ON TRANSCEIVER OPTIMIZATION FOR TIME-VARYING MULTIPATH CHANNEL ESTIMATION

Tim Whitworth, Mounir Ghogho, and Desmond C. McLernon

School of Electronic and Electrical Engineering,
The University of Leeds, LS2 9JT, UK.
(eentcw, m.ghogho, d.c.mclernon)@leeds.ac.uk

ABSTRACT

We address the problem of estimating doubly-selective channels using pilot clusters that are time-division multiplexed with the data. Channel estimation is carried out using different basis expansion models (BEMs), and direct MMSE channel estimation using the channel statistics. For a fixed number of pilot symbols, we attempt to optimize the power and placement of the pilot symbols used in transmission, and the number of BEM coefficients used in channel estimation, in the sense of minimizing the mean-square estimation error (MSE) that includes modelling error, which is normally neglected in existing work. Simulation results confirm that for a wide range of SNR and Doppler spread values, this optimization greatly reduces the MSE and the bit-error rate, and that modelling error should be taken into account when optimizing training. The effects of uncertainty in the channel statistics are also studied.

***Index Terms*—** Multipath Channels, Time-Varying Channels, Estimation, Training Design

1. INTRODUCTION

In many packet-based communications systems the discrete time complex baseband channel is modelled as an FIR filter, and the channel (filter) coefficients are assumed constant over the duration of the packet. In cases where the transmitter/receiver are moving with high speed, better modelling can be achieved by allowing the channel coefficients to randomly vary with time.

For time-invariant channels, the periodic insertion of pilot clusters was shown to be optimal [1]. For time-varying channels and zero-padded block transmission, using the complex exponential (CE) basis expansion model (BEM), it was shown in [2] that periodic insertion of zero-guarded pilot symbols was optimal, in the sense of minimizing the effect of noise. In [2], the basis exponentials were chosen to be orthogonal over the length of the block. This BEM has been termed the critically-sampled (CS)-CE model. In [3] and [4], a non-critically sampled (NCS)-CE model is used, which achieves better modelling accuracy by confining all the exponential frequencies to the range $[-\tilde{f}_{\max}, \tilde{f}_{\max}]$, where $\tilde{f}_{\max}$ is the maximum Doppler frequency. In [5], a different BEM, the discrete prolate spheroidal (DPS) sequence model, is used. This model is also band-limited, and has the advantage that the basis functions are orthogonal over the length of the block.

In this paper we focus on zero-padded block-based transmission over doubly selective channels. Rather than use the CS-CE as in [2] we use the NCS-CE approach of [4] and the DPS approach of [5], as both have better performance. We also use a BEM based on a truncation of the Karhunen-Loève (KL) expansion. We compare these BEM-based channel estimates with the direct MMSE estimation of the channel when the channel statistics are known.

The work in [4] optimizes the pilot power and placement, but only considers the variance due to the noise, and studies only frequency-flat channels. Here we consider the total MSE including channel modelling inaccuracies. The work in [5] also considers the total channel MSE, but does not carry out any pilot design optimization. Here we adjust both the power of the pilot symbols and their position within the packet in order to achieve the best possible channel estimation performance. We also allow the number of basis functions to vary (providing a tradeoff between the modelling accuracy and the ability of the receiver to estimate the unknown coefficients). We find that channel MSE and bit-error rate (BER) can be significantly reduced.

In addition to this, the case where the channel statistics are not known exactly is examined, in order to determine how best to select the training parameters in this case. Further, the case where the channel statistics are known exactly by the receiver but known roughly by the transmitter is considered.

Notation: x, $\mathbf{x}$, and $\mathbf{X}$ indicate a scalar, vector, and matrix respectively. Superscripts ^{H} and ^{T} denote Hermitian and transpose operators. $\mathbf{X}^\dagger$ is the pseudo-inverse of $\mathbf{X}$. $\mathbf{I}_N$ is the $(N \times N)$ identity matrix. $\Re(x)$ denotes the real part of x. $\mathrm{E}[x]$ is the expected value of the variable x. Finally $\mathrm{diag}(\mathbf{x})$ is a square, diagonal matrix whose diagonal is the vector $\mathbf{x}$, and $\mathrm{tr}\{.\}$ denotes the trace operator.

2. SYSTEM MODEL

The discrete-time baseband received signal for a doubly selective channel is:

$$y[n] = \sum_{\ell=0}^{N_t-1} h_\ell[n]\, s[n-\ell] + v[n], \quad n = 0, 1, \dots \quad (1)$$

where $h_\ell[n]$ is the channel response to an impulse input at time $n-\ell$, $s[n]$ is the input, and $v[n]$ is an AWGN with variance σ_v^2. The sequences $h_\ell[n]$ are assumed to be independent stationary Gaussian processes, with autocorrelation function (ACF) $\mathrm{E}\left[h_\ell[n]\, h_j^*[m]\right] = \delta(\ell - j)\sigma_\ell^2 \left[\mathbf{R}_{hh}\right]_{(n,m)}$. N_t is the length of the channel, and values of $h_\ell[n]$ for $\ell \geq N_t$ are assumed to be zero. The (normalized) Doppler-spread of the channel is also limited to the interval $[-f_{\max}, f_{\max}]$. The input, $s[n]$, is made by multiplexing pilot and data symbols, denoted by $p[n]$ and $d[n]$ respectively. The pilot symbols are known to the receiver, and the data symbols are drawn from a finite alphabet.

For block transmission with packet length N the received signal is

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{w}, \tag{2}$$

where $\mathbf{s}$ and $\mathbf{w}$ are $(N \times 1)$, and the $(N \times N)$ channel matrix $\mathbf{H}$ is lower-triangular with only the first N_t diagonals non-zero, corresponding to the length of the channel. The system uses zero-padded block transmission, in order to guard against inter-block interference.

As in [2], we will use pilot clusters where each cluster consists of a positive real symbol (called a pilot symbol) and $N_t - 1$ zeros (pilot zeros) padded on each side. In [6] this is referred to as a time-domain Kronecker delta (TDKD) pilot structure, which was shown to be optimal in terms of MMSE channel estimation for the CS-CE-BEM. This pilot structure decouples channel estimation and data detection, and this means that data recovery will be a sequential process. We will choose to end the packet with a pilot cluster, as in [2], so that the final $N_t - 1$ symbols of the packet will be pilot zeros to remove any inter-packet-interference. A packet will contain N_p pilot symbols, described by a placement vector, $\mathcal{P}$, indicating which of the input symbols is a pilot symbol, and a power vector, $\boldsymbol{\gamma} = \left[\gamma_0, \gamma_1, \ldots, \gamma_{N_p-1}\right]$, i.e: $p\left[\mathcal{P}\left[i\right]\right] = \sqrt{\gamma\left[i\right]}, \quad i \in 0 \ldots N_p - 1$. The placement vector will be ordered such that $\mathcal{P}\left[0\right] < \ldots < \mathcal{P}\left[N_p - 1\right]$ and the power vector must be non-negative. We also make the assumption that the total power available for the pilot symbols is fixed. By fixing the number of pilot symbols, we fix the maximum amount of data-rate loss due to training, and the number of non-zero data symbols will then be $N_d = N - (2N_t - 1)N_p$.

3. CHANNEL ESTIMATION

This section outlines two different methods of channel estimation, and the performance of each.

3.1. BEM Approach

The purpose of using a basis expansion model is to approximate each channel tap as the weighted sum of just a few 'basis functions', $u_k\left[n\right]$. The channel at the ℓ^{th} tap, $h_\ell\left[n\right]$, is modelled using N_f basis coefficients, $w_{\ell,k}$, i.e.

$$h_\ell\left[n\right] \approx \sum_{k=0}^{N_f-1} w_{\ell,k} u_k\left[n - \ell\right] \tag{3}$$

Using this method only $N_f N_t$ values need to be estimated.

Let $\mathbf{h}_\ell = \left[h_\ell\left[\ell\right], \ldots, h_\ell\left[N - 1 + \ell\right]\right]^T$ be the channel vector for the ℓ^{th} tap, since the values of $h_\ell\left[n\right]$ for $n < \ell$ are irrelevant in our zero-padded system, and define the k^{th} basis function as $\mathbf{u}_k = \left[u_k\left[0\right], \ldots, u_k\left[N-1\right]\right]^T$.

In the CS-CE model [2] $u_k\left[n\right] = e^{j2\pi kn/N}$, for $k \in \{-(N_f - 1)/2, \ldots, (N_f - 1)/2\}$ (for odd N_f) and therefore the sequences are orthogonal over $[0, \ldots, N-1]$. However, if the frequency response of the channel is limited to $\pm f_{\text{max}}$ then better channel modeling can be achieved by having all the exponentials lie within these limits. For that reason for the NCS-CE model $u_k\left[n\right] = e^{j2\pi f_{\text{max}} \frac{2k}{N_f-1} n}$ is used.

For the DPS model, the basis functions, $\mathbf{u}_k$, are the N_f most significant eigenvectors of a kernel matrix, as defined in [5]. The DPS model does not make use of the frequency response of the channel taps, only the Doppler spread is used. Another BEM is obtained by selecting the N_f most significant eigenvectors of the ACF, $\mathbf{R}_{hh}$, (which requires prior knowledge of the channel statistics), and we have called this the KL-BEM.

Under the condition that all pilot symbols have $N_t - 1$ leading and trailing pilot zeros, the task of estimating the channel is simplified into that of estimating N_t non-frequency selective channels (see [2]). Defining $\mathbf{U} = \left[\mathbf{u}_0, \ldots, \mathbf{u}_{N_f-1}\right]$ and $\mathbf{w}_\ell = \left[w_{(\ell,0)}, \ldots, w_{(\ell,N_f-1)}\right]^T$, the channel vectors can be written as $\mathbf{h}_\ell \approx \mathbf{U}\mathbf{w}_\ell$. The received signal can be written $\mathbf{y} = \sum_{\ell=0}^{N_t-1} \mathbf{S}_\ell \mathbf{h}_\ell + \mathbf{v} \approx \sum_{\ell=0}^{N_t-1} \mathbf{S}_\ell \mathbf{U}\mathbf{w}_\ell + \mathbf{v}$ where $\mathbf{S}_\ell$ is obtained by vertically shifting down $\text{diag}\left(\mathbf{s}\right)$ by ℓ rows. For block transmission the maximum time index is $N - 1$, whereas $\mathbf{h}_\ell$ is defined for time indices larger than this. However by defining $\mathbf{S}_\ell$ as above, which exploits the pilot zeros at the end of the packet, this problem can be ignored.

If the superscript $^{\mathcal{P}}(^{\mathcal{D}})$ indicates taking only the rows relating to the indices of the pilot (data) symbols, and $\mathbf{\Gamma} = \text{diag}\left(\boldsymbol{\gamma}\right)$, then the received values associated with the pilot symbols through the ℓ^{th} tap are:

$$\mathbf{y}^{\mathcal{P}+\ell} \approx \mathbf{S}_\ell^{\mathcal{P}+\ell} \mathbf{U}^{\mathcal{P}} \mathbf{w}_\ell + \mathbf{v}^{\mathcal{P}+\ell} = \mathbf{\Gamma}^{\frac{1}{2}} \mathbf{U}^{\mathcal{P}} \mathbf{w}_\ell + \mathbf{v}^{\mathcal{P}+\ell} \tag{4}$$

The least-squares estimate of $\mathbf{w}_\ell$ is:

$$\hat{\mathbf{w}}_\ell = (\mathbf{\Gamma}^{\frac{1}{2}} \mathbf{U}^{\mathcal{P}})^\dagger \mathbf{y}^{\mathcal{P}+\ell} \tag{5}$$

and from (5) the channel for the data symbols through the ℓ^{th} tap is estimated as

$$\hat{\mathbf{h}}_\ell^{\mathcal{D}} = \mathbf{U}^{\mathcal{D}} (\mathbf{\Gamma}^{\frac{1}{2}} \mathbf{U}^{\mathcal{P}})^\dagger \mathbf{y}^{\mathcal{P}+\ell} \tag{6}$$

If $\sigma_\ell^2 \mathbf{R}_{hh}$ is the $N \times N$ autocorrelation matrix of the ℓ^{th} channel tap, and if $\mathbf{R}_{h^{\mathcal{P}} h^{\mathcal{D}}}$ means taking the rows relating to $\mathcal{P}$ and the columns relating to $\mathcal{D}$, then the MSE, averaged over the data, can be simplified to:

$$\begin{aligned} \text{MSE}_{\hat{h}} &= \frac{1}{N_d} \text{tr}\left\{ \mathbf{\Phi} \mathbf{R}_{h^{\mathcal{P}} h^{\mathcal{P}}} \mathbf{\Phi}^H + \mathbf{R}_{h^{\mathcal{D}} h^{\mathcal{D}}} \right. \\ &\quad \left. - 2\Re(\mathbf{\Phi} \mathbf{R}_{h^{\mathcal{P}} h^{\mathcal{D}}}) + N_t \mathbf{\Psi} \sigma_v^2 \right\} \qquad (7) \\ \text{where} \quad \mathbf{\Phi} &= \mathbf{U}^{\mathcal{D}} (\mathbf{U}^{\mathcal{P}H} \mathbf{\Gamma} \mathbf{U}^{\mathcal{P}})^{-1} \mathbf{U}^{\mathcal{P}H} \mathbf{\Gamma}, \\ \mathbf{\Psi} &= \mathbf{U}^{\mathcal{D}} (\mathbf{U}^{\mathcal{P}H} \mathbf{\Gamma} \mathbf{U}^{\mathcal{P}})^{-1} \mathbf{U}^{\mathcal{D}H}, \end{aligned}$$

Without loss of generality the average total channel energy has been fixed to unity. Eq. (7) is valid for any basis functions which can be written in the form of (3), and for any channel autocorrelation function. If only noise is considered the MSE is given by the last term of in RHS of (7).

3.2. MMSE Approach

To avoid the modelling errors of the BEM-based methods, the MMSE estimate of $\mathbf{h}_\ell$ can be obtained directly as:

$$\tilde{\mathbf{h}}_\ell^{\mathcal{D}} = \mathbf{R}_{h^{\mathcal{D}} h^{\mathcal{P}}} \left(\mathbf{R}_{h^{\mathcal{P}} h^{\mathcal{P}}} + \mathbf{\Gamma}^{-1} \frac{\sigma_v^2}{\sigma_\ell^2} \right)^{-1} \mathbf{\Gamma}^{-\frac{1}{2}} \mathbf{y}^{\mathcal{P}+\ell} \tag{8}$$

Whilst this would be expected to outperform BEM method, it requires more knowledge about the channel, and it has higher complexity. The MSE of the direct MMSE estimate in (8) is given by:

$$\begin{aligned} \text{MSE}_{\tilde{h}} &= \frac{1}{N_d} \sum_{\ell=0}^{N_t-1} \sigma_\ell^2 \text{tr}\left\{ \mathbf{R}_{h^{\mathcal{D}} h^{\mathcal{D}}} - \mathbf{R}_{h^{\mathcal{D}} h^{\mathcal{P}}} (\mathbf{R}_{h^{\mathcal{P}} h^{\mathcal{P}}} \right. \\ &\quad \left. + \mathbf{\Gamma}^{-1} \sigma_v^2 / \sigma_\ell^2)^{-1} \mathbf{R}_{h^{\mathcal{P}} h^{\mathcal{D}}} \right\} \end{aligned} \tag{9}$$

4. TRAINING DESIGN

We place the pilot power and placement vectors under the constraints:
(C1) $\sum_i \gamma_i = N_p$
(C2) $0 \leq \gamma_i \leq N_p$
(C3) $\mathcal{P}[N_p - 1] = N - N_t$, i.e. the packet ends with a pilot symbol followed by $N_t - 1$ zeros.
(C4) $\mathcal{P}[i-1] + 2(N_t - 1) < \mathcal{P}[i] < \mathcal{P}[i+1] - 2(N_t - 1)$, i.e. pilot symbols are also separated by pilot zeros.

The average pilot symbol power is set to unity (due to **(C1)**) without loss of generality.

The problem of joint optimization can be written as

$$\{N_f, \boldsymbol{\gamma}, \mathcal{P}\} = \arg \min_{N_f, \boldsymbol{\gamma}, \mathcal{P}} \text{MSE} \tag{10}$$

where MSE is the value computed in (7) or (9). Prior to optimization N_f is chosen according to (as in [2], [5]):

$$N_f = 2 \lfloor f_{\text{max}} N \rfloor + 1, \tag{11}$$

In the optimization we allow the number of basis functions, N_f, used to approximate the channel to vary from $1, \ldots, N_p$.

In the case where $N_f = N_p$, the joint optimization of (7) can be simplified by noting that the matrix $\mathbf{U}^{\mathcal{P}}$ is square, and in this case the MSE can be simplified, since in (7) $\mathbf{\Phi}$ is no longer a function of $\boldsymbol{\gamma}$. The power vector can be solved analytically to give

$$\begin{aligned} \gamma_i &= \frac{N_p \sqrt{[\mathbf{A}]_{i,i}}}{\sum_{j=1}^{N_p} \sqrt{[\mathbf{A}]_{j,j}}} \\ \text{where} \quad \mathbf{A} &= (\mathbf{U}^{\mathcal{P}})^{-H} \mathbf{U}^{\mathcal{D}H} \mathbf{U}^{\mathcal{D}} (\mathbf{U}^{\mathcal{P}})^{-1} \end{aligned} \tag{12}$$

When $1 \leq N_f < N_p$ the MSE expression in (7) cannot be simplified. In this case, and for the MMSE method, a gradient descent method can be used in place of (12) (making sure that **(C1)** and **(C2)** are not broken). This cannot guarantee a global minimum since (7) has not been shown to be convex with respect to $\mathcal{P}$ when $N_f < N_p$.

Our simulations used a brute-force method to find the optimal value of $\mathcal{P}$ out of all the placement vectors satisfying **(C3)** and **(C4)**. The complexity of doing so is high however, and may be prohibitive for very large values of N. For comparison we also test the case where $(N_f, \mathcal{P})$ and $\boldsymbol{\gamma}$ are optimized iteratively, and where the search over $\mathcal{P}$ is not done exhaustively. One method along these lines was presented in [3], however here we have chosen to update $\mathcal{P}[i]$ by testing the MSE of $\mathcal{P}_{\text{new}}[i] = \mathcal{P}_{\text{old}}[i] \pm 1$, $i \in 0, \ldots, N_p - 1$ and keeping changes where the MSE is reduced. Further we fixed $\mathcal{P}[0] = N_t - 1$. For the parameters used for the simulations in Section 6 the number of pilot positions considered is around $1/1000^{\text{th}}$ of the total possible, and as the pilot power is only optimized for very few of these cases the total complexity was smaller still.

5. CHANNEL STATISTICS MISMATCH

The channel MSE value in (7) can still be used, by noting that $\mathcal{P}$ and $\boldsymbol{\gamma}$ are functions of the transmitter's estimates of the channel parameters, $\mathbf{U}$ (itself a function of f_{max} and N_f) is a function of the receiver's estimates of the channel parameters, and $\mathbf{R}_{hh}$ is the true channel ACF.

The channel MSE value given in (9) must be replaced by:

$$\begin{aligned} \text{MSE}_{\tilde{h}} &= \frac{1}{N_d} \sum_{\ell=0}^{N_t-1} \sigma_\ell^2 \text{tr} \left\{ \mathbf{\Xi}_\ell \mathbf{R}_{h^{\mathcal{P}} h^{\mathcal{P}}} \mathbf{\Xi}_\ell^H + \mathbf{R}_{h^{\mathcal{D}} h^{\mathcal{D}}} \right. \\ &\quad \left. - 2\Re(\mathbf{\Xi}_\ell \mathbf{R}_{h^{\mathcal{P}} h^{\mathcal{D}}}) + \mathbf{\Xi}_\ell \mathbf{\Gamma}^{-1} \mathbf{\Xi}_\ell^H \sigma_v^2 / \sigma_\ell^2 \right\} \\ \text{where } \mathbf{\Xi}_\ell &= \bar{\mathbf{R}}_{h^{\mathcal{D}} h^{\mathcal{P}}} \left(\bar{\mathbf{R}}_{h^{\mathcal{P}} h^{\mathcal{P}}} + \mathbf{\Gamma}^{-1} \sigma_v^2 / \sigma_\ell^2 \right)^{-1} \end{aligned} \tag{13}$$

and where $\bar{\mathbf{R}}_{hh}$ is the receiver's estimate of the channel ACF.

6. SIMULATION RESULTS

For all our simulations the channel autocorrelation function is assumed to be given by the widely accepted Jakes' model:

$$R_{h_\ell}(\tau) = \sigma_\ell^2 \text{J}_0 (2\pi f_{\text{max}} \tau), \quad \tau = 0, \pm 1, \ldots \tag{14}$$

where $\text{J}_0(.)$ is a zeroth order Bessel function. The channel taps follow an exponential delay profile, $\sigma_\ell^2 \propto e^{-\alpha \ell}$ where we have set $\alpha = 0.2$. The SNR (for a unit energy channel) has been defined as the average power of the pilot clusters divided by the noise power, i.e. $\text{SNR} = 1/((2N_t - 1)\sigma_v^2)$. For the training optimization simulations the system parameters were $N = 100$, $N_t = 4$, and $N_p = 4$. The BER was tested for a QPSK constellation over 20000 trials, using ML detection, where the instantaneous data power was equal to the average power of the pilot clusters (not necessarily the optimum ratio).

Fig 1 shows the results of different levels of optimization for the DPS-BEM. For comparison Fig 1 also shows the case where optimization only uses the noise (i.e. modelling error is not considered), and also the low complexity suboptimal case from Section 4. In the case where N_f is unoptimized the value is calculated using (11), giving $N_f = 3$. From this we see that optimizing the power provides little improvement to the MSE, indeed the difference between the joint optimization of $(N_f, \mathcal{P})$ and $(N_f, \mathcal{P}, \boldsymbol{\gamma})$ is negligible. However this does not mean that the values of $\mathcal{P}$ are the same, or even similar, as we will discuss later. Furthermore it is essential, especially at high SNR, to optimize N_f, rather than use (11). The low-complexity approximation performs very well, which would be useful when optimizing the parameters for larger values of N and N_p. Finally, the performance where only noise is considered in optimization is *far worse* than the optimal solution. It is also worse than the case where only $(\mathcal{P}, \boldsymbol{\gamma})$ are optimized, but using the modelling error in addition to the noise.

Comparing the performance of the different channel estimation methods (not shown due to space), we found that although the difference in performance is large prior to optimization, afterwards the difference is small. As expected the direct MMSE method has the lowest MSE at all SNR values, the KL-BEM very slightly outperforms the DPS-BEM, and the NCS-CE-BEM has the highest MSE.

Fig 2 shows the training parameters for the DPS-BEM after optimization. It is interesting to note that not all of the pilot blocks are used, i.e. some of the pilot values are set to zero. If those pilots which are set to zero are at the start/end of the packet, then they can essentially be ignored, and in this case it would improve the data rate by up to 16%. This is a large benefit of optimizing the pilot power, even though the MSE performance is roughly the same.

Figure 3 shows the performance when tx/rx mismatched f_{max}. Figure 4 shows the performance when only the transmitter has mismatched f_{max}. When f_{max} is overestimated at the transmitter, but known at the receiver, the channel can still be estimated with almost optimal MSE, but when $f_{\max}$ is under-estimated the receiver cannot rectify this. This is due to the fact that some of the pilot symbols have zero magnitude (see Fig 2), and they are not well-spread, and therefore should the receiver tries to choose a value of N_f larger than the transmitter anticipates the equation may be badly-scaled or under-determined. This would suggest that where the transmitter does not have accurate knowledge of the channel statistics it selects γ and $\mathcal{P}$ with value of N_f which is larger than optimal. With mismatched SNR (not shown due to space), as with $f_{\max}$, the BEMs are very resilient provided N_f is not chosen to be too small. The direct MMSE estimate is more sensitive, and requires a good estimate of σ_v^2.

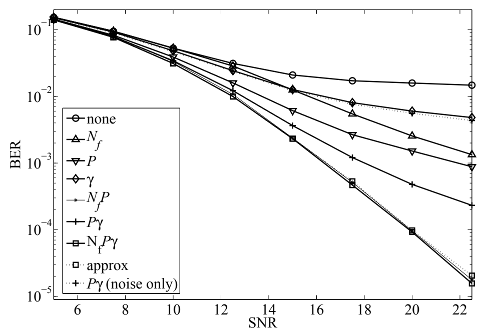

Fig. 1. The performance using the DPS-BEM estimate with different parameters optimized ($f_{\max} = 0.01$).

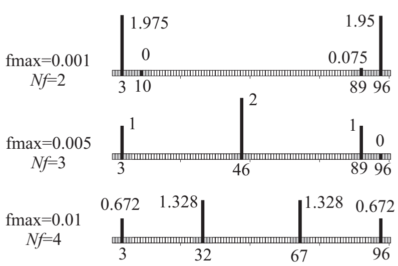

Fig. 2. The channel training parameters after optimization, for DPS-BEM (SNR $= 15$)

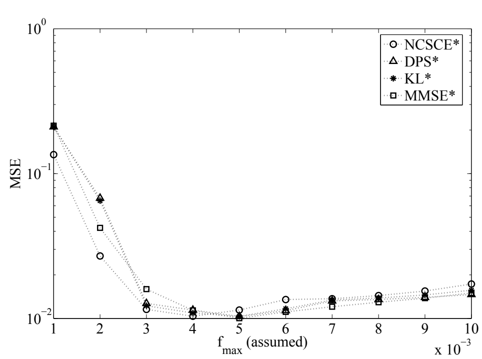

Fig. 3. The performance of different estimates with freq. mismatch at tx/rx ($f_{\max} = 0.005$, SNR=15dB)

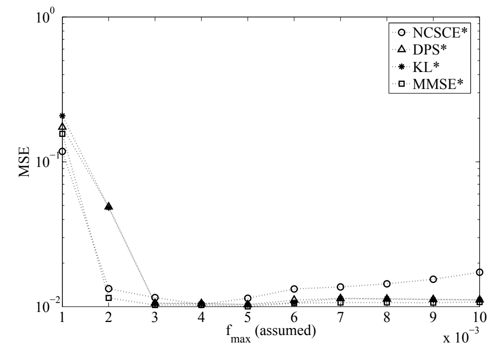

Fig. 4. The performance of different estimates with freq. mismatch at tx ($f_{\max} = 0.005$, SNR=15dB)

7. CONCLUSIONS

When using BEMs for time varying channels we have shown how important it is to properly set the power and placement of the pilot symbols in the packet. The channel MSE performance can be improved significantly by optimizing the system parameters, and this could lead to several benefits, such as improved BER (as we have shown) or a reduction in required transmitter power.

Taking into account modelling error in training design provides significant MSE/BER improvement compared to training design based on minimizing the noise effects only.

In terms of channel statistics mismatch the results suggest that when allocating pilot power and placement the transmitter should select a higher value of N_f than optimal, in order to give the receiver the flexibility to compensate for the transmitter's lack of channel knowledge. When the channel statistics are varying slowly the receiver will not have to waste much bandwidth sending information about the channel back to the transmitter. The receiver still requires a good estimate of $f_{\max}$.

The KL-BEM performs similarly to the DPS-BEM, with and without channel statistics mismatch, and so either could be used, however when the receiver has good knowledge of the channel statistics the direct MMSE estimation provides better MSE performance.

8. REFERENCES

[1] A. Adireddy, L. Tong and H. Viswanathan, "Optimal placement of training for frequency-selective block-fading channels," *Trans. on Info. Theory* 48(8), Aug. 2002.

[2] X. Ma, G. Giannakis, "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE Trans. on Sig. Proc.*, vol. 51 no. 5, May. 2003 pp. 1351–66.

[3] G. Leus, "On the estimation of rapidly time-varying channels," *Proc. EUSIPCO*, Sept. 2004.

[4] Z. Tang, G. Leus, "Time-Multiplexed Training for Time-Selective Channels," *to appear in IEEE Signal Processing Letters*.

[5] T. Zemen and C. F. Mecklenbräuker, "Time-variant channel estimation using discrete proloate spheroidal sequences," *Trans. on Sig. Proc.*, vol. 53 no. 9, Sept. 2005 pp. 3597–3607.

[6] A. P. Kannu and P. Schniter, "Desing and Analysis of MMSE Pilot-Aided Cyclic-Prefixed Block Transmissions for Doubly Selective Channels," *To be published*