Joint Power Loading of Data and Pilots in OFDM using Imperfect Channel State Information at the Transmitter

Chitaranjan P. Sukumar
University of California, Irvine
Address,
Email: cpelursu@uci.edu

Ricardo Merched
Dept. of Electrical and Computer Engineering
Universidade Federal do Rio de Janeiro, Brazil
Email: merched@lps.ufrj.br

Ahmed Eltawil
University of California, Irvine
Address,
Email: aeltawil@uci.edu

***Abstract*— The search for optimality in the design of channel precoders and training symbols in block processing communication systems is one of paramount importance. Finding the best tradeoff in terms of power distribution between information and pilot symbols for frequency selective channels, when channel estimation via feedback is available, however, has not been fully addressed. In this paper, we solve the problem of finding the optimal power distribution between pilots and data symbols in the mean-square-error (MSE) sense when a delayless error-free channel feedback path is available to the transmitter. The novel approach adaptively designs the optimal precoders and training vectors based on the frequency domain estimates of the channel.**

I. Introduction

Orthogonal Frequency Division Multiplexing (OFDM) is the de-facto technology of choice in a number of newly introduced commercial standards [1], [2]. In this context, one of the most effective ways of estimating frequency selective channels is by embedding pilots in the data stream. Once the channel state information (CSI) is measured at the receiver, further improvement in capacity and bit error rates can be achieved via its feedback to the transmitter. These can be obtained through the common notion of precoding[3].

In practice, the estimated CSI is imperfect and could be outdated by the time it is sent back to the transmitter. Recent efforts [4],[5],[6] that rely on channel estimates have focused on the design of data precoders in order to achieve optimality in terms of capacity, bit error rates, or symbol mean square error. References [7],[8],[9] have taken a more holistic approach and determined the optimum power allocation for data and pilots. In reference [7], the authors maximize the capacity in a Multiple Input Multiple Output (MIMO) scenario by trying to find the optimum power for training, as well as the optimum number of pilots. Also, the training power that maximizes the average spectral efficiency while maintaining a specific bit-error-rate (BER) has been obtained in [8] and [9], the former considering SISO links, while the latter in the context of MIMO Space Time Block Codes (STBC). To the best of our knowledge, in all instances that pursue optimum power distribution for pilots and data, the channel is assumed to be flat fading. Reference [10] discusses the optimal ratio of data and pilot power in a MIMO OFDM system by minimizing the symbol error rate, however power loading for each individual subcarriers is not considered.

In this paper, we use the previously estimated CSI to power load each individual data and training subcarriers in a frequency selective scenario, which leads to superior symbol mean square error performance when compared to the case where the CSI is used to design the precoder only.

Notation: We shall denote by * the complex conjugate transposition. Vectors are defined by lowercase letters, while matrices by capital letters. We denote the element in the m-th row and n-th column of $\boldsymbol{A}$ by $A(m,n)$. Also, the n-th element of a vector $\boldsymbol{z}$ will be represented as $z(n)$. The operator $diag(\cdot)$ captures the diagonal elements of a matrix into a vector of corresponding dimension. The same notation will be used to map a vector into a diagonal matrix. Also, we shall denote by $\mathsf{E}\boldsymbol{z}$ the expected value of a quantity $\boldsymbol{z}$.

II. Problem Formulation

Consider an L tap discrete single-input-single-output (SISO) channel which we assume to be invariant within a single OFDM frame (but changing slowly from one frame to another).We assume the transmitted vector is sent via M subcarriers, that is, let $\boldsymbol{x}_i$ be the i-th $M \times 1$ frequency domain OFDM frame. The IDFT of $\boldsymbol{x}_i$ is then taken and a cyclic prefix of length greater than or equal to L is added before the final transmission. At the receiver, the cyclic prefix is discarded and after the DFT is taken, we can express the i-th received frame as

$$\boldsymbol{y}_i = \boldsymbol{\Lambda}_{h_i}\boldsymbol{x}_i + \boldsymbol{n}_i, \tag{1}$$

where $\boldsymbol{y}_i$ is $M \times 1$, $\boldsymbol{\Lambda}_{h_i} = diag(\boldsymbol{\lambda}_{h_i})$, with $\boldsymbol{\lambda}_{h_i} = \sqrt{M}\boldsymbol{F}\begin{bmatrix} \boldsymbol{h}_i \\ \boldsymbol{0}_{(M-L)\times 1}\end{bmatrix}$, $\boldsymbol{h}_i$ is the $L \times 1$ channel and $\boldsymbol{F}$ is the $M \times M$ DFT matrix. The vector $\boldsymbol{n}_i$ is the DFT of the zero mean Gaussian noise vector with variance $\mathsf{E}\boldsymbol{n}_i\boldsymbol{n}_i^* = \sigma_n^2\boldsymbol{I}$.

978-1-4244-2324-8/08/$25.00 © 2008 IEEE.

Note that the covariance matrix $\mathsf{E}\boldsymbol{\lambda}_{h_i}\boldsymbol{\lambda}_{h_i}^*$ is given by

$$\begin{aligned}\mathsf{E}\boldsymbol{\lambda}_{h_i}\boldsymbol{\lambda}_{h_i}^* &= M\boldsymbol{F}\mathsf{E}\left[\begin{array}{c}\boldsymbol{h}_i\\ \boldsymbol{0}\end{array}\right]\left[\begin{array}{c}\boldsymbol{h}_i\\ \boldsymbol{0}\end{array}\right]^*\boldsymbol{F}^* = \\ &= M\boldsymbol{F}\left[\begin{array}{cc}\mathsf{E}\boldsymbol{h}_i\boldsymbol{h}_i^* & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{array}\right]\boldsymbol{F}^* \qquad (2)\end{aligned}$$

The variance $\mathsf{E}\boldsymbol{h}_i\boldsymbol{h}_i^*$ represents the second order statistics of the channel and are assumed known at both transmitter and receiver. Each vector $\boldsymbol{x}_i$ consists of D data symbols and P pilot symbols, so that $D + P = M$. For convenience, we will drop the time index. In general, an OFDM frame can written as

$$\boldsymbol{x} = \boldsymbol{A}'\boldsymbol{s} + \boldsymbol{t}', \qquad (3)$$

where $\boldsymbol{A}'$ is the $M \times D$ precoder matrix for the data symbols $\boldsymbol{s}$ which are picked from a known constellation and has mean zero and variance $\mathsf{E}\boldsymbol{s}\boldsymbol{s}^* = \sigma_s^2\boldsymbol{I}$. The vector $\boldsymbol{t}'$ contains pilot symbols. This is commonly known as Affine precoding [12],[13]. In the latter work, it has been noted that, in order to prevent data symbols from interfering with the channel estimation, pilots and data must be transmitted on different (nonoverlapping) subcarriers. This is achieved as follows. Let Q be the set of all subcarriers in a frame which carry data only, and let K be the set of all subcarriers carrying pilots. Then, the intersection of Q and K is a null set. For all $q \in Q$, we enforce $\boldsymbol{t}'(q) = 0$. Similarly, for all $k \in K$, we have $\boldsymbol{A}'(k,:) = \boldsymbol{0}_{1\times D}$, where $\boldsymbol{A}'(k,:)$ refers to the k-th row of $\boldsymbol{A}'$.

We assume that there exists an instantaneous error free feedback channel from receiver to the transmitter, which is responsible for informing the transmitter on the channel state information. Thus, given the received CSI, our goal is to design matrices $\boldsymbol{A}'$ and $\boldsymbol{t}'$ in order to minimize the symbol mean square error.

Due to the nonoverlapping structure of data and pilots, we can decompose (1) as

$$\boldsymbol{y}_d = \boldsymbol{\Lambda}_{h_d}\boldsymbol{A}\boldsymbol{s} + \boldsymbol{n}_d \qquad (4)$$

$$\boldsymbol{y}_t = \boldsymbol{\Lambda}_{h_t}\boldsymbol{t} + \boldsymbol{n}_t, \qquad (5)$$

where $\boldsymbol{y}_d$ is the $D \times 1$ vector formed from the D data subcarriers of $\boldsymbol{y}$, while $\boldsymbol{y}_t$ is the $P \times 1$ vector obtained from the P subcarriers of $\boldsymbol{y}$ carrying training symbols. The diagonal matrix $\boldsymbol{\Lambda}_{h_d}$ and $\boldsymbol{A}$ are obtained from $\boldsymbol{\Lambda}_h$ and $\boldsymbol{A}'$ respectively by removing the rows which correspond to the training subcarriers. Similarly, the diagonal matrix $\boldsymbol{\Lambda}_{h_t}$ and $\boldsymbol{t}$ are obtained from $\boldsymbol{\Lambda}_h$ and $\boldsymbol{t}'$ respectively, after eliminating the rows corresponding to data subcarriers. The vectors $\boldsymbol{n}_d$ and $\boldsymbol{n}_t$ represent the noise in the data and pilot subcarriers respectively.

III. Channel Estimation

Note that equation (5) can be written as

$$\boldsymbol{y}_t = \boldsymbol{\Lambda}_t\boldsymbol{\lambda}_{h_t} + \boldsymbol{n}_t, \qquad (6)$$

where $\boldsymbol{\Lambda}_t = diag(\boldsymbol{t})$ and $\boldsymbol{\lambda}_{h_t} = diag(\boldsymbol{\Lambda}_{h_t})$. Thus, given the above model, the least squares (LS) estimate of $\boldsymbol{\lambda}_{h_t}$ is given by

$$\hat{\boldsymbol{\lambda}}_{h_t} = (\boldsymbol{\Lambda}_t^*\boldsymbol{\Lambda}_t)^{-1}\boldsymbol{\Lambda}_t^*\boldsymbol{y}_t, \qquad (7)$$

so that the m-th element of $\hat{\boldsymbol{\lambda}}_{h_t}$ is given by

$$\hat{\lambda}_{h_t}(m) = \frac{y_t(m)}{t(m)} = \left(\frac{t(m)\lambda_{h_t}(m) + n_t(m)}{t(m)}\right) \qquad (8)$$

Hence, once $\hat{\boldsymbol{\lambda}}_{h_t}$ is estimated, it is fedback to the transmitter for the purpose of designing the precoder and training matrices. Let the corresponding estimation error be denoted by $\tilde{\boldsymbol{\lambda}}_{h_t} = \boldsymbol{\lambda}_{h_t} - \hat{\boldsymbol{\lambda}}_{h_t}$. The associated error covariance for the training subcarriers is then given by [11]

$$\mathsf{E}\tilde{\boldsymbol{\lambda}}_{h_t}\tilde{\boldsymbol{\lambda}}_{h_t}^* = (\sigma_n^{-2}\boldsymbol{\Lambda}_t^*\boldsymbol{\Lambda}_t)^{-1}. \qquad (9)$$

In this paper, we shall assume that the remaining channel information (the part corresponding to data subcarriers) is obtained via interpolation from the training channel estimates. That is,

$$\hat{\boldsymbol{\lambda}}_{h_d} = \boldsymbol{B}\hat{\boldsymbol{\lambda}}_{h_t}, \qquad (10)$$

where $\boldsymbol{B}$ is a $D\times P$ interpolating matrix, whose entries depend on the interpolating technique used.

We shall denote the estimation error of $\boldsymbol{\lambda}_{h_d}$ by $\tilde{\boldsymbol{\lambda}}_{h_d}$. Clearly, we have that

$$\boldsymbol{\lambda}_{h_d} = \hat{\boldsymbol{\lambda}}_{h_d} + \tilde{\boldsymbol{\lambda}}_{h_d}. \qquad (11)$$

IV. Symbol MMSE

By virtue of (11), the model in (4) can readily be written as

$$\boldsymbol{y}_d = \hat{\boldsymbol{\Lambda}}_{h_d}\boldsymbol{A}\boldsymbol{s} + \underbrace{\tilde{\boldsymbol{\Lambda}}_{h_d}\boldsymbol{A}\boldsymbol{s} + \boldsymbol{n}_d}_{\boldsymbol{w}}, \qquad (12)$$

where $\hat{\boldsymbol{\Lambda}}_{h_d} = diag(\hat{\boldsymbol{\lambda}}_{h_d})$, $\tilde{\boldsymbol{\Lambda}}_{h_d} = diag(\tilde{\boldsymbol{\lambda}}_{h_d})$ and we have defined the effective noise $\boldsymbol{w}$ as

$$\boldsymbol{w} = \tilde{\boldsymbol{\Lambda}}_{h_d}\boldsymbol{A}\boldsymbol{s} + \boldsymbol{n}_d. \qquad (13)$$

Note that $\boldsymbol{w}$ contains the effect of channel estimation error, as well as noise. The symbol estimation can be pursued in several different ways. In this paper, we shall assume for simplicity, its weighed least squares estimate given by [11]

$$\hat{\boldsymbol{s}} = (\boldsymbol{A}^*\hat{\boldsymbol{\Lambda}}_{h_d}^*\boldsymbol{R}_w^{-1}\hat{\boldsymbol{\Lambda}}_{h_d}\boldsymbol{A})^{-1}\boldsymbol{A}^*\hat{\boldsymbol{\Lambda}}_{h_d}^*\boldsymbol{R}_w^{-1}\boldsymbol{y}_d, \qquad (14)$$

where $\boldsymbol{R}_w \triangleq \mathsf{E}\boldsymbol{w}\boldsymbol{w}^*$. The associated mean square error is given by [11]

$$tr(\mathsf{E}\tilde{\boldsymbol{s}}\tilde{\boldsymbol{s}}^*) = tr(\boldsymbol{A}^*\hat{\boldsymbol{\Lambda}}_{h_d}^*\boldsymbol{R}_w^{-1}\hat{\boldsymbol{\Lambda}}_{h_d}\boldsymbol{A})^{-1}, \qquad (15)$$

where $\tilde{\boldsymbol{s}}$ is the error in estimating $\boldsymbol{s}$, while $\boldsymbol{R}_w$

$$\begin{aligned}\boldsymbol{R}_w &= \mathsf{E}(\tilde{\boldsymbol{\Lambda}}_{h_d}\boldsymbol{A}\boldsymbol{s} + \boldsymbol{n}_d)(\tilde{\boldsymbol{\Lambda}}_{h_d}\boldsymbol{A}\boldsymbol{s} + \boldsymbol{n}_d)^*, \qquad (16)\\ &= \mathsf{E}\left(\tilde{\boldsymbol{\Lambda}}_{h_d}\boldsymbol{A}\mathsf{E}(\boldsymbol{s}\boldsymbol{s}^* \mid \tilde{\boldsymbol{\Lambda}}_{h_d})\boldsymbol{A}^*\tilde{\boldsymbol{\Lambda}}_{h_d}^*\right) + \sigma_n^2\boldsymbol{I}\\ &= \sigma_s^2\mathsf{E}(\tilde{\boldsymbol{\Lambda}}_{h_d}\boldsymbol{A}\boldsymbol{A}^*\tilde{\boldsymbol{\Lambda}}_{h_d}^*) + \sigma_n^2\boldsymbol{I} \qquad (17)\end{aligned}$$

where we have used the fact that $\boldsymbol{s}$ is independent of $\tilde{\boldsymbol{\Lambda}}_{h_d}$ and $\boldsymbol{n}_d$. For the above trace to be minimized, it is sufficient to pick its argument as a diagonal matrix [14]. This condition is satisfied, by choosing $\boldsymbol{A}$ as diagonal as well. Therefore, using (17), the mean square error cost function in (15) can be written as

$$\sum_{q=1}^{D} \mathsf{E}|s(q)-\hat{s}(q)|^2 = \sum_{q=1}^{D} \frac{\sigma_s^2 \mathsf{E}|\tilde{\lambda}_{h_d}(q)|^2 |A(q,q)|^2 + \sigma_n^2}{\sigma_s^2 |\hat{\lambda}_{h_d}(q)|^2 |A(q,q)|^2} \quad (18)$$

Now, before proceeding with the minimization of (18), we still need calculate the term $\mathsf{E}|\tilde{\lambda}_{h_d}(q)|^2$. Expanding it, we get

$$\begin{aligned}\mathsf{E}|\tilde{\lambda}_{h_d}(q)|^2 &= \mathsf{E}[\lambda_{h_d}(q)-\hat{\lambda}_{h_d}(q)][\lambda_{h_d}(q)-\hat{\lambda}_{h_d}(q)]^* \\ &= \mathsf{E}[\lambda_{h_d}(q)-\boldsymbol{B}(q,:)\hat{\boldsymbol{\lambda}}_{h_t}][\lambda_{h_d}(q)-\boldsymbol{B}(q,:)\hat{\boldsymbol{\lambda}}_{h_t}]^* \\ &= \mathsf{E}|\lambda_{h_d}(q)|^2 - 2Re\mathsf{E}(\lambda_{h_d}(q)\hat{\boldsymbol{\lambda}}_{h_t}^*)\boldsymbol{B}^*(q,:) + \\ &\quad \boldsymbol{B}(q,:)\mathsf{E}(\hat{\boldsymbol{\lambda}}_{h_t}\hat{\boldsymbol{\lambda}}_{h_t}^*)\boldsymbol{B}^*(q,:) \end{aligned} \quad (19)$$

where we have used (10) in the second line. The notation $Re(z)$ represents the real part of a complex variable z. To determine the second term, we first find the l-th element of $\mathsf{E}(\lambda_{h_d}(q)\hat{\boldsymbol{\lambda}}_{h_t}^*)$. That is,

$$\begin{aligned}\mathsf{E}\lambda_{h_d}(q)\hat{\lambda}_{h_t}(l)^* &= \mathsf{E}\lambda_{h_d}(q)\left(\frac{t(l)\lambda_{h_t}(l)+n_t(l)}{t(l)}\right)^* \\ &= \mathsf{E}\lambda_{h_d}(q)\lambda_{h_t}^*(l), \end{aligned} \quad (20)$$

where we have used the fact that $\lambda_{h_t}(l)$ is given by (8) and also that $\lambda_{h_d}(q)$ and $n_t(l)$ are uncorrelated. To calculate the entries of $\mathsf{E}(\hat{\boldsymbol{\lambda}}_{h_t}\hat{\boldsymbol{\lambda}}_{h_t}^*)$, we note that the cross correlation between the estimated channel at the m-th subcarrier and the estimated channel at the l-th subcarrier is given by

$$\begin{aligned}\mathsf{E}\hat{\lambda}_{h_t}(m)\hat{\lambda}_{h_t}(l)^* &= \mathsf{E}\left(\frac{t(m)\lambda_{h_t}(m)+n_t(m)}{t(m)}\right) \\ &\qquad \left(\frac{t(l)\lambda_{h_t}(l)+n_t(l)}{t(l)}\right)^* \\ &= \begin{cases} \mathsf{E}|\lambda_{h_t}(m)|^2 + \frac{\sigma_n^2}{|t(m)|^2} \text{ if } m=l \\ \mathsf{E}\lambda_{h_t}(m)\lambda_{h_t}^*(l) \text{ otherwise.}\end{cases}\end{aligned} \quad (21)$$

Using these results in (19) we can now write it as

$$\begin{aligned}\mathsf{E}|\tilde{\lambda}_{h_d}(q)|^2 &= \mathsf{E}|\lambda_{h_d}(q)|^2 - 2Re\mathsf{E}(\lambda_{h_d}(q)\boldsymbol{\lambda}_{h_t}^*)\boldsymbol{B}^*(q,:) + \\ &\quad \boldsymbol{B}(q,:)\mathsf{E}\boldsymbol{\lambda}_{h_t}\boldsymbol{\lambda}_{h_t}^*\boldsymbol{B}^*(q,:) + \boldsymbol{B}(q,:)\boldsymbol{D}\boldsymbol{B}^*(q,:),\end{aligned} \quad (22)$$

where $\boldsymbol{D} = diag\left(\left[\begin{array}{cccc}\frac{\sigma_n^2}{|t(1)|^2} & \frac{\sigma_n^2}{|t(2)|^2} & \cdots & \frac{\sigma_n^2}{|t(P)|^2}\end{array}\right]\right)$. The last two terms in the above equation were obtained from the last term in (19) using (21). To calculate $\mathsf{E}(\lambda_{h_d}(q)\boldsymbol{\lambda}_{h_t}^*)$ note that

$$\begin{aligned}&\mathsf{E}\lambda_{h_d}(q)\boldsymbol{\lambda}_{h_t}^* = \\ &\left[\begin{array}{cccc}\mathsf{E}\lambda_{h_d}(q)\lambda_{h_t}^*(1) & \mathsf{E}\lambda_{h_d}(q)\lambda_{h_t}^*(2) & \cdots & \mathsf{E}\lambda_{h_d}(q)\lambda_{h_t}^*(P)\end{array}\right]\end{aligned} \quad (23)$$

Also, note that $\mathsf{E}\boldsymbol{\lambda}_{h_t}\boldsymbol{\lambda}_{h_t}^*$ can be written as

$$\begin{aligned}&\mathsf{E}\boldsymbol{\lambda}_{h_t}\boldsymbol{\lambda}_{h_t}^* = \\ &\left[\begin{array}{cccc}\mathsf{E}|\lambda_{h_t}(1)|^2 & \mathsf{E}\lambda_{h_t}(1)\lambda_{h_t}^*(2) & \cdots & \mathsf{E}\lambda_{h_t}(1)\lambda_{h_t}^*(P) \\ \mathsf{E}\lambda_{h_t}(2)\lambda_{h_t}^*(1) & \mathsf{E}|\lambda_{h_t}(1)|^2 & \cdots & \mathsf{E}\lambda_{h_t}(2)\lambda_{h_t}^*(P) \\ \vdots & \ddots & & \\ \mathsf{E}\lambda_{h_t}(P)\lambda_{h_t}^*(1) & \mathsf{E}\lambda_{h_t}(P)\lambda_{h_t}^*(2) & \cdots & \mathsf{E}|\lambda_{h_t}(P)|^2\end{array}\right]\end{aligned} \quad (24)$$

In both of the above equations, each entry can be found by reading off the corresponding elements of (2).

Note that in (22) we have described how the power of the individual pilots affect the symbol mean square error in (18).

V. Optimization

We are now ready to state our optimization problem as

$$\begin{aligned}\min_{A(q,q),t(k)} \sum_{q=1}^{D} \frac{\sigma_s^2\mathsf{E}|\tilde{\lambda}_{h_d}(q)|^2|A(q,q)|^2+\sigma_n^2}{\sigma_s^2|\hat{\lambda}_{h_d}(q)|^2|A(q,q)|^2} \\ s.t. \quad \sigma_s^2\sum_{\ell=1}^{D}|A(\ell,\ell)|^2 + \sum_{k=1}^{P}|t(k)|^2 = P_t,\end{aligned} \quad (25)$$

where P_t is the total transmit power. Here, we would like to point out that the above optimization problem has been stated in terms of $|\hat{\lambda}_{h_d}(q)|^2$, which is the current estimated channel at the receiver. Now,as this optimization has to be performed at the transmitter before the OFDM symbol is transmitted, it is impossible for the transmitter to have $|\hat{\lambda}_{h_d}(q)|^2$. It only has outdated information about the same (from previously fedback information from the receiver). We argue here that as the channel has been assumed to be slowly varying, it is an acceptable approximation to substitute $|\hat{\lambda}_{h_d}(q)|^2$ with this outdated information. So from this point onward, $|\hat{\lambda}_{h_d}(q)|^2$ represents the outdated channel information available at the transmitter. Note that this is valid only for solving the optimization problem and the upcoming expressions for power loading. As far as the receiver architecture in equation (14) goes, $|\hat{\lambda}_{h_d}(q)|^2$ maintains its meaning as the current estimated channel at the receiver. In order to solve the optimization problem in (25) we first obtain the lagrangian

$$\begin{aligned}\mathfrak{L} &= \sum_{q=1}^{D}\frac{\sigma_s^2\mathsf{E}|\tilde{\lambda}_{h_d}(q)|^2|A(q,q)|^2+\sigma_n^2}{\sigma_s^2|\hat{\lambda}_{h_d}(q)|^2|A(q,q)|^2} \\ &\quad + \mu(\sigma_s^2\sum_{\ell=1}^{D}|A(\ell,\ell)|^2 + \sum_{k=1}^{P}|t(k)|^2 - P_t)\end{aligned} \quad (26)$$

where μ is the lagrangian multiplier. Differentiating with respect to some $|A(l,l)|^2$ where $1 \le l \le D$ gives,

$$|A_i(l,l)|^2 = \sqrt{\frac{\sigma_n^2}{\mu\sigma_s^4|\hat{\lambda}_{h_d}(l)|^2}}. \quad (27)$$

(A similar expression is obtained for cyclic prefixed block transmissions via a least squares receiver in [15]).

Substituting (22) into (25) we obtain

$$\mathfrak{L} = \sum_{q=1}^{D} \frac{\sigma_s^2 \left\{ \left(z(q) + \sum_{l=1}^{P} \frac{B^2(q,l)\sigma_n^2}{|t(l)|^2} \right) |A(q,q)|^2 \right\} + \sigma_n^2}{\sigma_s^2 |\hat{\lambda}_{h_d}(q)|^2 |A(q,q)|^2} + \mu(\sigma_s^2 \sum_{\ell=1}^{D} |A(\ell,\ell)|^2 + \sum_{k=1}^{P} |t(k)|^2 - P_t), \quad (28)$$

where $z(q)$ contains the terms of $\mathsf{E}|\tilde{\lambda}_{h_d}(q)|^2$ which are not functions of $\boldsymbol{t}$, i.e.,

$$\boldsymbol{z}(q) = \mathsf{E}|\lambda_{h_d}(q)|^2 - 2Re\mathsf{E}\lambda_{h_d}(q)\boldsymbol{\lambda}_{h_t}^{*}\boldsymbol{B}^{*}(q,:) + \boldsymbol{B}(q,:)\mathsf{E}\boldsymbol{\lambda}_{h_t}\boldsymbol{\lambda}_{h_t}^{*}\boldsymbol{B}^{*}(q,:), \quad (29)$$

and the term $\sum_{l=1}^{P} \frac{B^2(q,l)\sigma_n^2}{|t(l)|^2}$ is another way of writing the last term in equation (22). Differentiating (28) with respect to $|t(m)|^2$, $1 \leq m \leq P$, we have

$$\frac{\partial \mathfrak{L}}{\partial |t(m)|^2} = -\sum_{q=1}^{D} \frac{B^2(q,m)\sigma_n^2}{|t(m)|^4 |\hat{\lambda}_{h_d}(q)|^2} + \mu \quad (30)$$

so that equating it to zero implies that

$$|t(m)|^2 = \sqrt{\sum_{q=1}^{D} \frac{B^2(q,m)\sigma_n^2}{\mu |\hat{\lambda}_{h_d}(q)|^2}}. \quad (31)$$

To solve for μ we substitute we (27) and (31) in the power constraint, which gives

$$\sigma_s^2 \sum_{\ell=1}^{D} \sqrt{\frac{\sigma_n^2}{\mu \sigma_s^4 |\hat{\lambda}_{h_d}(\ell)|^2}} + \sum_{k=1}^{P} \sqrt{\sum_{q=1}^{D} \frac{B^2(q,k)\sigma_n^2}{\mu |\hat{\lambda}_{h_d}(q)|^2}} = P_t \quad (32)$$

which enables us to calculate

$$\mu^{\frac{1}{2}} = \frac{\sum_{\ell=1}^{D} \sqrt{\frac{\sigma_n^2}{|\hat{\lambda}_{h_d}(\ell)|^2}} + \sum_{k=1}^{P} \sqrt{\sum_{q=1}^{D} \frac{B^2(q,k)\sigma_n^2}{|\hat{\lambda}_{h_d}(q)|^2}}}{P_t}. \quad (33)$$

Note that in both equations (27) and (31), the square of the channel norm appears in the denominator. If the channel at a particular subcarrier experiences a deep fade, the norm of the channel tends to zero. When this happens the equations (27), (31) and (33) are no longer true. To prevent this from happening, we follow a strategy similar to that of [15], in that those subcarriers must be dropped from sets K and Q as applicable, before the optimization is performed. Unfortunately, while this problem is solved, subcarrier dropping results in loss of capacity.

VI. Performance

In this section, we illustrate that the joint power loading of data and pilots can offer significant advantages over power loading the data symbols only. We assume a multipath ($L = 4$) Rayleigh fading channel with Doppler of 50 Hz. The number of subcarriers is set to 1024, while the cyclic prefix length to 64. QPSK is used for modulating the data symbols. The total bandwidth is 10 MHz and 94 of the 1024 subcarriers are used for equispaced pilots. Moreover, for simplicity, linear interpolation is used to find the channel for the data subcarriers. A delayless error free feedback path is assumed between receiver and transmitter and the receiver utilizes the training subcarriers embedded in the transmission to estimate the channel. While the feedback path is delayless, here we do not assume that the CSI update takes place every time the receiver estimates the channel.

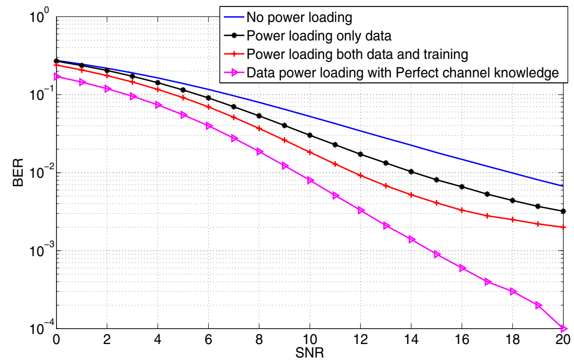

Fig. 1. BER curves when feedback is always available at transmitter.

We shall assume four different scenarios: (i) First, equal power is assigned to both pilots and data. This corresponds to a situation where we make no use of the CSI; (ii) Second, only data is power loaded, by making use of (27). Of course, the scaling factor μ in (27) is calculated using

$$\mu^{\frac{1}{2}} = \frac{\sum_{q=1}^{D} \sqrt{\frac{\sigma_n^2}{|\hat{\lambda}_{h_d}(q)|^2}}}{P_d}, \quad (34)$$

where P_d is the power given to the data symbols (set arbitrarily) while all pilots are assigned equal power. This corresponds to the precoder that minimizes the symbol mean square error for cyclic prefixed systems using least squares receiver in [15]; (iii) In the third scenario, the optimal power for the pilots and data are calculated via(27), (31) and (33); (iv) Finally, the data is power loaded according to (27) and (34) but perfect channel knowledge is assumed at both transmitter and receiver. Thus for this scenario, we do not need to transmit any pilot symbols. The receiver performs the same calculations for all cases and hence knows perfectly how much power to expect in each subcarrier in each case. Moreover, the CSI at the transmitter can be updated after every frame or every J frames depending on system configuration.

In Fig. 1, it is assumed that as soon as the channel estimate is computed at the receiver, it will be available immediately at the transmitter. We notice that at BER 10^{-2} the joint data and training algorithm produces a 2 db gain over the case where only data is power loaded, and 6 dB in the case where the CSI is unused. The graph also indicates that we are approximately 3 dB away from the perfect scenario. Figure 2 shows how the performance of the various algorithms degrades as the same CSI is used for a large number of frames. We plot the SNR required at the receiver for each algorithm to achieve a BER of 10^{-2} for different values of J. A value of 0 for J implies that

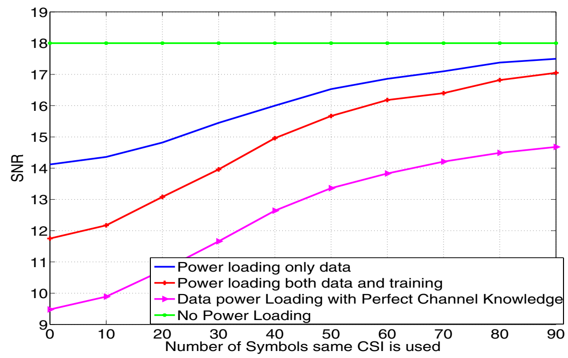

Fig. 2. SNR required by each algorithm to achieve BER of 10^{-2} when the same CSI is used for J symbols.

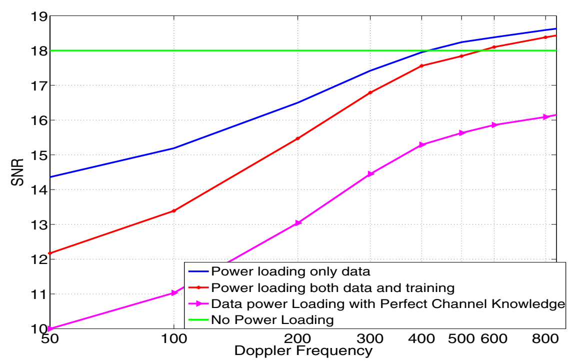

Fig. 3. SNR required by each algorithm to achieve BER of 10^{-2} for different Doppler frequencies.

the CSI is available to the transmitter as soon as the receiver estimates it. As expected, the SNR required to obtain the same BER increases with time due to outdated CSI.The effect of different Doppler spreads is illustrated in Fig. (3). Again, we plot the SNR required to achieve a BER of 10^{-2}. We have assumed that the transmitter uses the same CSI for 10 frames before it is updated. As the Doppler frequency increases we require higher SNRs to reach the same performance. This is because for high Dopplers, the channel changes too fast, during the 10 frames in which no CSI update is received. In the conventional power loading case, performance becomes worse than the case of no power loading at 415 Hz. For the proposed joint power loading algorithm, however, this condition is reached at 559 Hz Doppler.

VII. Conclusion

In this paper we have presented an algorithm that jointly power loads data and training subcarriers to achieve minimum symbol MSE for a least squares receiver. This algorithm was shown to achieve a gain of 2 dB over the conventional power loading scheme while requiring the same amount of feedback. We also show that gains can still be achieved if the CSI used is not updated for a large number of symbols. Finally, we demonstrated that the new algorithm compares favorably with the state of the art techniques for a wide range of Doppler frequencies. We conclude that using feedback information to jointly design the precoder and training matrices provides superior BER performance over techniques which power loads only the data symbols.

VIII. Acknowledgements

This work was supported in part under grant number NIJ/DOJ 2006-IJ-CX-K044.

References

[1] I. Koffman and V. Roman, “Broadband wireless access solutions based on OFDM access in IEEE 802.16,” *IEEE Communications Magazine*, vol. 40, no. 4, pp. 96-103, Apr. 2002.

[2] H.Bolcskei, “MIMO-OFDM wireless systems: basics, perspectives, and challenges,” *IEEE Wireless Communications*, vol .13, no. 4, pp. 31-37, Aug. 2006.

[3] A. Scaglione, G.B. Giannakis, and S. Barbarossa, “Redundant filterbank precoders and equalizers. I. Unification and optimal designs,” *IEEE Transactions on Signal Processing*, vol. 47, no. 7, pp. 1988-2006, Jul 1999.

[4] T. Yoo and A. Goldsmith, “Capacity of fading MIMO channels with channel estimation error,” *2004 IEEE International Conference on Communications* , vol. 2, pp. 808-813, June 2004.

[5] N. Khaled, G. Leus, C. Desset, and H.De Man, “A robust joint linear precoder and decoder MMSE design for slowly time-varying MIMO channels,” *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004.*, vol. 4, pp. 485-488, May 2004.

[6] F. Rey, M. Lamarca, and G. Vazquez, “Robust power allocation algorithms for MIMO OFDM systems with imperfect CSI,” *IEEE Transactions on Signal Processing*, vol. 53, no. 3, pp. 1070-1085, March 2005.

[7] B. Hassibi and B.M. Hochwald, “How much training is needed in multiple-antenna wireless links?,” *IEEE Transactions on Information Theory* , vol. 49, no. 4, pp. 951-963, April 2003.

[8] X. Cai and G.B. Giannakis, “Adaptive PSAM accounting for channel estimation and prediction errors,” *IEEE Transactions on Wireless Communications* , vol. 4, no. 1, pp. 246-256, Jan. 2005.

[9] D.V. Duong and G.E. Oien, “Optimal Pilot Spacing and Power in Rate-Adaptive MIMO Diversity Systems with Imperfect CSI,” *IEEE Transactions on Wireless Communications* , vol. 6, no. 3, pp. 845-851, March 2007.

[10] T. Kim and J.G. Andrews, “ Balancing Pilot and Data Power for Adaptive MIMO-OFDM Systems,” *IEEE Global Telecommunications Conference, 2006. GLOBECOM ’06.* , pp. 1-5, Nov. 2006.

[11] A. H. Sayed, *Fundamentals of Adaptive Filtering* New York: Wiley, 2003.

[12] J.H. Manton, I.Y. Mareels, and Y. Hua, “Affine precoders for reliable communications,” *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 2749-2752 vol.5, 2000.

[13] S. Ohno and G.B. Giannakis, “Optimal training and redundant precoding for block transmissions with application to wireless OFDM,” *IEEE Transactions on Communications*, vol. 50, no. 12, pp. 2113-2123, Dec 2002.

[14] A. Vosoughi and A. Scaglione, “Everything you always wanted to know about training: guidelines derived using the affine precoding framework and the CRB,” *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 940-954, March 2006

[15] Y. Ding and T.N. Davidson, Z. Luo; K.M. Wong, “Minimum BER block precoders for zero-forcing equalization,” *IEEE Transactions on Signal Processing* , vol. 51, no. 9, pp. 2410-2423, Sept. 2003.