(\mathbf{A})_{k,\ell} = \sqrt{\frac{N}{N+\nu}} s_d(k-\ell) \tag{7}$$

where $\mathbf{s}_d = \mathbf{F}_d \mathbf{a}_d$ and $\mathbf{F}_d$ consists of the $N + \nu - M$ columns of $\mathbf{F}^+$ corresponding to the data carrier positions I_d. We introduce the matrix $\mathbf{C}$ to write (4) in a more compact form

$$\mathbf{r} = \mathbf{Ch} + \mathbf{w} \tag{8}$$

where $\mathbf{C} = \mathbf{B} + \mathbf{A}$.

A. EM Estimation

The EM algorithm [17] is an iterative method to obtain an ML estimate of a parameter vector $\boldsymbol{\theta}$ based on an observation $\mathbf{r}$, where $\mathbf{r}$ depends on unobserved data $\mathbf{y}$. Each iteration consists of an expectation (E) step, and a maximization (M) step. In the E-step, the log-likelihood $\log p(\mathbf{r}|\mathbf{y}, \boldsymbol{\theta})$ is averaged over the unobserved data, given the observation $\mathbf{r}$ and the last estimate of $\boldsymbol{\theta}$

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}_{k-1}) = \int \log p(\mathbf{r}|\mathbf{y}, \boldsymbol{\theta}) p(\mathbf{y}|\mathbf{r}, \hat{\boldsymbol{\theta}}_{k-1}) d\mathbf{y} \tag{9}$$

where k is the iteration index and $\hat{\boldsymbol{\theta}}_{k-1}$ denotes the estimate of $\boldsymbol{\theta}$ obtained in the previous iteration. The M-step comprises the maximization of (9) with respect to $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}}_k = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}_{k-1}). \tag{10}$$

B. Step 1: The Initial Channel Estimate

It is clear from (9) that, in order to start up the EM-algorithm, an initial estimate of the channel should be available. In this correspondence, we consider a data-aided estimation method to initialize the iterative algorithm. To obtain the estimator, we follow a similar reasoning as in [16]: we consider an invertible transform independent of the parameter to be estimated that results in a part of the observation to be data-free. However, in contrast with [16], we do not make any approximations. As it is impossible to find a linear transform independent of the channel vector $\mathbf{h}$ that makes the last ν samples of the observation interval data-free, a truly data-free observation consists of $M - \nu$ samples only.

Let us consider the following invertible transform:

$$\mathbf{r}' = \left(\begin{array}{c|c} \mathbf{F} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I}_\nu \end{array} \right) \left(\begin{array}{c|c} \mathbf{I}_N & \begin{smallmatrix} \mathbf{I}_\nu \\ \mathbf{0} \end{smallmatrix} \\ \hline \mathbf{0} & \mathbf{I}_\nu \end{array} \right) \mathbf{r} \triangleq \mathbf{T}\mathbf{r} \tag{11}$$

where $\mathbf{I}_K$ is the $K \times K$ identity matrix. According to this transform, the last ν samples from the observation interval are added to the first ν samples [as indicated in Fig. 1(b)]; this restores the orthogonality (over the first N samples) between the carriers of the OFDM system. Then an N-point FFT is applied to the first N samples (i.e., we convert these samples to the frequency domain) while the last ν samples are not transformed. As the carriers are orthogonal, data carriers do not interfere with pilot carriers. Hence, if we use as observation subset the $M - \nu$ FFT outputs at the pilot carrier positions only, these observations are data-free and ML estimation of the channel is simple.

The $M - \nu$ observations corresponding to the set I_p of pilot carrier positions can be collected in the vector $\mathbf{r}'_p$:

$$\mathbf{r}'_p = \mathbf{B}'\mathbf{h} + \mathbf{w}' \tag{12}$$

where the noise component $\mathbf{w}'$ is zero-mean Gaussian distributed with autocorrelation matrix $\mathbf{R}'$ given by

$$(\mathbf{R}')_{k,k'} = N_0 \left(\delta_{k,k'} + \frac{1}{N} \sum_{\ell=0}^{\nu-1} e^{-j2\pi \frac{(n_k - n_{k'})\ell}{N}} \right) \tag{13}$$

and $\mathbf{B}' = \mathbf{B}'_p + \mathbf{B}'_g$ is a $(M - \nu) \times L$ matrix. The matrix $\mathbf{B}'_p$ corresponds to the contributions from the pilot carriers, i.e.,

$$(\mathbf{B}'_p)_{k,\ell} = \sqrt{\frac{N}{N+\nu}} b_c(k) e^{-j2\pi \frac{k\ell}{N}} \quad k \in I_p, \ell = 0, \ldots, L-1 \tag{14}$$

and $\mathbf{B}'_g$ to the contributions from the guard interval pilots, i.e.,

$$\mathbf{B}'_g = \sqrt{\frac{N}{N+\nu}} \mathbf{F}_{\nu,p} \mathbf{B}_{g,\nu} \tag{15}$$

where $(\mathbf{F}_{\nu,p})_{k,\ell} = (1/\sqrt{N}) e^{-j2\pi(k\ell/N)}$, $k \in I_p$, $\ell = 0, \ldots, \nu-1$ and $(\mathbf{B}_{g,\nu})_{k,\ell} = b_g(|k - \ell|_\nu)$.

The ML estimate of $\mathbf{h}$ based on the observation $\mathbf{r}'_p$ is defined as [19]

$$\hat{\mathbf{h}}_{ML} = \arg\max_{\mathbf{h}} p\left(\mathbf{r}'_p | \mathbf{h}\right). \tag{16}$$

The ML estimate of $\mathbf{h}$ is easily found to be

$$\hat{\mathbf{h}}_{ML} = (\mathbf{B}'^+ \mathbf{R}'^{-1} \mathbf{B}')^{-1} \mathbf{B}'^+ \mathbf{R}'^{-1} \mathbf{r}'_p \tag{17}$$

and its MSE is given by

$$\mathrm{MSE} = \mathrm{trace}\left((\mathbf{B}'^+ \mathbf{R}'^{-1} \mathbf{B}')^{-1} \right). \tag{18}$$

The MSE is proportional to N_0/E_s, so there will be no error floor for $N_0/E_s \to 0$. This MSE still depends on the pilot symbols $\mathbf{b}_g$ and $\mathbf{b}_c$. The averaging of the MSE (18) over all possible pilot sequences is not straightforward, so we derive a lower bound instead. Since the inverse of a matrix is a matrix convex function, Jensen's inequality for matrices [20] can be applied: $E[(\mathbf{B}'^+ \mathbf{R}'^{-1} \mathbf{B}')^{-1}] \geq (E[\mathbf{B}'^+ \mathbf{R}'^{-1} \mathbf{B}'])^{-1}$. We assume that the pilot symbols are selected in a pseudorandom way. For the computation of $E[\mathbf{B}'^+ \mathbf{R}'^{-1} \mathbf{B}']$, we approximate $\mathbf{R}'$ by the diagonal matrix $N_0((N+\nu)/N)\mathbf{I}_{M-\nu}$ because $(\mathbf{R}')_{k,k'} \ll (\mathbf{R}')_{k,k}$, for $k' \neq k$. In that case, $E[\mathbf{B}'^+ \mathbf{R}'^{-1} \mathbf{B}']$ is essentially equal to

$$\left(E[\mathbf{B}'^+ \mathbf{R}'^{-1} \mathbf{B}'] \right)_{\ell,\ell'}$$
$$= \left(\frac{N}{N+\nu} \right)^2 \frac{E_s}{N_0} \sum_{k \in I_p} e^{j2\pi \frac{k(\ell-\ell')}{N}}$$
$$+ \left(\frac{N}{N+\nu} \right)^2 \frac{1}{N_0} \frac{1}{N} \sum_{k \in I_p} \sum_{m,m'=0}^{\nu-1} e^{j2\pi \frac{k(m-m')}{N}} R_g(m, m', \ell, \ell') \tag{19}$$

where $R_g(m, m', \ell, \ell') = E[b_g^*(|m - \ell|_\nu) b_g(|m' - \ell'|_\nu)]$. When the pilot symbols are evenly distributed over the carriers and $M - \nu$ divides N, the first term in (19) reduces to $(N/(N+\nu))^2 (E_s/N_0)(M - \nu)\delta_{\ell,\ell'}$. The computation of the second term in (19) is carried out in the Appendix. Hence, $E[\mathbf{B}'^+ \mathbf{R}'^{-1} \mathbf{B}']$ is essentially equal to $(N/(N + \nu))(E_s/N_0)(M - \nu)\mathbf{I}_L$,[1] from which it follows that the MSE lower bound is given by

$$\mathrm{MSE}_{\mathrm{LB}} = \mathrm{trace}\left((E[\mathbf{B}'^+ \mathbf{R}'^{-1} \mathbf{B}'])^{-1} \right) = \frac{N + \nu}{N} \frac{N_0}{E_s} \frac{L}{M - \nu}, \tag{20}$$

i.e., the MSE lower bound is inversely proportional to the number of pilot carriers. In Section IV we will show that the actual MSE is close to the lower bound.

The matrices $\mathbf{B}'$ and $\mathbf{R}'$ depend only on the known pilot symbols and the known positions of the data carriers and the pilot carriers. Hence, $\mathbf{B}'$ and $\mathbf{R}'$ are known at the receiver and $(\mathbf{B}'^+ \mathbf{R}'^{-1} \mathbf{B}')^{-1} \mathbf{B}'^+ \mathbf{R}'^{-1}$ can be precomputed. Therefore, the estimate (17) can be obtained with low complexity.

C. Step 2: Decision Directed Channel Estimation

The obtained channel estimate (17) is used to start up the EM-algorithm. In our case, we need to compute in the E-step (9) of the EM-algorithm, the average of the log likelihood $\log p(\mathbf{r}|\mathbf{a}_d, \hat{\mathbf{h}}_{k-1})$ over the unknown data vector $\mathbf{a}_d$ given the observation $\mathbf{r}$ and the last obtained estimate of the channel vector $\hat{\mathbf{h}}_{k-1}$. The vector of received samples $\mathbf{r}$ (8), given the channel vector $\mathbf{h}$ and the data symbols $\mathbf{a}_d$, is Gaussian distributed with mean $\mathbf{C}\mathbf{h}$ and autocorrelation matrix $N_0 \mathbf{I}_{N+\nu}$, so the log likelihood $\log p(\mathbf{r}|\mathbf{a}_d, \mathbf{h})$ is given by

$$\log p(\mathbf{r}|\mathbf{a}_d, \mathbf{h}) = -\frac{1}{N_0} (\mathbf{r} - \mathbf{C}\mathbf{h})^+ (\mathbf{r} - \mathbf{C}\mathbf{h}). \tag{21}$$

The averaging of (21) over the unknown data vector yields

$$E_{\mathbf{a}_d}\left[\log p(\mathbf{r}|\mathbf{a}_d, \hat{\mathbf{h}}_{k-1}) \right] = -\frac{1}{N_0}$$
$$\times (\mathbf{r}^+\mathbf{r} - \mathbf{r}^+\tilde{\mathbf{C}}\mathbf{h} - \mathbf{h}^+\tilde{\mathbf{C}}^+\mathbf{r} + \mathbf{h}^+\tilde{\mathbf{R}}_\mathbf{C}\mathbf{h}) \tag{22}$$

---

[1]For the special case that $\nu = L - 1$ we have shown in the Appendix that $E[\mathbf{B}'^+ \mathbf{R}'^{-1} \mathbf{B}']$ has two non-diagonal elements, but they are much smaller than the diagonal elements, so we neglect them.

where

$$\tilde{\mathbf{C}} = E_{\mathbf{a}_d}[\mathbf{C}|\mathbf{r}, \hat{\mathbf{h}}_{k-1}] \tag{23}$$

$$\tilde{\mathbf{R}}_{\mathbf{C}} = E_{\mathbf{a}_d}[\mathbf{C}^+\mathbf{C}|\mathbf{r}, \hat{\mathbf{h}}_{k-1}]. \tag{24}$$

Taking into account that $\mathbf{C} = \mathbf{B} + \mathbf{A}$, the former equations can be rewritten by

$$\tilde{\mathbf{C}} = \mathbf{B} + E_{\mathbf{a}_d}[\mathbf{A}|\mathbf{r}, \hat{\mathbf{h}}_{k-1}] \tag{25}$$

$$\tilde{\mathbf{R}}_{\mathbf{C}} = \mathbf{B}^+\mathbf{B} + E_{\mathbf{a}_d}[\mathbf{B}^+\mathbf{A} + \mathbf{A}^+\mathbf{B} + \mathbf{A}^+\mathbf{A}|\mathbf{r}, \hat{\mathbf{h}}_{k-1}]. \tag{26}$$

The posterior distribution of the data symbols $\mathbf{a}_d$ given the observation $\mathbf{r}$ and the last obtained estimate of the channel vector $\hat{\mathbf{h}}_{k-1}$, is given by

$$p(\mathbf{a}_d|\mathbf{r}, \hat{\mathbf{h}}_{k-1}) \propto p(\mathbf{r}|\mathbf{a}_d, \hat{\mathbf{h}}_{k-1})p(\mathbf{a}_d). \tag{27}$$

We assume that all data sequences are equiprobable. Using (21), where we substitute $\mathbf{h}$ by its estimate $\hat{\mathbf{h}}_{k-1}$, the posterior distribution of the data symbols (27) can be rewritten as

$$p(\mathbf{a}_d|\mathbf{r}, \hat{\mathbf{h}}_{k-1}) \propto e^{-(\mathbf{a}_d-\mathbf{m}_a)^+\mathbf{R}_a^{-1}(\mathbf{a}_d-\mathbf{m}_a)} \tag{28}$$

where

$$\mathbf{R}_a = \frac{N+\nu}{N}N_0\left(\mathbf{F}_d^+\hat{\mathbf{H}}_{k-1}^+\hat{\mathbf{H}}_{k-1}\mathbf{F}_d\right)^{-1} \tag{29}$$

$$\mathbf{m}_a = \frac{1}{N_0}\mathbf{R}_a\mathbf{F}_d^+\hat{\mathbf{H}}_{k-1}^+(\mathbf{r} - \mathbf{B}\hat{\mathbf{h}}_{k-1}). \tag{30}$$

The $(N+\nu) \times N$ matrix $\hat{\mathbf{H}}_{k-1}$ mentioned in (29), is defined as

$$(\hat{\mathbf{H}}_{k-1})_{\ell,m} = \hat{h}_{k-1}(\ell - m). \tag{31}$$

Note that $\hat{h}_{k-1}(\ell) = 0$ for $\ell < 0$ or $\ell \geq L$. Further, the matrix $\hat{\mathbf{H}}_{k-1}^+\hat{\mathbf{H}}_{k-1}$ is a Toeplitz matrix. For large N we can approximate $\hat{\mathbf{H}}_{k-1}^+\hat{\mathbf{H}}_{k-1}$ by a circulant matrix. The eigenvectors of a circulant $N \times N$ matrix are given by the columns of the $N \times N$ FFT matrix $\mathbf{F}$. The matrix $\mathbf{F}_d$ consists of a subset of those eigenvectors so $\mathbf{R}_a$ can be approximated by a diagonal matrix. This means that the data symbols $\mathbf{a}_d$, given $\mathbf{r}$ and $\hat{\mathbf{h}}_{k-1}$ can be considered as (approximately) statistically independent.

To obtain $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{R}}_{\mathbf{C}}$ we need to compute

$$E\left[a_d(m)|\mathbf{r}, \hat{\mathbf{h}}_{k-1}\right] = \sum_{a_d(m)} a_d(m)p\left(a_d(m)|\mathbf{r}, \hat{\mathbf{h}}_{k-1}\right) \tag{32}$$

and

$$E\left[a_d(m)a_d^*(m')|\mathbf{r}, \hat{\mathbf{h}}_{k-1}\right]$$
$$= \begin{cases} E_s, & m = m' \\ E\left[a_d(m)|\mathbf{r}, \hat{\mathbf{h}}_{k-1}\right]\left(E\left[a_d(m')|\mathbf{r}, \hat{\mathbf{h}}_{k-1}\right]\right)^*, & m \neq m' \end{cases} \tag{33}$$

for $m, m' = 0, \ldots, N + \nu - M - 1$. Substituting $a_d(m)$ and $a_d(m)a_d^*(m')$ by their respective expected values (32) and (33) in $\mathbf{C}$ and $\mathbf{C}^+\mathbf{C}$ yields $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{R}}_C$.

The new estimate $\hat{\mathbf{h}}_k$ obtained in the M-step can be written as a closed-form expression

$$\hat{\mathbf{h}}_k = (\tilde{\mathbf{R}}_C)^{-1}\tilde{\mathbf{C}}^+\mathbf{r}. \tag{34}$$

The algorithm terminates once the estimate has reached convergence.

To evaluate the MSE performance of the EM-based estimator, we will compare it with the all pilots case. In [16], it is shown that the

MSE of the ML estimate for the all pilots case also equals its CRB. The ML estimation in this case is based on the observation (8), where the matrix $\mathbf{C}$ is filled with pilots only and therefore *a priori* known at the receiver, i.e., the data symbols are replaced by known pilots. This results in the MSE (and CRB) given by

$$\text{MSE}_{\text{all pilots}} = \text{trace}\left(N_0(\mathbf{C}^+\mathbf{C})^{-1}\right). \tag{35}$$

Similarly, as in Section III-B, we want to obtain a lower bound for the average of the MSE (35) over all possible pilot sequences. Again, we apply Jensen's inequality for matrices [20]: $E[(\mathbf{C}^+\mathbf{C})^{-1}] \geq (E[\mathbf{C}^+\mathbf{C}])^{-1}$. We assume that the pilots are selected in a pseudorandom way. The average $E[\mathbf{C}^+\mathbf{C}]$ is given by $(E[\mathbf{C}^+\mathbf{C}])_{\ell,\ell'} = (N/N + \nu)E_s(N + \nu)\delta_{\ell,\ell'}$, from which it follows that the MSE lower bound is equal to

$$\text{MSE}_{\text{all pilots},LB} = \frac{N+\nu}{N}\frac{N_0}{E_s}\frac{L}{N+\nu} = \frac{L}{N}\frac{N_0}{E_s}. \tag{36}$$

The MSE lower bound of the all pilots case is proportional to the length L of the channel and inversely proportional to the FFT size N.

IV. NUMERICAL RESULTS

In this section, we evaluate the performance of both the frequency domain estimator and the iterative channel estimator. Without loss of generality, we assume the comb-type pilot arrangement [21] is used for the pilots transmitted on the carriers. The considered channel is a frequency selective Rayleigh fading channel consisting of $L = 8$ channel taps with equal variance. The channel impulse response is normalized: $\sum_{\ell=0}^{L-1} E[|h(\ell)|^2] = 1$. The pilot symbols are randomly generated and BPSK modulated. We assume that the pilots are equally spaced over the carriers, i.e., the positions of the pilot carriers are $I_p = \{n_0 + m\epsilon | m = 0, \ldots, M - \nu - 1\}$, where $\epsilon = \text{floor}(N/(M - \nu))$, $n_0 \in \{0, \ldots, \rho\}$ and $\rho = N - 1 - (M - \nu - 1)\epsilon$. The simulation results in this correspondence are based on the estimates (17) (for the frequency domain estimator) and (34) (for the EM-algorithm), and are averaged out over 10 000 different channel realizations and pilot symbol values.

A. Frequency Domain Estimator

First, we evaluate the performance of the frequency domain estimator, which is used in the initialization phase of the iterative estimator.

In Fig. 2, the MSE of the frequency-domain estimator is shown as function of E_s/N_0. In addition, the MSE of the subset estimator from [16] and the MSE from the estimator from [14] are shown. The estimator from [14] is slightly modified to take into account all pilot symbols, and not only the guard interval pilots. The analytical expressions for this modified estimator are given in [7, (21)]. As can be observed, the proposed frequency-domain estimator does not suffer from an error floor at high E_s/N_0, in contrast with the subset estimator and the estimator from [14]. Further, the CRB for data-aided channel estimation, derived in [16], is shown. It can be observed that the MSE of the frequency domain estimator is close to the CRB, and is inversely proportional to E_s/N_0.

To further evaluate the MSEs and the CRB from Fig. 2, we consider the normalized MSE (NMSE) and the normalized CRB (NCRB), defined as NMSE $=$ SNR $\cdot$ MSE and NCRB $=$ SNR $\cdot$ CRB, where SNR $= (N/(N + \nu))(E_s/N_0)$. From Fig. 3, it follows that at low E_s/N_0, the subset estimator slightly outperforms the frequency-domain estimator. This follows from the fact that the subset estimator uses a larger subset of observations than the frequency domain estimator (i.e., M observations for the subset estimator versus $M - \nu$ ob-
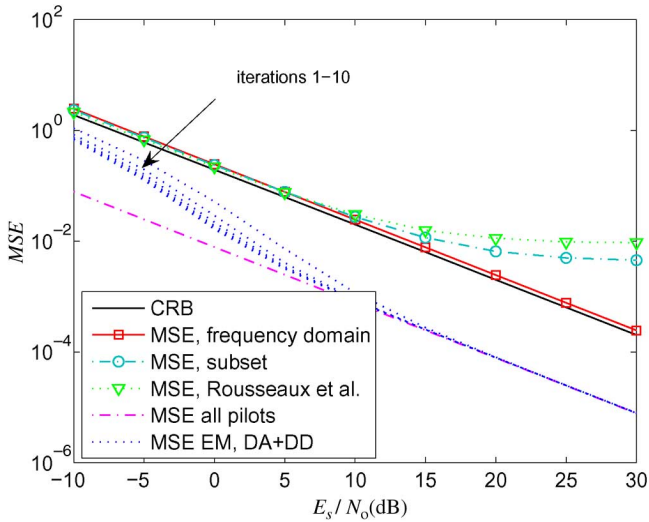
Fig. 2. MSE performance and CRB, $\nu = 7$, $N = 1024$, $M = 40$.

Fig. 4. Influence of M on the normalized MSE, $\nu = 7$, $N = 1024$.
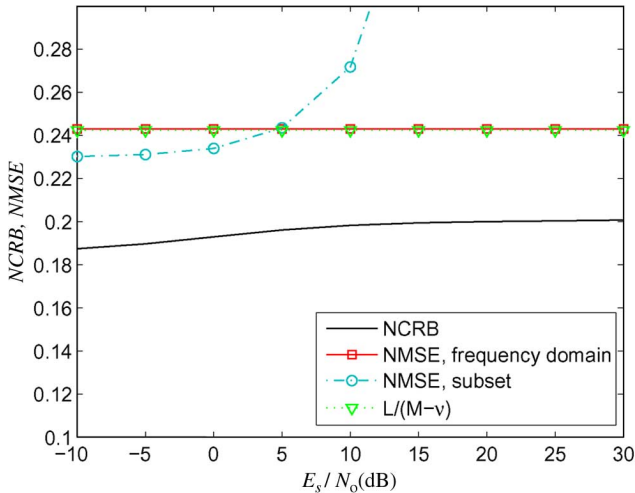

Fig. 3. Normalized MSE and CRB, $\nu = 7$, $N = 1024$, $M = 40$.

servations for the frequency domain estimator). The NMSE of the frequency domain estimator is constant with E_s/N_0, whereas the NMSE of the subset estimator strongly increases for high E_s/N_0. This can be explained as in the frequency domain estimator, the observations are data-free and therefore the data symbols have no influence on the performance of the estimator, whereas in the subset estimator, the data symbols cause an increasing amount of interference, resulting in the error floor in Fig. 2. From Fig. 3, it also can be observed that the MSE resulting from the frequency-domain estimator is close to the CRB from [16]. Further, it follows from the figure that the NMSE of the frequency-domain estimator is close to the theoretical lower bound $L/(M - \nu)$ corresponding to (20), as was shown in Section III-B. This verifies the approximations made to obtain the lower bound (20).

Fig. 4 shows the influence of the number of pilots on the MSE of the frequency-domain estimator. We first assume that the pilot carriers are equally spaced. As expected [see (20)], the MSE is essentially equal to $(L/(M - \nu))\text{SNR}^{-1}$ for a wide range of M, i.e., the MSE is inversely proportional to the number of pilot carriers. For large M, the pilot spacing becomes $\epsilon = 2$ (for $N/4 < M - \nu < N/2 = 512$) and $\epsilon = 1$ (for $M - \nu > N/2 = 512$); in that case pilots are not evenly
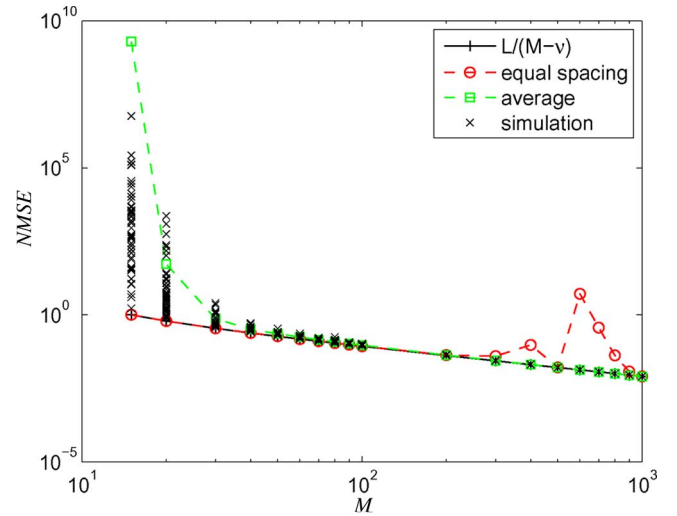
spread over the carriers but grouped in one part of the spectrum, such that (19) can no longer be approximated by $\text{SNR}(M - \nu)\delta_{\ell,\ell'}$. This causes the peaks in the figure.

The influence of random pilot carrier positions on the frequency domain estimator performance is also shown in Fig. 4. The MSE is shown for 50 randomly generated pilot carrier positions, along with the average over the simulations. For small M, we observe that the performance of the frequency domain estimator strongly depends on the pilot positions, whereas for large M, the frequency domain estimator becomes essentially independent of the pilot positions. This indicates that for small M, the second approximation (i.e., the pilots are evenly distributed) to obtain the MSE lower bound (20) is no longer valid. Hence, for small M, fixed, equally spaced pilot positions are preferred. For large M, equally spaced pilot positions are not suitable because of the peak in the MSE. Therefore, at large M, random pilot positions are advised.

B. Iterative Channel Estimator

In this correspondence, we consider the case of uncoded transmission, but the results can easily be extended to coded transmission, as long as the decoder provides the posterior probabilities of the transmitted bits given the observation (as is the case for turbo codes, LDPC codes, convolutional codes decoded using the BCJR algorithm, etc.). The EM algorithm uses the data-aided (DA) frequency domain estimator to initialize the estimation, and then iteratively updates the channel estimates in a soft decision directed (DD) way. The MSE results shown are the results obtained after convergence, unless mentioned otherwise.

Fig. 2 shows the MSE of the EM algorithm as function of E_s/N_0 for different numbers of iterations. Also shown are the MSE of the frequency domain estimator and the all pilots estimator. As mentioned in Sections III-B and III-C, the MSE of the frequency domain estimator coincides with $(L/(M - \nu))\text{SNR}^{-1}$ and the MSE of the all pilot estimator with $(L/(N + \nu))\text{SNR}^{-1}$, where $\text{SNR} = (N/(N + \nu))(E_s/N_0)$. The MSE of the EM algorithm converges to the MSE of the all pilots estimator for high E_s/N_0, while for low E_s/N_0 this is not the case. However, there is still an improvement in performance if we compare with the MSE of the frequency domain estimator. From Fig. 2, we observe that less than 10 iterations of the EM-algorithm are necessary to obtain convergence. In the rest of the simulations, we therefore use 10 iterations as a stopping criterion for the EM-algorithm.
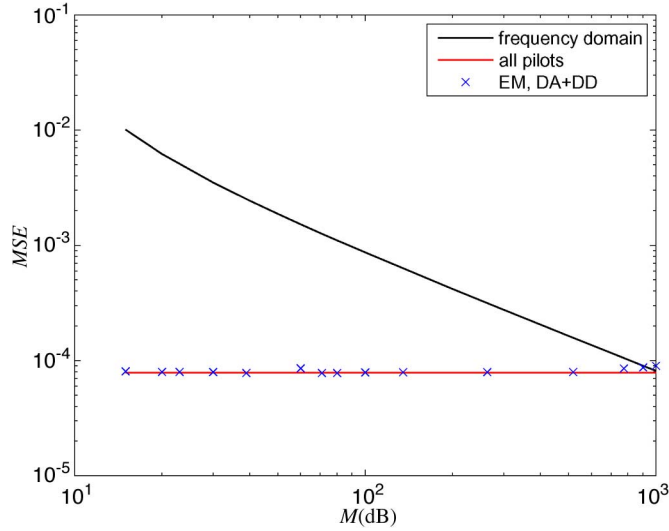
Fig. 5. Influence of M on the MSE, $\nu = 7$, $N = 1024$, $E_s/N_0 = 20$ dB.

Fig. 6. Influence of the channel estimation errors on the BER, BPSK, $N = 1024$, $\nu = 7$, $M = 40$.

Fig. 5 shows the influence of the number M of pilot symbols on the performance of the EM algorithm. Again, the analytical MSE performance of the frequency domain estimator (20) and the all pilot estimator (35) are added to the figure. As mentioned above, the MSE of the all pilot estimator coincides with $(L/(N+\nu))\text{SNR}^{-1}$ and does not depend on the number M of pilot symbols. We observe that the MSE of the EM algorithm is very close to the MSE of the all pilot estimator and is almost independent of M for a large range of M. This proves that the approximations made to obtain (35) are good approximations.

Fig. 6 shows the BER performance when using the EM algorithm and the frequency domain estimator. The transmitted data symbols consist of BPSK symbols. The performance of a receiver with perfect channel knowledge is also added. The BER of a receiver with perfect channel knowledge, averaged over the distribution of the channel was computed analytically and is given by [22]

$$\text{BER} = \frac{1}{2}\left(1 - \sqrt{\frac{\alpha}{1+\alpha}}\right) \qquad (37)$$

where $\alpha = (N/(N+\nu))^2(E_s/N_0)$. The BER curve corresponding to the EM estimator is close to the BER of the receiver with perfect channel knowledge for the considered range of E_s/N_0. The iterative algorithm results in a significantly lower BER compared to the frequency domain estimator.

V. CONCLUSIONS

In this correspondence, we have proposed an iterative channel estimation algorithm for KSP-OFDM based on the EM-algorithm using pilots that are present in the guard interval as well as on pilot carriers. To initialize the EM algorithm, we propose a DA channel estimator, that operates in the frequency domain. This estimator exploits only the $M - \nu$ FFT outputs at the pilot carrier positions. In contrast with the estimator from [14] and the subset estimator from [16], the proposed estimator does not suffer from an error floor in the MSE performance at high SNR, because the pilot carriers are not affected by the data carriers. At low SNR, the proposed estimator performs only slightly worse than the subset estimator. The MSE of this frequency domain estimator is inversely proportional to the SNR. Further, we have analytically derived a lower bound on the MSE which is inversely proportional to the number of pilot carriers and essentially proportional to $L/(M-\nu)$. The MSE of the frequency domain estimator is very close to its lower
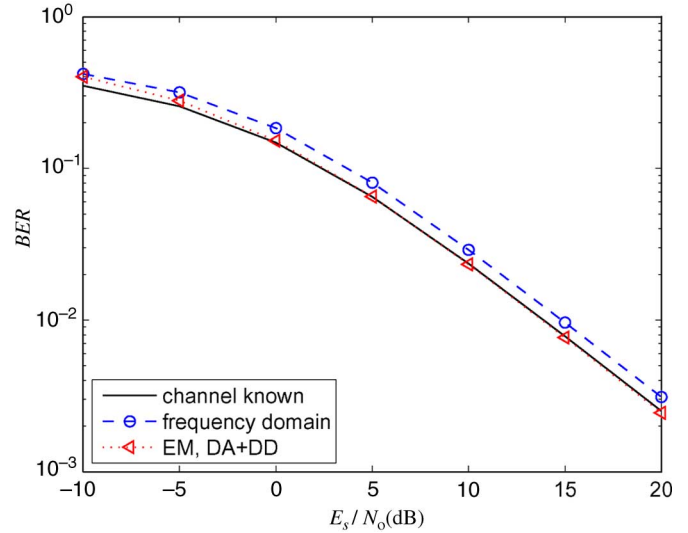
bound. At low M, the MSE strongly depends on the pilot carrier positions, whereas at high M, the performance is essentially independent of the pilot positions. Further, the MSE is essentially independent of the FFT size. However, the frequency domain estimator still causes a strong degradation of the BER.

After the initialization phase, the EM algorithm switches back to the time domain and iteratively updates the channel estimate. Every iteration, the received signal is averaged over the unknown data symbols by using the channel estimate from the previous iteration, and a new estimate of the channel is obtained. This process is repeated until the EM algorithm reaches convergence. At high SNR, the MSE of the EM algorithm coincides with the MSE of the all pilot estimator, which is proportional to L/N. Simulation results also show that the performance of the algorithm is almost independent of the number of pilot symbols for sufficiently high SNR, in contrast with the frequency domain estimator used for initialization. Further, the EM algorithm causes virtually no BER degradation as compared to the case of perfect channel knowledge.

APPENDIX

When the pilot symbols are equally spaced over the carriers, the second term of (19) can be expressed as

$$\left(\frac{N}{N+\nu}\right)^2 \frac{1}{N_0} \frac{1}{N} \sum_{m,m'=0}^{\nu-1} \frac{1 - e^{j2\pi(m-m')}}{1 - e^{j2\pi\frac{(m-m')}{M-\nu}}} R_g(m, m', \ell, \ell'). \qquad (38)$$

We assume that the total number of pilot carriers $M - \nu$ is larger than $\nu - 1$, so (38) reduces to

$$\left(\frac{N}{N+\nu}\right)^2 \frac{1}{N_0} \frac{M-\nu}{N} \sum_{m}^{\nu-1} E\left[b_g^*(|m-\ell|_\nu) b_g(|m-\ell'|_\nu)\right]. \qquad (39)$$

Now we only need to determine for which values of ℓ and ℓ', $E[b_g^*(|m - \ell|_\nu) b_g(|m - \ell'|_\nu)] \neq 0$: When $\nu > L - 1$, we only have a contribution for $\ell = \ell'$, so the second term of (19) only has a contribution for the diagonal elements. In the special case that $\nu = L - 1$, we have contributions for $\ell = \ell'$ and $(\ell, \ell') = (0, L - 1)$ and $(L - 1, 0)$. For those values of ℓ and ℓ', the second term of (19) is equal to

$$\nu\left(\frac{N}{N+\nu}\right)^2 \frac{E_s}{N_0} \frac{M-\nu}{N}. \qquad (40)$$

REFERENCES

[1] J. A. C. Bingham, "Multicarrier modulation for data transmission, an idea whose time has come," *IEEE Commun. Mag.*, vol. 28, no. 5, pp. 5–14, May 1990.

[2] *Transmission and Multiplexing (TM); Access Transmission Systems on Metallic Access Cables; Very High Speed Digital Subscriber Line (VDSL); Part 2: Transceiver Specification*, ETSI TS 101 270-2, Feb. 2001.

[3] *Digital Audio Broadcasting (DAB); DAB to Mobile Portable and Fixed Receivers*, ETS 300 401, Feb. 1995.

[4] H. Sari, G. Karam, and I. Jeanclaude, "Transmission techniques for digital terrestrial TV broadcasting," *IEEE Commun. Mag.*, vol. 33, no. 2, pp. 100–109, Feb. 1995.

[5] R. van Nee, G. Awater, M. Morikura, H. Takanashi, M. Webster, and K. W. Halford, "New high-rate wireless LAN standards," *IEEE Commun. Mag.*, vol. 37, no. 12, pp. 82–88, Dec. 1999.

[6] B. Muquet, Z. Wang, G. B. Giannakis, M. de Courville, and P. Duhamel, "Cyclic prefixing or zero padding for wireless multicarrier transmissions," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 2136–2148, Dec. 2002.

[7] H. Steendam and M. Moeneclaey, "Different guard interval techniques for OFDM: Performance comparison," in *Proc. 6th Int. Workshop Multi-Carrier Spread Spectrum (MC-SS)*, Herrsching, Germany, May 2007, pp. 11–24.

[8] T. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Trans. Commun.*, vol. 45, no. 12, pp. 1613–1621, Dec. 1997.

[9] L. Deneire, B. Gyselinckx, and M. Engels, "Training sequence versus cyclic prefix—A new look on single carrier communication," *IEEE Commun. Lett.*, vol. 5, no. 7, pp. 292–294, Jul. 2001.

[10] U. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers*. New York: Plenum, 1997.

[11] D. Van Welden, H. Steendam, and M. Moeneclaey, "Time delay estimation for KSP-OFDM systems in multipath fading channels," in *Proc. 20th Personal, Indoor, Mobile Radio Communications Symp. (PIMRC)*, Tokyo, Japan, Sep. 2009, pp. 3064–3068.

[12] E. de Carvalho and D. T. M. Slock, "Maximum-likelihood blind FIR multi-channel estimation with Gaussian prior for the symbols," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Apr. 1997, vol. 5, pp. 3593–3596.

[13] O. Rousseaux, G. Leus, P. Stoica, and M. Moonen, "Gaussian maximum-likelihood channel estimation with short training sequences," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2945–2955, Nov. 2005.

[14] O. Rousseaux, G. Leus, and M. Moonen, "Estimation and equalization of doubly selective channels using known symbol padding," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 979–990, Mar. 2006.

[15] R. Cendrillon and M. Moonen, "Efficient equalizers for single and multicarrier environments with known symbol padding," in *Proc. Int. Symp. Signal Processing Its Applications (ISSPA)*, Aug. 2001, pp. 607–610.

[16] H. Steendam, M. Moeneclaey, and H. Bruneel, "An ML-based estimate and the Cramer-Rao bound for data-aided channel estimation in KSP-OFDM," *EURASIP J. Wireless Commun. Netw.*, 2008, Article DOI 10.1155/2008/186809.

[17] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. Series B*, vol. 39, no. 1, pp. 1–38, 1997.

[18] X. Ma, H. Kobayashi, and S. C. Schwartz, "EM-based channel estimation algorithms for OFDM," *EURASIP J. Appl. Signal Process.*, vol. 2004, no. 10, pp. 1460–1477, 2004.

[19] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York: Wiley, 2001.

[20] R. Ravikanth and S. P. Meyn, "Bounds on achievable performance in the identification and adaptive control of time-varying systems," *IEEE Trans. Autom. Control*, vol. 44, no. 4, pp. 670–682, Apr. 1999.

[21] F. Tufvesson and T. Maseng, "Pilot assisted channel estimation for OFDM in mobile cellular systems," in *Proc. IEEE Veh. Tech. Conf.*, Phoenix, AZ, May 1997, pp. 1639–1643.

[22] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.

Robust Precoding With Bayesian Error Modeling for Limited Feedback MU-MISO Systems

Michael Joham, Paula M. Castro, Luis Castedo, and Wolfgang Utschick

Abstract—We consider the robust precoder design for multiuser multiple-input single-output (MU-MISO) systems where the channel state information (CSI) is fed back from the single antenna receivers to the centralized transmitter equipped with multiple antennas. We propose to compress the feedback data by projecting the channel estimates onto a vector basis, known at the receivers and the transmitter, and quantizing the resulting coefficients. The channel estimator and the basis for the rank reduction are jointly optimized by minimizing the mean-square error (MSE) between the true and the rank-reduced CSI. Expressions for the conditional mean and the conditional covariance of the channel are derived which are necessary for the robust precoder design. These expressions take into account the following sources of error: channel estimation, truncation for rank reduction, quantization, and feedback channel delay. As an example for the robust problem formulation, vector precoding (VP) is designed based on the expectation of the MSE conditioned on the fed-back CSI. Our results show that robust precoding based on fed-back CSI clearly outperforms conventional precoding designs which do not take into account the errors in the CSI.

Index Terms—Bayesian methods, channel estimation, channel state information, limited rate feedback, vector broadcast channel.

I. INTRODUCTION

We consider a MU-MISO system or vector broadcast channel (BC), i.e., a multiple antennas transmitter and several single-antenna receivers. For this setup, dirty paper coding (DPC) schemes designed according to signal-to-interference-plus-noise ratio (SINR) criteria are able to approach the sum capacity [1], [2]. Similar to [3]–[5], these contributions, however, only consider the ideal case where the CSI at the transmitter is perfectly known. In the more practical case, where only an estimate of the CSI is available at the transmitter, the capacity region of the vector BC has not been found yet. First, the application

M. Joham and W. Utschick are with the Associate Institute for Signal Processing, Technische Universität München, 80290 Munich, Germany (e-mail: joham@tum.de; utschick@tum.de).

P. M. Castro and L. Castedo are with the Department of Electronics and Systems, University of A Coruña, A Coruña, 15071 Spain (e-mail: pcastro@udc.es; luis@udc.es).

Color versions of one or more of the figures in this correspondence are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2010.2052046