Superimposed Pilot Symbols for Channel Estimation in OFDM Systems

Tao Cui and Chintha Tellambura
Department of Electrical and Computer Engineering
University of Alberta
Edmonton, AB, Canada T6G 2V4
Email: {taocui, chintha}@ece.ualberta.ca

***Abstract*— We consider superimposing pilot symbols on to data symbols for channel estimation for Orthogonal Frequency Division Multiplexing (OFDM) systems. We first derive maximum-likelihood (ML) and minimum-mean square error (MMSE) iterative channel estimators. Modeling the time domain signal as Gaussian, we derive an ML channel estimator by averaging the likelihood function for both data and channel impulse response (CIR) over the resulting Gaussian vector. Two data detectors are also proposed by eliminating the CIR from the likelihood function. The resulting integer least squares problem can be efficiently solved using a sphere decoder (SD). Furthermore, the Cramer-Rao bound (CRB) for the superimposed channel and data estimation is derived. The equispaced pilot placement is optimal in superimposed training. The ideal performance benchmarks are reached by our proposed estimators. Their performance is comparable to that of a separated training scheme, but they offer a higher data rate.**

I. Introduction

Orthogonal Frequency Division Multiplexing (OFDM) provides high spectral efficiency, but the channel impulse response (CIR) must be estimated accurately for coherent detection. Pilot-based channel estimation for OFDM has thus been widely studied [1], where pilots and data symbols are placed in separate subcarriers by periodic insertion of pilot symbols. The receiver estimates the channel at the pilot subcarriers first, and these estimates are interpolated to estimate the channel at the data subcarriers. In mobile radio environments, the time-varying channel requires closely-spaced pilot symbols, resulting in a significant bandwidth loss. Semi-blind and blind equalization and channel estimation methods, however, need several OFDM symbols for channel estimation and exhibit both high complexity and phase ambiguities.

On the other hand, pilot symbols can be added to data symbols to enable CIR estimation without sacrificing the data rate. This idea was first proposed for analog communication in [2] and was later extended to digital single carrier systems in [3]. Recently, the idea of superimposed training has received renewed attention [4]–[6]. In [4], a two-dimensional Wiener filter is employed to obtain the initial frequency domain channel estimate using second order statistics. The initial CIR estimate is then used to recover data symbols. The resulting data are used to iteratively update the channel estimation with a new Wiener filter. However, this scheme is sensitive to the signal-to-pilot power ratio (SPR). In [5], periodic pilots are added to data symbols in time domain before transmission, and first order statics are exploited to identify the CIR. As adding pilots can increase the peak-to-average power ratio (PAPR), superimposed pilots must be carefully chosen to mitigate this problem. In [6], the data vector is distorted so that its discrete fourier transform (DFT) at the pilot frequencies is zero, which cancels the performance degradation by the embedded unknown data.

In this paper, we derive channel estimators and data detectors for superimposed training in both the time and frequency domains. We propose channel estimators based on iterative maximum likelihood (ML) and minimum-mean square error (MMSE). ML and MMSE algorithms are used to obtain the initial estimate of the time domain CIR. The data and CIR estimates are then updated using a decision-directed algorithm. The iterative estimators perform well in low SPR. Taking the time domain OFDM signals as complex Gaussian, we derive an ML channel estimator by averaging the data-CIR likelihood function over the resulting Gaussian vector. The resulting ML channel estimator is solved using gradient based algorithms. On the other hand, two data detectors are also proposed by eliminating the CIR from the likelihood function. The resulting integer least squares problem can be efficiently solved with a sphere decoder (SD) [7]. We also derive the Cramer-Rao bound (CRB) for the superimposed channel and data estimation. Optimal pilot placement is obtained by minimizing the CRB. We find that the equispace condition still holds in superimposed training.

Notation: $(\cdot)^T$ and $(\cdot)^H$ denote transpose and conjugate transpose. The set of complex K-dimensional vectors is denoted by $\mathcal{C}^K$. A signal constellation and its cardinality are denoted by $\mathcal{Q}$ and $|\mathcal{Q}|$. The $N \times N$ DFT matrix is $\mathbf{F} = 1/\sqrt{N}[e^{j\frac{2\pi}{N}kl}]$, $k, l \in 0, 1, \cdots, N-1$, $j = \sqrt{-1}$. A matrix $\mathbf{A}_D$ denotes a diagonal matrix whose diagonal entries are the components of vector $\mathbf{A}$.

II. System Model

We consider the discrete-time equivalent baseband model of an OFDM system over frequency-selective channels. Data are mapped into a finite constellation $\mathcal{Q}$. We consider a generalized training strategy in which the transmitted symbol X_k at the k-th subcarrier is a linear combination of a pilot symbol and a data symbol

$$X_k = \sqrt{\phi_k}p_k + \sqrt{\psi_k}s_k,\ k = 0, 1, \ldots, N-1 \quad (1)$$

where $p_k \in \mathcal{Q}$ is the known pilot, $s_k \in \mathcal{Q}$ is a zero-mean randomly distributed data symbol, and both p_k and s_k have the unity average power. The coefficients ϕ_k and ψ_k specify the power of the pilot and data symbols, respectively. The SPR for the k-th subcarrier is defined as $\text{SPR}_k = \phi_k/\psi_k$. The power $\mathcal{E}_k = \phi_k + \psi_k$ is the total power for the k-th subcarrier and $\mathcal{E} = \sum_{k=0}^{N-1} \mathcal{E}_k$ is the total power for an OFDM symbol. If $\psi_k = 0$, for $k \in I_p$, (1) reduces to the separated training scheme in [1], where I_p denotes the index set of N_p pilot subcarriers. Transmit symbols X_k's are modulated by an inverse DFT (IDFT), and the resulting time domain signal samples are

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{\jmath 2\pi kn/N}, \; n = 0, \cdots, N-1. \quad (2)$$

Note after IDFT, pilots and data symbols are superimposed in both time domain and frequency domain. A guard interval, inserted to prevent inter-frame interference, includes a cyclic prefix of $\{x(N_g-1), \cdots, x(N-1)\}$ where N_g is the number of samples in the guard interval. These samples are appropriately pulse shaped to construct the time domain signal $x(t)$ for transmission.

The composite response including transmit and receive pulse shaping and the physical channel response between the transmitter and receiver may be modeled as

$$h(t) = \sum_{l=0}^{L-1} h_l(t)\delta(t - \tau_l) \quad (3)$$

where $h_l(t) \sim \mathcal{CN}(0, \sigma_l^2)$, and τ_l is the delay of the l-th tap. Typically, it is assumed that $\tau_l = lT_s$, and this results in a finite impulse response filter with an effective length L. We consider that the channel taps $h_l(t)$ remain constant in each block so that inter-carrier-interference (ICI) is negligible. Assuming perfect synchronization, the received signal after sampling can be represented as

$$\begin{aligned} y_n &= \sum_{l=0}^{L-1} h_l x_{n-l} + w_n \\ &= \frac{1}{\sqrt{N}} \sum_{l=0}^{L-1} h_l \sum_{k=0}^{N-1} X_k e^{\jmath 2\pi k(n-l)/N} + w_n \end{aligned} \quad (4)$$

where w_n is Additive White Gaussian Noise (AWGN). After removing the guard interval and performing DFT demodulation, we can get

$$\begin{aligned} Y_k &= \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} y_n e^{-\jmath 2\pi kn/N} \\ &= X_k H_k + W_k \end{aligned} \quad (5)$$

where $H_k = 1/N \sum_{l=0}^{L-1} h_l e^{-\jmath 2\pi kl/N}$, and $W_k = 1/\sqrt{N} \sum_{n=0}^{N-1} w_n e^{-\jmath 2\pi nk/N}$ with mean zero and variance σ_n^2. We find $\mathbf{H} = [H_0, \ldots, H_{N-1}]^T = \mathbf{F}_L \mathbf{h}$, where $\mathbf{h} = [h_0, \ldots, h_{L-1}]^T$ is the CIR and $\mathbf{F}_L$ is a $N \times L$ submatrix of the DFT matrix $\mathbf{F}$. We can vectorize (5) as

$$\mathbf{Y} = \mathbf{X}_D \mathbf{F}_L \mathbf{h} + \mathbf{W} = (\mathbf{\Phi P}_D + \mathbf{\Psi S}_D) \mathbf{F}_L \mathbf{h} + \mathbf{W} \quad (6)$$

or equivalently

$$\mathbf{Y} = \mathbf{H}_D \mathbf{X} + \mathbf{W} = \mathbf{H}_D (\mathbf{\Phi P} + \mathbf{\Psi S}) = \mathbf{H}_D \mathbf{F} \mathbf{x} + \mathbf{W} \quad (7)$$

where $\mathbf{X} = [X_0, \cdots, X_{N-1}]^T$, $\mathbf{x} = [x_0, \cdots, x_{N-1}]^T$, $\mathbf{Y} = [Y_0, \ldots, Y_{N-1}]^T$, $\mathbf{P} = [p_0, \ldots, p_{N-1}]^T$, $\mathbf{S} = [s_0, \ldots, s_{N-1}]^T$, $\mathbf{\Phi} = \text{diag}\{\sqrt{\phi_0}, \cdots, \sqrt{\phi_{N-1}}\}$ and $\mathbf{\Psi} = \text{diag}\{\sqrt{\psi_0}, \cdots, \sqrt{\psi_{N-1}}\}$.

III. Iterative Channel Estimators

A. Iterative ML estimator

Under the assumption of AWGN, and using (6), the probability density function (pdf) of received signal $\mathbf{Y}$ conditioned on $\mathbf{h}$ and $\mathbf{S}$ is

$$f(\mathbf{Y}|\mathbf{h}, \mathbf{S}) = \frac{1}{(\pi\sigma_n^2)^N} \exp\left\{-\frac{1}{\sigma_n^2} \|\mathbf{Y} - \mathbf{X}_D \mathbf{F}_L \mathbf{h}\|^2\right\}. \quad (8)$$

Its dependence on $\mathbf{S}$ can be removed by summing $f(\mathbf{Y}|\mathbf{h}, \mathbf{S})$ over all $\mathbf{S}$ and obtaining the marginal probability $f(\mathbf{Y}|\mathbf{h}) = \sum_{\mathbf{S} \in \mathcal{Q}^N} f(\mathbf{S}) f(\mathbf{Y}|\mathbf{h}, \mathbf{S})$. The CIR can thus be estimated maximizing $f(\mathbf{Y}|\mathbf{h})$. However, this approach becomes intractable when N is large, and it is difficult to find the global maximum of $f(\mathbf{Y}|\mathbf{h})$. Instead, we use a suboptimal approach by considering the approximation $e^x = 1 + x$ when x is small. We write $f(\mathbf{Y}|\mathbf{h})$ as

$$f(\mathbf{Y}|\mathbf{h}) \approx \sum_{\mathbf{S} \in \mathcal{Q}^N} \frac{1}{(\pi\sigma_n^2|\mathcal{Q}|)^N} \left\{1 - \frac{1}{\sigma_n^2} \|\mathbf{Y} - \mathbf{X}_D \mathbf{F}_L \mathbf{h}\|^2\right\}. \quad (9)$$

Therefore, the initial CIR estimate can be obtained by minimizing

$$g(\mathbf{h}) = \sum_{\mathbf{S} \in \mathcal{Q}^N} \|\mathbf{Y} - \mathbf{X}_D \mathbf{F}_L \mathbf{h}\|^2. \quad (10)$$

Taking the partial derivative of (10) with respect to $\mathbf{h}$ and setting the result to zero, we find

$$\hat{\mathbf{h}}^0 = \left(\mathbf{F}_L^H \mathbf{\Lambda} \mathbf{F}_L\right)^{-1} \mathbf{F}_L^H \mathbf{P}_D^H \mathbf{\Phi} \mathbf{Y} \quad (11)$$

where $\mathbf{\Lambda} = \text{diag}\{\mathcal{E}_0, \ldots, \mathcal{E}_{N-1}\}$. We use $E\{s_k\} = 0$ in (11).

Starting from $\hat{\mathbf{h}}^0$, a decision directed (DD) technique can be used to improve the performance of both channel estimation and data detection. From (8), the joint CIR and data estimator is given by

$$\{\mathbf{h}, \mathbf{S}\} = \underset{\mathbf{h} \in \mathcal{C}^L, \mathbf{S} \in \mathcal{Q}^N}{\arg\min} \|\mathbf{Y} - (\mathbf{\Phi P}_D + \mathbf{\Psi S}_D) \mathbf{F}_L \mathbf{h}\|^2. \quad (12)$$

In the i-th iteration, data symbols $\mathbf{S}$ can be estimated via

$$\hat{\mathbf{S}}^i = \mathcal{M}_{\mathcal{Q}}\left(\mathbf{\Psi}^{-1}\left[\left(\mathbf{H}_D^{i-1}\right)^{-1} \mathbf{Y} - \mathbf{\Phi P}\right]\right) \quad (13)$$

where $\mathbf{H}_D^{i-1} = \text{diag}\{\mathbf{F}_L \hat{\mathbf{h}}^{i-1}\}$, and $\mathcal{M}_{\mathcal{Q}}(\cdot)$ quantizes $(\cdot)$ to the nearest element in $\mathcal{Q}$. The CIR estimation follows as

$$\hat{\mathbf{h}}^i = \left(\mathbf{F}_L^H \mathbf{X}_D^H \mathbf{X}_D \mathbf{F}_L\right)^{-1} \mathbf{F}_L^H \mathbf{X}_D^H \mathbf{Y} \quad (14)$$

where $\mathbf{X}_D = \mathbf{\Phi P}_D + \mathbf{\Psi S}_D^i$.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2005 proceedings.

B. Iterative MMSE estimator

In the initial estimation stage, $\hat{\mathbf{S}}$ is not available. We then assume the term $\mathbf{V} = \mathbf{\Psi}\mathbf{S}_D\mathbf{F}_L\mathbf{h}$ in (6) to be random interference and obtain

$$\mathbf{R}_{hY} = E\{\mathbf{h}\mathbf{Y}^H\} = \mathbf{R}_h\mathbf{F}_L^H\mathbf{P}_D^H\mathbf{\Phi}^H \tag{15}$$

and

$$\mathbf{R}_{YY} = E\{\mathbf{Y}\mathbf{Y}^H\} = \mathbf{\Phi}\mathbf{P}_D\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H\mathbf{P}_D^H\mathbf{\Phi}^H + \mu\mathbf{\Psi}^2 + \sigma_n^2\mathbf{I}_N \tag{16}$$

where $\mathbf{R}_h = E\{\mathbf{h}\mathbf{h}^H\}$ and $\mu = \sum_{l=0}^{L-1}\sigma_l^2/N$. The linear MMSE channel estimation is therefore given by

$$\hat{\mathbf{h}}^0 = \mathbf{R}_{hY}\mathbf{R}_{YY}^{-1}\mathbf{Y}. \tag{17}$$

Simulation results show that the initial channel estimate by (17) has a smaller error floor than that of (11). The iterative stage is similar to the ML channel estimator. Given $\hat{\mathbf{h}}^{i-1}$, $\hat{\mathbf{S}}^i$ can be estimated from (13). Instead of using (14), the CIR estimate can be MMSE updated as

$$\hat{\mathbf{h}}^i = \mathbf{R}_h\mathbf{F}_L^H\mathbf{X}_D^H\left(\mathbf{X}_D\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H\mathbf{X}_D^H + \sigma_n^2\mathbf{I}_N\right)^{-1}\mathbf{Y} \tag{18}$$

where $\mathbf{X}_D = \mathbf{\Phi}\mathbf{P}_D + \mathbf{\Psi}\mathbf{S}_D^i$. Since both of the iterative detectors depend on the quality of the initial channel estimation, they perform well in high SPR. Note that both complexity and performance increase for more iterations, but the performance improvement rate diminishes.

IV. Maximum Likelihood Channel Estimator

Using (6), computing $f(\mathbf{Y}|\mathbf{h})$ is intractable by summing $f(\mathbf{Y}|\mathbf{h},\mathbf{S})$ over all $\mathbf{S}$. In this section, we take a different approach by using (7). The likelihood function conditional on $\mathbf{x}$ and $\mathbf{H}_D$ is given by

$$f(\mathbf{Y}|\mathbf{x},\mathbf{H}_D) = \frac{1}{(\pi\sigma_n^2)^N}\exp\left\{-\frac{1}{\sigma_n^2}\left\|\mathbf{Y}-\mathbf{H}_D\mathbf{F}\mathbf{x}\right\|^2\right\}. \tag{19}$$

The time domain transmitted signal x_n can be modelled as complex Gaussian via the central limit theorem when N is large. The covariance matrix and mean of $\mathbf{x}$ (7) are given by

$$\bar{\mathbf{x}} = E\{\mathbf{x}\} = \mathbf{F}^H\mathbf{\Phi}\mathbf{P},\ \mathbf{R}_x = E\{(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}})^H\} = \mathbf{F}^H\mathbf{\Psi}^2\mathbf{F}. \tag{20}$$

The mean depends on the fixed, superimposed pilots. The average of $f(\mathbf{Y}|\mathbf{x},\mathbf{H}_D)$ with respect to $\mathbf{x}$ gives the marginal likelihood function $f(\mathbf{Y}|\mathbf{H}_D)$, which can be expressed as

$$f(\mathbf{Y}|\mathbf{H}_D) = \int f(\mathbf{Y}|\mathbf{x},\mathbf{H}_D)f(\mathbf{x})d\mathbf{x} \tag{21}$$

where $f(\mathbf{x})$ is the pdf of $\mathbf{x}$. We evaluate (21) as

$$\begin{aligned} f(\mathbf{Y}|\mathbf{H}_D) &= \frac{1}{\det(\sigma_n^2\mathbf{I}_N + \mathbf{H}_D\mathbf{\Psi}^2\mathbf{H}_D^H)} \\ &\times \exp\{-[\mathbf{P}^H\mathbf{\Phi}^H(\mathbf{\Psi}^2+\sigma_n^2(\mathbf{H}_D^H\mathbf{H}_D)^{-1})^{-1}\mathbf{\Phi}\mathbf{P} \\ &+\sigma_n^2\mathbf{Y}^H(\mathbf{H}_D^H\mathbf{H}_D\mathbf{\Psi}^2+\sigma_n^2\mathbf{I}_N)^{-1}\mathbf{Y} \\ &+2\sigma_n^2\mathrm{Re}(\mathbf{Y}^H\mathbf{H}_D(\mathbf{\Psi}^2\mathbf{H}_D^H\mathbf{H}_D+\sigma_n^2\mathbf{I}_N)^{-1}\mathbf{\Phi}\mathbf{P})]\}. \end{aligned} \tag{22}$$

After some manipulations, we find the log likelihood function $\Lambda(\mathbf{Y}|\mathbf{H}_D) = \ln f(\mathbf{Y}|\mathbf{H}_D)$ as

$$\Lambda(\mathbf{Y}|\mathbf{H}_D) = -\sigma_n^2\sum_{k=0}^{N-1}\left(\frac{|Y_k-\phi_kH_kp_k|^2}{\psi_k|H_k|^2+\sigma_n^2} + \ln\left(\psi_k|H_k|^2+\sigma_n^2\right)\right). \tag{23}$$

Maximizing (23) is equivalent to minimizing

$$g(\mathbf{h}) = -\Lambda(\mathbf{Y}|\mathbf{H}_D). \tag{24}$$

By minimizing each term of (24) individually, we get the initial estimate of $\hat{H}_k$. The time-domain CIR estimate $\hat{\mathbf{h}}$ is $\mathbf{F}_L^H\hat{\mathbf{H}}$. Using the resulting initial channel estimate, the CIR $\mathbf{h}$ can be estimated by minimizing (24) via the gradient descent and related algorithms. However, the gradient based algorithms cannot guarantee the global minimum. To improve the channel estimate, the iterative ML and MMSE estimators (14) and (18) may be applied. Since the local minimum of (24) seems fairly close the global minimum, just one iteration is enough to achieve the CRB.

For the separated training scheme in [1] with the index set of pilot subcarriers I_p, minimizing (24) reduces to

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}\in\mathcal{C}^L}\sum_{k\in I_p}|Y_k - H_kp_k|^2 \tag{25}$$

which is identical to the conventional least squares (LS) channel estimator, a natural conclusion since (24) is derived in the ML sense.

V. Maximum Likelihood Data Detectors

In this section, we derive two data detectors. If the transmitted data $\mathbf{X}_D$ were known, the channel response $\mathbf{h}$ would be estimated using

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}\in\mathcal{C}^L}\left\|\mathbf{X}_D^{-1}\mathbf{Y}-\mathbf{F}_L\mathbf{h}\right\|^2. \tag{26}$$

The LS solution to (26) is

$$\hat{\mathbf{h}} = \mathbf{F}_L^H\mathbf{X}_D^{-1}\mathbf{Y}. \tag{27}$$

Using this CIR estimate (27), the data detector is given by

$$\begin{aligned} \hat{\mathbf{S}} &= \arg\min_{\mathbf{S}\in\mathcal{Q}^N}\left\|\mathbf{X}_D^{-1}\mathbf{Y}-\mathbf{F}_L\mathbf{F}_L^H\mathbf{X}_D^{-1}\mathbf{Y}\right\|^2 \\ &= \arg\min_{\mathbf{S}\in\mathcal{Q}^N}\check{\mathbf{X}}^H\mathbf{Y}_D^H\left(\mathbf{I}_N-\mathbf{F}_L\mathbf{F}_L^H\right)\mathbf{Y}_D\check{\mathbf{X}} \\ &= \arg\min_{\mathbf{S}\in\mathcal{Q}^N}\|\mathbf{U}_1\mathbf{Y}_D\check{\mathbf{X}}\|^2 \end{aligned} \tag{28}$$

where $\check{\mathbf{X}} = [X_0^{-1},\ldots,X_{N-1}^{-1}]^T$, and $\mathbf{U}_1$ is the Cholesky decomposition of $\mathbf{I}_N - \mathbf{F}_L\mathbf{F}_L^H$. Using (1), and assuming a unitary constellation, we find

$$X_k^{-1} = \frac{\sqrt{\phi_k}p_k - \sqrt{\psi_k}s_k}{\phi_k-\psi_k},\ k=0,1,\ldots,N-1. \tag{29}$$

Define $\check{\mathbf{\Phi}} = \mathrm{diag}\{\frac{\sqrt{\phi_0}}{\phi_0-\psi_0},\ldots,\frac{\sqrt{\phi_{N-1}}}{\phi_{N-1}-\psi_{N-1}}\}$ and $\check{\mathbf{\Psi}} = \mathrm{diag}\{\frac{\sqrt{\psi_0}}{\phi_0-\psi_0},\ldots,\frac{\sqrt{\phi_{N-1}}}{\psi_{N-1}-\psi_{N-1}}\}$. ¿From (29), we have $\check{\mathbf{X}} =$

$\check{\mathbf{\Phi}}\mathbf{P} - \check{\mathbf{\Psi}}\mathbf{S}$. Substituting this equation into (28), it reduces to

$$\hat{\mathbf{S}} = \arg\min_{\mathbf{S}\in\mathcal{Q}^N} \|\mathbf{R}_1 - \mathbf{G}_1\mathbf{S}\|^2 \tag{30}$$

where $\mathbf{R}_1 = \mathbf{U}_1\mathbf{Y}_D\check{\mathbf{\Phi}}\mathbf{P}$, and $\mathbf{G}_1 = \mathbf{U}_1\mathbf{Y}_D\check{\mathbf{\Psi}}$. We refer to the detector (30) as MLDD1.

If the CIR $\mathbf{h}$ is complex Gaussian, the received signal (6) is also complex Gaussian with zero mean, and the covariance matrix is given by

$$\mathbf{R}_Y = \mathbf{X}_D\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H\mathbf{X}_D^H + \sigma_n^2\mathbf{I}_N. \tag{31}$$

The pdf of $\mathbf{Y}$ conditioned on $\mathbf{S}$ is thus

$$f(\mathbf{Y}|\mathbf{S}) = \frac{1}{\det(\mathbf{R}_Y)}\exp\left\{-\mathbf{Y}^H\mathbf{R}_Y^{-1}\mathbf{Y}\right\}. \tag{32}$$

$\log\det(\mathbf{R}_Y) = \log\det(\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H\mathbf{X}_D^H\mathbf{X}_D + \sigma_n^2\mathbf{I}_N)$ is independent of $\mathbf{X}_D$. Therefore, maximizing (32) is equivalent to solving

$$\hat{\mathbf{S}} = \arg\min_{\mathbf{S}\in\mathcal{Q}^N} \mathbf{Y}^H\left(\mathbf{X}_D^{-1}\right)^H\left(\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H + \sigma_n^2(\mathbf{X}_D^H\mathbf{X}_D)^{-1}\right)^{-1}\mathbf{X}_D^{-1}\mathbf{Y} \tag{33}$$

Eq. (33) can be solved exactly using branch and bound as the blind estimator for both unitary and non-unitary constellations. Following the same approach in (28) and using the same definition, (33) can be simplified as

$$\hat{\mathbf{S}} = \arg\min_{\mathbf{S}\in\mathcal{Q}^N} \|\mathbf{R}_2 - \mathbf{G}_2\mathbf{S}\|^2 \tag{34}$$

where $\mathbf{R}_2 = \mathbf{U}_2\mathbf{Y}_D\check{\mathbf{\Phi}}\mathbf{P}$, $\mathbf{G}_2 = \mathbf{U}_2\mathbf{Y}_D\check{\mathbf{\Psi}}$ and $\mathbf{U}_2$ is the Cholesky decomposition of $(\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H + \sigma_n^2\mathbf{\Lambda}^{-1})^{-1} = \mathbf{U}_2^H\mathbf{U}_2$. We refer to the detector (34) as MLDD2.

Since $\mathbf{S}$ is from a discrete set $\mathcal{Q}$, (30) and (34) are well-known as integer least squares problems. SD [7] solves (30) and (34) with moderate complexity.

VI. Cramer-Rao Bound and Optimal Pilots Placement

Though several channel estimation algorithms have been proposed for a superimposed training scheme [3]–[5], no optimal pilots placement scheme has been given. We thus optimize pilots by minimizing the CRB on the mean square error (MSE) of channel estimation with superimposed training.

The CRB for the MSE of channel estimation can be derived as

$$\mathrm{CRB}_h = \left(\frac{1}{\sigma_n^2}\mathbf{F}_L^H(\mathbf{\Phi}^2 + \mathbf{\Psi}^2)\mathbf{F}_L + \mathbf{R}_h^{-1}\right)^{-1}. \tag{35}$$

Optimal pilots are designed by minimizing the CRB with respect to the placement and power constraint of the pilots and data symbols. Assuming that the subcarrier index set for non-zero ϕ_k is I_p with N_p elements, and the index set for the rest of the subcarriers is I_d with N_d elements, the power constraint $\sum_k \psi_k = \mathcal{D}$ and $\sum_{k\in I_p}\phi_k = \mathcal{P}$, where $\mathcal{D}$ and $\mathcal{P}$ are the total power on pilots and data symbols respectively, the problem becomes

$$(I_p, \phi_k, I_d, \psi_k) = \arg\min_{\sum_k \psi_k = \mathcal{D}, \sum_{k\in I_p}\phi_k = \mathcal{P}} \mathrm{tr}\left\{\left(\frac{1}{\sigma_n^2}\mathbf{F}_L^H(\mathbf{\Phi}^2 + \mathbf{\Psi}^2)\mathbf{F}_L + \mathbf{R}_h^{-1}\right)^{-1}\right\}. \tag{36}$$

The lower bound on the MSE is attained if and only if $\mathbf{C} = \mathbf{F}_L^H(\mathbf{\Phi}^2 + \mathbf{\Psi}^2)\mathbf{F}_L$ is diagonal. The (i, j)th $(0 \le i, j \le L-1)$ entry of $\mathbf{C}$ can be written as

$$[\mathbf{C}]_{i,j} = \begin{cases} \mathcal{D} + \mathcal{P} & i = j \\ \sum_{k\in I_p}\phi_k e^{-\jmath 2\pi k(j-i)/N} + \sum_k \psi_k e^{-\jmath 2\pi k(j-i)/N} & i \ne j \end{cases}. \tag{37}$$

Therefore, we require

$$\sum_{k\in I_p}\phi_k e^{-\jmath 2\pi k(j-i)/N} + \sum_k \psi_k e^{-\jmath 2\pi k(j-i)/N} = (\mathcal{D} + \mathcal{P})\delta(i-j). \tag{38}$$

Eq. (38) is satisfied if the following conditions are satisfied

C1) $\mathcal{E}_k = C_1, \forall k \in I_p$ and $\mathcal{E}_k = C_2, \forall k \in I_d$.

C2) $d = N/N_p \in \mathcal{Z}$ and $I_p = \{d_0 + kd, k = 0, 1, \ldots, N_p - 1\}$, $d_0 \in \{0, 1, \ldots, d-1\}$.

C1) means that the total power at both each subcarrier in I_p and each subcarrier in I_d must be equipowered. C2) means that the superimposed pilots must be equispaced. In the separated training scheme, C1) and C2 agree with the equispace and equipower conditions. A remarkable property of C1) and C2 is that the CRB does not depend on the power allocated to the pilots and the location of pilots but depends only on the total power allocated to each subcarrier. However, the training power determines the SNR required to achieve the CRB. The more training power, the lower SNR is needed.

VII. Simulation Results

We test uncoded OFDM with $N = 32$ subcarriers and binary phase-shift keying (BPSK). A COST 207 6-ray channel model with the power delay profile $[0.189, 0.379, 0.239, 0.095, 0.061, 0.037]$ is considered. Each path is an independently generated complex Gaussian random process. We compare the performance of 8 equispaced superimposed pilots with that of 4 equispaced separated pilots, and denote them as SU and SE respectively; they allocated the same total training power. In SU, the total power at each subcarrier is 1, and superimposed pilot subcarriers have $\phi_k = \phi$. Since SU and SE have different bandwidth efficiencies, we compare them in terms of the effective SNR defined by $\frac{\mathcal{E}}{N_d N_0}$, where N_d is the number of information symbols in an OFDM symbol. IML and IMMSE are used to denote the iterative estimators in Section III, and the notation $i = n$ denotes the performance in the n-th iteration. Ideal detectors, assuming the availability of perfect CIR knowledge, are used as benchmarks.

Fig. 1 shows the MSE of channel estimation with $\phi = 0.2$. The iterative estimators cannot converge since the initial CIR estimates are bad. MLCE shows an error floor at high SNR because the gradient based algorithm cannot guarantee the

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2005 proceedings.

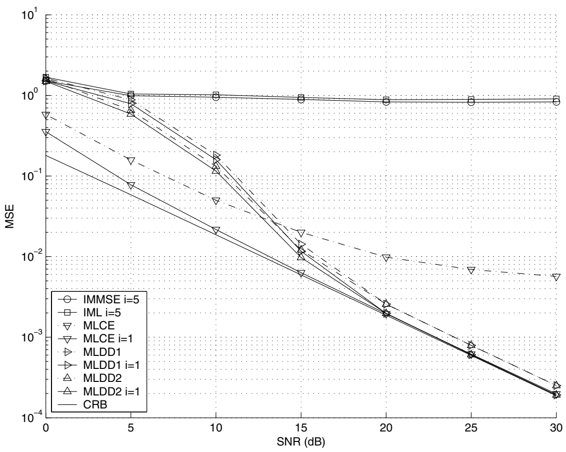

Fig. 1. MSE comparison with different estimators in a BPSK OFDM system with $N = 32$ and $\phi = 0.2$ for superimposed training.

global minimum. However, with only one iteration, MLCE achieves the MSE CRB in high SNR. Since the initial data detection in MLDDs at superimposed pilot subcarriers is not reliable, we first use the detected data at non-superimposed subcarriers to estimate the CIR and perform one iteration to update both data and CIR estimators. They also achieve the CRB in high SNR.

Figs. 2, 3 compare the MSE of channel estimation and the resulting BER for $\phi = 0.7$. The performance of the frequency domain channel estimator in [4] after 4 iterations is denoted as HFC. After 4 iterations, both IML and IMMSE converge to the CRB in high SNR (Fig. 2). However, they both exhibit error floors. IMMSE has a smaller error floor than IML in each iteration. Their MSE and BER performance are much better than those of HFC with the same number of iterations. Similarly, MLCE and MLDDs achieve the MSE CRB in high SNR (Fig. 2). In Fig. 3, SU with perfect CIR performs 1 dB worse than SE with perfect CIR at $\mathrm{BER} = 10^{-3}$. MLCE and MLDDs have the same performance in high SNR. They perform close to SE with MMSE channel estimation but have a 0.5-dB loss over SU with perfect CIR.

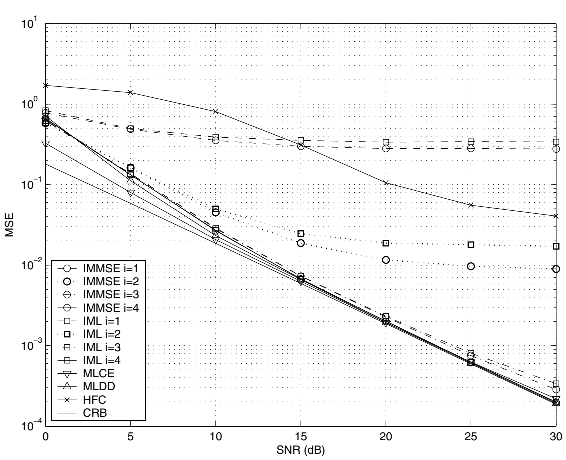

Fig. 2. MSE for different estimators in a BPSK OFDM system with $N = 32$ and $\phi = 0.7$ for superimposed training.

Comparing Fig. 1 with Fig. 2, we find that MLCE and MLDDs achieve the CRB at 20 dB with $\phi = 0.2$ and at 15 dB with $\phi = 0.7$. But the CRBs are identical. This verifies that the training power only determines the SNR threshold to achieve the CRB. However, the analysis of the dependence of threshold on ϕ seems to be intractable.

VIII. Conclusion

We have investigated superimposed training for OFDM and derived iterative ML and MMSE channel estimators. We derived the ML channel estimator by averaging the likelihood function over the distribution of the transmitted signal. Two ML data detectors have also derived by eliminating the CIR first. The resulting integer quadratic optimization problems are efficiently solved using an SD. Moreover, we show that equispaced superimposed pilot symbols are optimal for minimizing the CRB. Our proposed channel estimators and data detectors achieve the CRB in high SNR. They perform close to separated training with perfect CIR but with a higher data rate. The training scheme in this paper can be readily generalized to two dimensional training sequences and extended to coded systems where iterative detection strategies are employed.

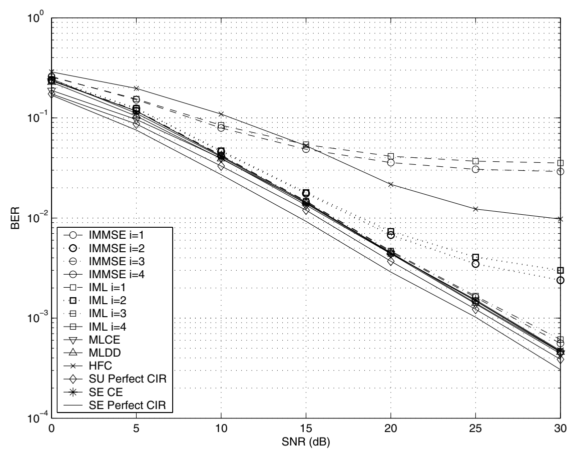

Fig. 3. BER for different estimators in a BPSK OFDM system with $N = 32$ and $\phi = 0.7$ for superimposed training.

References

[1] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 4, pp. 1207–1215, Jul. 2000.

[2] C. Kastenholz and W. Birkemeier, "A simultaneous information transfer and channel-sounding modulation technique for wide-band channels," *IEEE Trans. Commun.*, vol. 13, no. 2, pp. 162 – 165, Jun 1965.

[3] B. Farhang-Boroujeny, "Pilot-based channel identification: proposal for semi-blind identification of communication channels," *Electronics Letters*, vol. 31, no. 13, pp. 1044 – 1046, June 1995.

[4] C. K. Ho, B. Farhang-Boroujeny, and F. Chin, "Added pilot semi-blind channel estimation scheme for OFDM in fading channels," in *Proc. of GLOBECOM '01*, vol. 5, Nov. 2001, pp. 3075 – 3079.

[5] N. Chen and G. T. Zhou, "A superimposed periodic pilot scheme for semi-blind channel estimation of OFDM systems," in *Proc. of IEEE 10th Digital Signal Processing Workshop*, Oct. 2002, pp. 362 – 365.

[6] M. Ghogho, D. McLernon, E. Alameda-Hernandez, and A. Swami, "Channel estimation and symbol detection for block transmission using data-dependent superimposed training," *IEEE Signal Processing Lett.*, vol. 12, no. 3, pp. 226 – 229, Mar. 2005.

[7] O. Damen, A. Chkeif, and J. C.Belfiore, "Lattice code decoder for space-time codes," *IEEE Commun. Lett.*, vol. 4, no. 5, pp. 161 – 163, May 2000.

