Hindawi Publishing Corporation
EURASIP Journal on Wireless Communications and Networking
Volume 2007, Article ID 32460, 12 pages
doi:10.1155/2007/32460

Research Article

Capacity of MIMO-OFDM with Pilot-Aided Channel Estimation

Ivan Cosovic and Gunther Auer

DoCoMo Euro-Labs, Landsberger Straße 312, 80687 München, Germany

Received 31 October 2006; Revised 9 July 2007; Accepted 4 October 2007

Recommended by A. Alexiou

An analytical framework is established to dimension the pilot grid for MIMO-OFDM operating in time-variant frequency selective channels. The optimum placement of pilot symbols in terms of overhead and power allocation is identified that maximizes the training-based capacity for MIMO-OFDM schemes without channel knowledge at the transmitter. For pilot-aided channel estimation (PACE) with perfect interpolation, we show that the maximum capacity is achieved by placing pilots with maximum equidistant spacing given by the sampling theorem, if pilots are appropriately boosted. Allowing for realizable and possibly suboptimum estimators where interpolation is not perfect, we present a semianalytical method which finds the best pilot allocation strategy for the particular estimator.

1. INTRODUCTION

Systems employing multiple transmit and receive antennas, known as multiple-input multiple-output (MIMO) systems, promise significant gains in channel capacity [1–3]. Together with orthogonal frequency division multiplexing (OFDM), MIMO-OFDM is selected for the wireless local area network (WLAN) standard IEEE 802.11n [4], and for beyond 3rd generation (B3G) mobile communication systems [5].

As multiple signals are transmitted from different transmit antennas simultaneously, coherent detection requires accurate channel estimates of all transmit antennas' signals at the receiver. The most common technique to obtain channel state information is via pilot-aided channel estimation (PACE) where known training symbols termed pilots are multiplexed with data. For PACE, channel estimates are exclusively generated by means of pilot symbols, and these estimates are then processed for the detection of data symbols as if they were the true channel response. A sophisticated pilot design should strike a balance between the attainable accuracy of the channel estimate and the resources consumed by pilot symbols. An appropriate means to optimize this trade-off is to maximize the channel capacity of pilot-aided schemes. As the training overhead grows proportionally to the number of transmitted spatial streams [6], the attainable capacity gains for MIMO-OFDM are traded with the bandwidth and energy consumed by a growing number of pilot symbols to estimate the MIMO channels.

A lower bound for the attainable capacity for multiple antenna systems with pilot-aided channel estimation for the block fading channel was derived in [7]. The capacity lower bound was then used to optimize the energy allocation and the fraction of resources consumed by pilots. The capacity lower bound of [7] was extended to single carrier systems operating in frequency-selective channels [8–10], and to spatially correlated MIMO channels [11]. Furthermore, in [8, 12] the capacity achieving pilot design for MIMO-OFDM over frequency-selective channels was studied. For OFDM, pilot symbols inserted in the frequency domain (before OFDM modulation) sample the channel, allowing to recover the channel response for data bearing subcarriers by means of interpolation. This implies that, besides pilot overhead and power allocation, also the placement of pilots is to be optimized. It was found that equidistant placement of pilot symbols not only minimizes the mean squared error (MSE) of the channel estimates [13], but also maximizes the capacity [8].

All previous work deriving the capacity for training-based schemes for single-carrier [7–11] as well as for multicarrier systems [8, 12] considered time-invariant channels, where the channel was assumed static for the block of transmitted symbols. Furthermore, work on OFDM was limited to perfect interpolation [8, 12]. That is, additive white Gaussian noise (AWGN) is the only source of channel estimation errors. This implies that, in the absence of noise channel estimates perfectly match the true channel response. However,

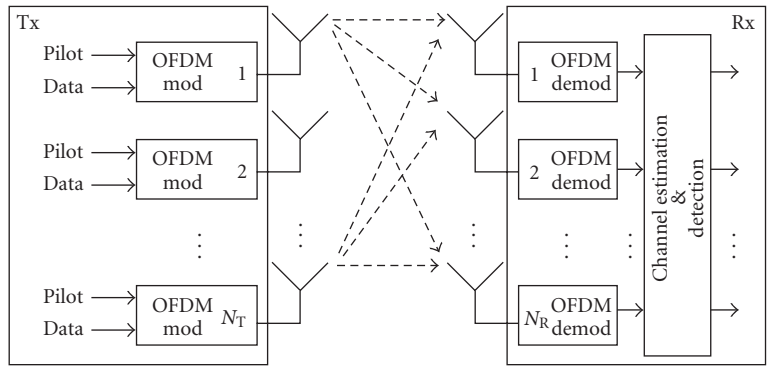

Figure 1: MIMO-OFDM system employing N_T transmit and N_R receive antennas.

from the sampling theorem it is well known that an infinite number of pilots is necessary for perfect interpolation [14]. In the context of OFDM, perfect interpolation is observed only if the channel model is sample spaced [15], that is, all channel taps are integer multiples of the sampling duration. In practice, however, pilot sequences are finite while channel taps are nonsample spaced, ultimately leading to reduced capacity bounds. Whereas in related work for OFDM only time-invariant channels and perfect interpolation are considered, in this work real world problems are taken into account, and a more realistic capacity bound is achieved.

Furthermore, previous work on maximizing the training based capacity was exclusively dedicated to minimum mean squared error (MMSE) channel estimation [7–12]. Applied to PACE the MMSE criterion finds the best tradeoff between the attainable interpolation accuracy and the mitigation of noise [16]. While MMSE estimates are favorable in terms of performance, knowledge of the 2nd order statistics is required for implementation, together with a computationally expensive matrix inversion [16]. Unlike previous work, our objective is *not* to find the optimum estimator that achieves capacity, rather we aim to identify the optimum pilot design that maximizes capacity for a given and possibly suboptimum estimator.

The present paper addresses the above mentioned limitations of previous work; its main contributions are summarized as follows.

(i) The work from [7], conducted for block fading channels, is extended to time-variant frequency-selective channels. Assuming perfect interpolation and placing pilot symbols with maximum possible distance in time and frequency still satisfying the sampling theorem, the capacity is shown to approach that for the block fading channel, provided that the size of the block is chosen according to the maximum pilot spacing imposed by the sampling theorem.

(ii) Previous work in [7, 8, 12] is extended to arbitrary linear estimators. Capacity-achieving pilot design for realizable, and possibly suboptimum, channel estimation schemes is therefore possible, as for example, interpolation by finite impulse response (FIR) filtering [16–18], discrete Fourier transform (DFT) based interpolation [19, 20], or linear interpolation [21]. We derive a closed form expression of the training based-capacity for perfect interpolation, and propose a semianalytical procedure for practical estimation techniques.

(iii) For a particular class of estimators, namely FIR interpolation filters, we demonstrate that the pilot grid that maximizes capacity is mostly independent of the chosen channel model, as long as the maximum channel delay and the maximum Doppler frequency are within a certain range. This is an appealing property, as a sophisticated pilot design should be valid for an as wide as possible range of channel conditions.

The remainder of this paper is structured as follows. In Section 2, the system model is introduced. In Section 3, the estimation error model is established and analyzed, whereas bounds on the achievable capacity of the optimized pilot design are derived in Section 4. Numerical examples in Section 5 verify the developed framework in terms of pilot boost and overhead, as well as number of transmit antennas.

2. SYSTEM MODEL

Consider a MIMO-OFDM system with N_T transmit and N_R receive antennas as illustrated in Figure 1. We assume that N_T spatial streams are transmitted and that channel knowledge is not available at the transmitter. Denote with N_c the number of used subcarriers, and with L the number of OFDM symbols per frame. OFDM modulation is performed by N_{DFT}-point ($N_{DFT} \geq N_c$) inverse DFT (IDFT), followed by insertion of a cyclic prefix (CP) of N_{CP} samples. Assuming perfect orthogonality in time and frequency, the received signal of subcarrier n of the ℓth OFDM symbol block and νth receive antenna is given by

$$Y_{n,\ell}^{(\nu)} = \sum_{\mu=1}^{N_T} \sqrt{\frac{E_d}{N_T}} X_{n,\ell}^{(\mu)} H_{n,\ell}^{(\mu,\nu)} + Z_{n,\ell}^{(\nu)}, \quad 0 \leq n < N_c, \quad 0 \leq \ell < L, \quad 0 \leq \nu < N_R. \tag{1}$$

In (1), $X_{n,\ell}^{(\mu)}$, $H_{n,\ell}^{(\mu,\nu)}$, and $Z_{n,\ell}^{(\nu)}$, denote the normalized transmitted symbol over transmit antenna μ with $E\{|X_{n,\ell}^{(\mu)}|^2\} = 1$, the

channel transfer function (CTF) between transmit antenna μ and receive antenna ν, and AWGN at the νth receive antenna with zero mean and variance N_0, respectively. An energy per transmitted data symbol of E_d/N_T and a normalized average channel gain, $E\{|H_{n,\ell}^{(\mu,\nu)}|^2\} = \sigma_H^2 = 1$ is assumed.

The discrete CTF $H_{n,\ell}^{(\mu,\nu)}$, is obtained by sampling $H^{(\mu,\nu)}(f, t)$ at frequency $f = n/T$ and time $t = \ell T_{sym}$, where $T_{sym} = (N_c + N_{CP})T_{spl}$ and $T = N_c T_{spl}$ represent the OFDM symbol duration with and without the cyclic prefix, and T_{spl} is the sample duration. Considering a frequency selective time-variant channel, modeled by a tapped delay line with Q_0 nonzero taps with channel impulse response (CIR), $h^{(\mu,\nu)}(\tau, t) = \sum_{q=1}^{Q_0} h_q^{(\mu,\nu)}(t)\delta(t - \tau_q^{(\mu,\nu)})$, the CTF is described by

$$H_{n,\ell}^{(\mu,\nu)} = H^{(\mu,\nu)}\left(\frac{n}{T}, \ell T_{sym}\right) = \sum_{q=1}^{Q_0} h_{q,\ell}^{(\mu,\nu)} \exp\left(-j2\pi\tau_q^{(\mu,\nu)}\frac{n}{T}\right), \tag{2}$$

where $h_{q,\ell}^{(\mu,\nu)} = h_q^{(\mu,\nu)}(\ell T_{sym})$ denotes the complex valued channel tap q, assumed to be constant over one OFDM symbol block, with associated tap delay $\tau_q^{(\mu,\nu)}$. Intersymbol interference is avoided by ensuring that $N_{CP}T_{spl} \geq \tau_{max}$, where τ_{max} denotes the maximum delay of the CIR. Then an arbitrary CIR is supported for which all channel taps $h_{q,\ell}^{(\mu,\nu)}$ are contained within the range $0 \leq \tau_q^{(\mu,\nu)} \leq \tau_{max}$, and the received signal is given by (1).

The 2nd-order statistics are determined by the two-dimensional (2D) correlation function $R^{(\mu,\nu)}[\Delta_n, \Delta_\ell] = E\{H_{n,\ell}^{(\mu,\nu)}(H_{n+\Delta_n,\ell+\Delta_\ell}^{(\mu,\nu)})^*\}$, composed of two independent correlation functions in frequency and time, $R^{(\mu,\nu)}[\Delta_n, \Delta_\ell] = R_f^{(\mu,\nu)}[\Delta_n]R_t^{(\mu,\nu)}[\Delta_\ell]$. Both $R_f^{(\mu,\nu)}[\Delta_n] = E\{H_{n,\ell}^{(\mu,\nu)}(H_{n+\Delta_n,\ell}^{(\mu,\nu)})^*\}$ and $R_t^{(\mu,\nu)}[\Delta_\ell] = E\{H_{n,\ell}^{(\mu,\nu)}(H_{n,\ell+\Delta_\ell}^{(\mu,\nu)})^*\}$ are strictly band-limited [18]. That is, the inverse Fourier transform of $R_f^{(\mu,\nu)}[\Delta_n]$ described by the power delay profile is essentially nonzero in the range $[0, \tau_{max}]$, where τ_{max} is the maximum channel delay. Likewise, the Fourier transform of $R_t^{(\mu,\nu)}[\Delta_\ell]$ describing time variations due to mobile velocities is given by the Doppler power spectrum, nonzero within $[-f_{D,max}, f_{D,max}]$, where $f_{D,max}$ is the maximum Doppler frequency. No further assumptions regarding the distribution of $h^{(\mu,\nu)}(\tau, t)$ are imposed. To this end, the CIR may possibly be nonsample spaced, that is, tap delays $\tau_q^{(\mu,\nu)}$ in (2) may *not* be placed at integer multiples of the sampling duration.

In order to recover the transmitted information, pilot symbols are commonly used for channel estimation. Channel estimation schemes for MIMO-OFDM based on the least squares (LS) and MMSE criterion are studied in [22, 23] and [6, 24, 25], respectively. We assume that channel state information about all $N_T \times N_R$ channels is required at the receiver. To enable this, pilots belonging to different transmit antennas are orthogonally separated in time and/or frequency. Thus, the problem of MIMO-OFDM channel estimation breaks down to estimating the channel of a single antenna OFDM system. Note, there are other possibilities to orthogonally separate the pilots, but they lead to higher complexity and/or at least the same pilot overhead [6]. Furthermore, pilots belonging to the same transmit antenna are equidistantly spaced in time and frequency within the OFDM frame [17]. This is motivated by the findings in [8], where it is shown that equidistant placement of pilots minimizes the harmonic mean of the MSE of channel estimates over all subcarriers and thus maximizes the capacity. Figure 2 illustrates the resulting placement of pilots for four spatial streams, arranged in a rectangular shaped pattern. Extension to other regular pilot patters, such as a diamond shaped grid [26], is straightforward.

Resources are constraint to bandwidth and energy. Unlike the orthogonally separated pilots, data symbols are spatially multiplexed. One frame is assigned $N_p^{(\mu)}$ pilot and N_d data symbols per spatial dimension which amounts to (cf., Figure 2)

$$N_c L = N_d + \sum_{\mu=1}^{N_T} N_p^{(\mu)}. \tag{3}$$

The resulting pilot overhead of the μth antenna is defined by

$$\Omega_p^{(\mu)} = \frac{N_p^{(\mu)}}{N_c L}. \tag{4}$$

With an energy per transmitted data symbol of E_d/N_T, the total transmit energy over all N_T antennas equals

$$E_{tot} = E_d N_d + \sum_{\mu=1}^{N_T} E_p^{(\mu)} N_p^{(\mu)}, \tag{5}$$

where $E_p^{(\mu)}$ is the energy per pilot symbol of the μth transmit antenna.

The accuracy of the channel estimates may be improved by a pilot boost $S_p^{(\mu)}$. With an energy per transmitted pilot set to $E_p^{(\mu)} = S_p^{(\mu)} E_d$, the signal-to-noise ratio (SNR) at the input of the channel estimation unit is improved by a factor of $S_p^{(\mu)}$. On the other hand, the useful transmit energy of the payload information is reduced, if the overall transmit energy in (5) is kept constant. The energy dedicated to pilot symbols is determined by the pilot overhead per antenna $\Omega_p^{(\mu)}$, and the pilot boost per antenna $S_p^{(\mu)}$. Including the pilot overhead, the ratio of the energy per symbol E_d of a system with pilots, to the energy per symbol E_0 of an equivalent system with the same frame size $N_c \times L$, same transmit energy E_{tot}, but without pilot symbols is

$$\frac{E_d}{E_0} = \frac{1}{1 + \sum_{\mu=1}^{N_T} \Omega_p^{(\mu)}\left(S_p^{(\mu)} - 1\right)}. \tag{6}$$

The ratio E_d/E_0 is a measure for the pilot insertion loss relative to a reference system assuming no overhead due to pilots. Note, (6) is obtained exploiting (4), (5), and the constraint $E_{tot} = N_c L E_0$.

In the following, we assume that for each transmit antenna the same number of pilot symbols and the same boosting level are used, that is, $N_p = N_p^{(\mu)}$, $\Omega_p = \Omega_p^{(\mu)}$, and $S_p = S_p^{(\mu)}$,

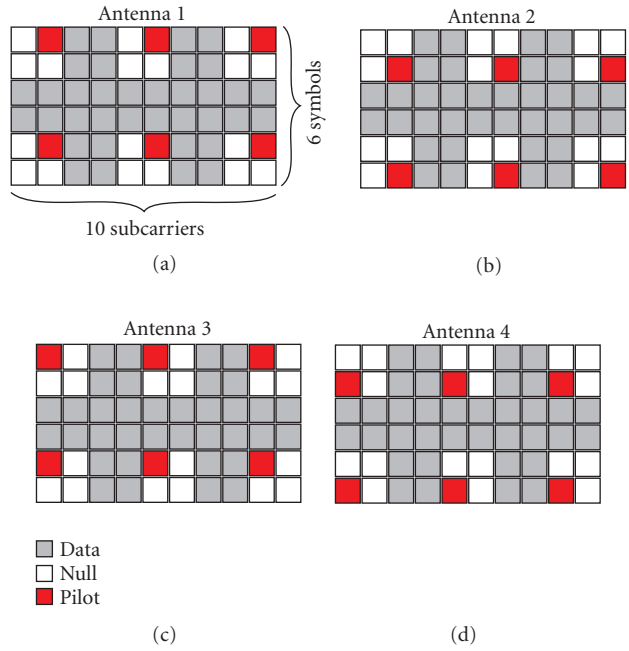

FIGURE 2: Example for placing orthogonal pilots over antenna-specific OFDM subframes, $N_T = 4$, $N_p^{(1)} = N_p^{(2)} = N_p^{(3)} = N_p^{(4)} = 6$, and $N_d = 36$.

$\mu = \{1, \ldots, N_T\}$. Now, the total pilot overhead in (4) amounts to $N_T\Omega_p$ and (6) simplifies to

$$\frac{E_d}{E_0} = \frac{1}{1 + N_T\Omega_p(S_p - 1)}. \tag{7}$$

3. ESTIMATION ERROR MODELING

The channel estimation unit outputs an estimate of the CTF, $H_{n,\ell}^{(\mu,\nu)}$, denoted by $\hat{H}_{n,\ell}^{(\mu,\nu)} = \mathbf{w}^H \tilde{\boldsymbol{y}}^{(\mu,\nu)}$. Let M_f and M_t denote the number of pilot symbols in frequency and time used to generate $\hat{H}_{n,\ell}^{(\mu,\nu)}$. The $M_tM_f \times 1$ column vector $\tilde{\boldsymbol{y}}^{(\mu,\nu)}$ contains the received pilots from transmit antenna μ to receive antenna ν. The $M_tM_f \times 1$ column vector $\mathbf{w}$ represents an arbitrary linear estimator.

3.1. Parametrization of the MSE

Channel estimation impairments are quantified by the MSE of the estimation error $\varepsilon_{n,\ell}^{(\mu,\nu)} = H_{n,\ell}^{(\mu,\nu)} - \hat{H}_{n,\ell}^{(\mu,\nu)}$. With the assumption that all transmit and receive antennas are mutually uncorrelated, the MSE is independent of μ and ν, denoted by $\sigma_\varepsilon^2[n,\ell] = E[|\varepsilon_{n,\ell}^{(\mu,\nu)}|^2]$. The MSE of an arbitrary 2D pilot-aided scheme is given by

$$\begin{aligned}\sigma_\varepsilon^2[n,\ell] &= E\left[\,|\,\varepsilon_{n,\ell}^{(\mu,\nu)}\,|^2\right] = E\left[\,|\,H_{n,\ell}^{(\mu,\nu)} - \hat{H}_{n,\ell}^{(\mu,\nu)}\,|^2\right] \\ &= E\left[\,|\,H_{n,\ell}^{(\mu,\nu)}\,|^2\right] - 2\Re\left\{\mathbf{w}^H \mathbf{r}_{\tilde{\boldsymbol{y}}H}^{(\mu,\nu)}[n,\ell]\right\} + \mathbf{w}^H \mathbf{R}_{\tilde{\boldsymbol{y}}\tilde{\boldsymbol{y}}}^{(\mu,\nu)} \mathbf{w}.\end{aligned} \tag{8}$$

The 2D correlation functions $\mathbf{r}_{\tilde{\boldsymbol{y}}H}^{(\mu,\nu)}[n,\ell] = E\{\tilde{\boldsymbol{y}}^{(\mu,\nu)}(H_{n,\ell}^{(\mu,\nu)})^*\}$ and $\mathbf{R}_{\tilde{\boldsymbol{y}}\tilde{\boldsymbol{y}}}^{(\mu,\nu)} = E\{\tilde{\boldsymbol{y}}^{(\mu,\nu)}(\tilde{\boldsymbol{y}}^{(\mu,\nu)})^H\}$ represent the cross-correlation between $\tilde{\boldsymbol{y}}^{(\mu,\nu)}$ and the desired response $H_{n,\ell}^{(\mu,\nu)}$, and the autocorrelation matrix of the received pilots, $\tilde{\boldsymbol{y}}^{(\mu,\nu)}$, respectively, [16]. The autocorrelation matrix is composed of $\mathbf{R}_{\tilde{\boldsymbol{y}}\tilde{\boldsymbol{y}}}^{(\mu,\nu)} = \mathbf{R}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}}^{(\mu,\nu)} + \mathbf{I}/\gamma_p$, where $\mathbf{R}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}} = E\{\tilde{\boldsymbol{h}}^{(\mu,\nu)}(\tilde{\boldsymbol{h}}^{(\mu,\nu)})^H\}$ is the autocorrelation matrix of the CTF at pilot positions excluding the AWGN term, and $\mathbf{I}$ denotes the identity matrix, all of dimension $M_fM_t \times M_fM_t$. With the pilot insertion loss of (7), the SNR at pilot positions amounts to $\gamma_p = S_pE_d/N_0 = \gamma_0 S_p/(1 + N_T\Omega_p(S_p - 1))$, where $\gamma_0 = E_0/N_0$ denotes the SNR of a reference system assuming perfect channel knowledge and no overhead due to pilots.

The MSE in (8) is dependent on n and ℓ. In order to allow for a tractable model, we choose to average the MSE over the entire sequence, so $\sigma_\varepsilon^2[n,\ell] \to \sigma_\varepsilon^2$.

The channel estimates generated by a linear estimator $\mathbf{w}$ can be decomposed into a signal and noise part, denoted by $\hat{H}_{n,\ell}^{(\mu,\nu)} = \mathbf{w}^H\tilde{\boldsymbol{h}}^{(\mu,\nu)} + \mathbf{w}^H\tilde{\boldsymbol{z}}^{(\nu)}$, where $\tilde{\boldsymbol{h}}^{(\mu,\nu)}$ and $\tilde{\boldsymbol{z}}^{(\nu)}$ account for CTF and AWGN vectors at pilot positions. Likewise, the estimation error $\varepsilon_{n,\ell}^{(\mu,\nu)}$ can be separated into an interpolation error $H_{n,\ell}^{(\mu,\nu)} - \mathbf{w}^H\tilde{\boldsymbol{h}}^{(\mu,\nu)}$ and a noise error $\mathbf{w}^H\tilde{\boldsymbol{z}}^{(\nu)}$. Assuming that CTF $H_{n,\ell}^{(\mu,\nu)}$ and AWGN $Z_{n,\ell}^{(\nu)}$ are uncorrelated, the MSE also separates into a noise and interpolation error:

$$\begin{aligned}\sigma_\varepsilon^2 &= E\left[\,|\,H_{n,\ell}^{(\mu,\nu)} - \mathbf{w}^H\tilde{\boldsymbol{h}}^{(\mu,\nu)} - \mathbf{w}^H\tilde{\boldsymbol{z}}^{(\nu)}\,|^2\right] \\ &= E\left[\,|H_{n,\ell}^{(\mu,\nu)} - \mathbf{w}^H\tilde{\boldsymbol{h}}^{(\mu,\nu)}\,|^2\right] + E\left[\,|\,\mathbf{w}^H\tilde{\boldsymbol{z}}^{(\nu)}\,|^2\right].\end{aligned} \tag{9}$$

We note that this separation of the MSE is possible for any linear estimator. The noise part $\sigma_{\mathrm{n}}^2 = E[|\mathbf{w}^H \tilde{\boldsymbol{z}}^{(\nu)}|^2]$ is inversely proportional to the SNR and is given by

$$\sigma_{\mathrm{n}}^2 = \frac{\mathbf{w}^H\mathbf{w}}{\gamma_{\mathrm{p}}} = \frac{1}{G_{\mathrm{n}}\gamma_0}\cdot\frac{1+N_{\mathrm{T}}\Omega_{\mathrm{p}}(S_{\mathrm{p}}-1)}{S_{\mathrm{p}}}, \tag{10}$$

where $G_{\mathrm{n}} = 1/(\mathbf{w}^H\mathbf{w})$ defines the estimator gain. According to (8) the variance of the interpolation error is determined by

$$\begin{aligned}\sigma_{\mathrm{i}}^2 &= E\Big[\,|H_{n,\ell}^{(\mu,\nu)} - \mathbf{w}^H\tilde{\boldsymbol{h}}^{(\mu,\nu)}|^2\Big] \\ &= E\Big[\,|H_{n,\ell}^{(\mu,\nu)}|^2\Big] - 2\Re\Big\{\mathbf{w}^H\mathbf{r}_{\tilde{\boldsymbol{h}}H}^{(\mu,\nu)}\Big\} + \mathbf{w}^H\mathbf{R}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}}^{(\mu,\nu)}\mathbf{w}.\end{aligned} \tag{11}$$

3.2. Equivalent system model and effective SNR

In order to derive a model taking into account channel estimation errors, we assume a receiver that processes the channel estimates $\widehat{H}_{n,\ell}^{(\mu,\nu)}$ as if these were the true CTF. The effect of channel estimation errors on the received signal in (1) is described by the equivalent system model:

$$\Upsilon_{n,\ell}^{(\mu,\nu)} = \sum_{\mu=1}^{N_{\mathrm{T}}}\sqrt{\frac{E_{\mathrm{d}}}{N_{\mathrm{T}}}}X_{n,\ell}^{(\mu)}\breve{H}_{n,\ell}^{(\mu,\nu)} + \underbrace{\sum_{\mu=1}^{N_{\mathrm{T}}}\sqrt{\frac{E_{\mathrm{d}}}{N_{\mathrm{T}}}}X_{n,\ell}^{(\mu)}\varepsilon_{n,\ell}^{(\mu,\nu)} + Z_{n,\ell}^{(\nu)}}_{\eta_{n,\ell}^{(\mu,\nu)}}, \tag{12}$$

where $\eta_{n,\ell}^{(\mu,\nu)}$ denotes the effective noise term with zero mean and variance $\sigma_\eta^2 = N_0 + E_{\mathrm{d}}\sigma_\varepsilon^2$. Apart from the increased noise term $\eta_{n,\ell}^{(\mu,\nu)}$, channel estimation impairments affect the equivalent system model (12) by distortions in the signal part of $\breve{H}_{n,\ell}^{(\mu,\nu)} = \mathbf{w}^H\tilde{\boldsymbol{h}}^{(\mu,\nu)}$. The useful signal energy observed at the receiver is given by

$$\sigma_{\breve{H}}^2 = E\Big[\,|\mathbf{w}^H\tilde{\boldsymbol{h}}^{(\mu,\nu)}|^2\Big] = \mathbf{w}^H\mathbf{R}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}}\mathbf{w}. \tag{13}$$

Now the effective SNR including channel estimation of the equivalent system model (12) yields

$$\gamma = E_{\mathrm{d}}\frac{\sigma_{\breve{H}}^2}{\sigma_\eta^2} = \frac{E_{\mathrm{d}}\sigma_{\breve{H}}^2}{N_0 + E_{\mathrm{d}}\sigma_\varepsilon^2}. \tag{14}$$

An important difference to the equivalent system model devised by [7] is the definition of $\sigma_{\breve{H}}^2$ in (13). In [7] the estimate $\widehat{H}_{n,\ell}^{(\mu,\nu)} = \mathbf{w}^H\tilde{\boldsymbol{y}}^{(\mu,\nu)}$ replaces $\breve{H}_{n,\ell}^{(\mu,\nu)} = \mathbf{w}^H\tilde{\boldsymbol{h}}^{(\mu,\nu)}$ in (12), thus containing contributions from the CTF and noise, so that $\sigma_{\widehat{H}}^2 = \sigma_{\breve{H}}^2 + N_0/E_{\mathrm{d}}$. Instead, our model with $\sigma_{\breve{H}}^2$ in (13) exclusively captures the signal part of the channel estimate. As the model of [7] was tailored for an MMSE estimator with $\sigma_{\widehat{H}}^2 = 1 - \sigma_\varepsilon^2$, meaningful results are produced. However, the model of [7] implies that the noise term contained in $\sigma_{\widehat{H}}^2$ contributes to the useful signal energy of the effective SNR γ in (14). This becomes problematic at low SNR, when the equivalent system model is to be applied to other than MMSE estimators. For instance, consider an unbiased estimator with $\mathbf{w}^H\mathbf{R}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}}\mathbf{w} = 1$ and unitary estimator gain $\mathbf{w}^H\mathbf{w} = 1$. Then with the model of [7]: $\sigma_{\widehat{H}}^2 = 1 + N_0/E_{\mathrm{d}}$, so for low SNR, $N_0 \to \infty$, the effective SNR in (14) approaches $\gamma = 1/2$, which is clearly a contradiction. On the other hand, by using (13) we get $\sigma_{\breve{H}}^2 = 1$ and $\gamma \to 0$ as $N_0 \to \infty$, so our model in (12) produces meaningful results for low SNR. In any case, in the high SNR regime with $N_0 \ll E_{\mathrm{d}}$ both models converge and we get $\sigma_{\breve{H}}^2 \approx \sigma_{\widehat{H}}^2$.

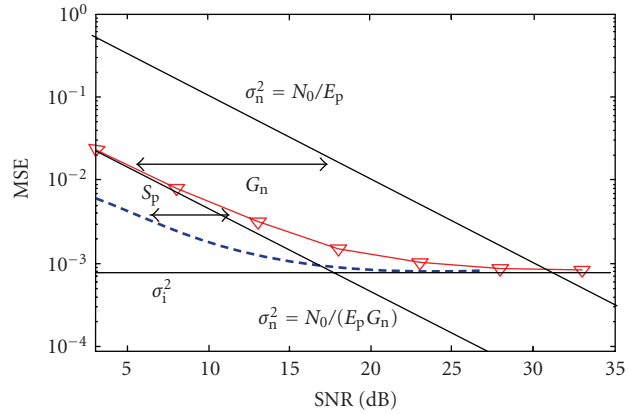

Figure 3: Parametrization of an MSE estimation curve.

Inserting (7) into the effective SNR γ in (14) and computing the ratio γ_0/γ quantifies the SNR degradation due to channel estimation errors, given by

$$\Delta\gamma = \frac{\gamma_0}{\gamma} = \frac{1}{\sigma_{\breve{H}}^2}(1 + N_{\mathrm{T}}\Omega_{\mathrm{p}}(S_{\mathrm{p}}-1) + \sigma_\varepsilon^2\gamma_0). \tag{15}$$

Substituting the MSE $\sigma_\varepsilon^2 = \sigma_{\mathrm{n}}^2 + \sigma_{\mathrm{i}}^2$ into (15), with σ_{n}^2 being expressed in the parametrized form of (10), the loss in SNR due to channel estimation can be transformed to

$$\Delta\gamma = \frac{1}{\sigma_{\breve{H}}^2}(1 + N_{\mathrm{T}}\Omega_{\mathrm{p}}(S_{\mathrm{p}}-1))\cdot\left(1 + \frac{1}{G_{\mathrm{n}}S_{\mathrm{p}}}\right) + \frac{\sigma_{\mathrm{i}}^2}{\sigma_{\breve{H}}^2}\gamma_0. \tag{16}$$

In the following, a fixed estimator $\mathbf{w}$ is considered where the estimator coefficients are computed once and are not adapted for changing channel conditions. Then, according to (16) the performance penalty due to channel estimation is fully determined by two SNR independent parameters, the estimator gain G_{n} and the interpolation error σ_{i}^2. On the other hand, allowing for an SNR dependent estimator, $\mathbf{w} = \mathbf{w}(\gamma)$, the parameters G_{n} and σ_{i}^2 would be strictly speaking only valid for one particular SNR value γ. A prominent example for an SNR dependent estimator is the MMSE estimator, known as Wiener filter [27]. In this case, the SNR for which G_{n} and σ_{i}^2 lead to a maximum $\Delta\gamma$ in (16) should be used, so to maintain a certain performance under worst case conditions.

The MSE of a fixed estimator $\mathbf{w}$ is plotted in Figure 3. At low SNR, the MSE is dominated by the noise error σ_{n}^2.

Hence, the MSE linearly decreases with the SNR. At high SNR the MSE experiences an error floor caused by the SNR independent interpolation error. A pilot boost is only effective to reduce the noise part of the MSE in (10), while the interpolation error, σ_{i}^2 in (11), remains unaffected. This is shown in Figure 3, where a pilot boost shifts the MSE S_{p} dB to the left.

4. CAPACITY ANALYSIS

The ergodic channel capacity that includes channel estimation and pilot insertion losses when the channel is not known at transmitter can be lower bounded by [7]

$$C \geq (1 - N_{\mathrm{T}}\Omega_{\mathrm{p}})E\left[\log_2 \det\left(\mathbf{I}_{N_{\mathrm{R}}} + \frac{\mathbf{H}_{n,\ell}\mathbf{H}_{n,\ell}^H}{N_{\mathrm{T}}}\frac{\gamma_0}{\Delta\gamma}\right)\right], \quad (17)$$

where $\mathbf{I}_{N_{\mathrm{R}}}$ is the $N_{\mathrm{R}} \times N_{\mathrm{R}}$ identity matrix and the CTF is defined by

$$\mathbf{H}_{n,\ell} = \begin{pmatrix} H_{n,\ell}^{(1,1)} & \cdots & H_{n,\ell}^{(N_{\mathrm{T}},1)} \\ \vdots & \ddots & \vdots \\ H_{n,\ell}^{(1,N_{\mathrm{R}})} & \cdots & H_{n,\ell}^{(N_{\mathrm{T}},N_{\mathrm{R}})} \end{pmatrix}. \quad (18)$$

In (17), the expectation is taken over the frequency and time dimension of $\mathbf{H}_{n,\ell}$, that is, over indices n and ℓ. The capacity penalty due to the pilot-aided channel estimation is characterized by two factors: the SNR loss due to estimation errors, $\Delta\gamma$ from (15) or (16), and the loss in spectral efficiency due to resources consumed by pilot symbols, $N_{\mathrm{T}}\Omega_{\mathrm{p}}$. Inserting (16) into (17) we obtain

$$C \geq (1-N_{\mathrm{T}}\Omega_{\mathrm{p}})E\left[\log_2 \det\left(\mathbf{I}_{N_{\mathrm{R}}}+\frac{\mathbf{H}_{n,\ell}\mathbf{H}_{n,\ell}^H}{N_{\mathrm{T}}} \cdot \frac{\gamma_0\sigma_{\breve{H}}^2}{(1+N_{\mathrm{T}}\Omega_{\mathrm{p}}(S_{\mathrm{p}}-1))\cdot(1+1/(G_{\mathrm{n}}S_{\mathrm{p}}))+\sigma_{\mathrm{i}}^2\gamma_0}\right)\right]. \quad (19)$$

An important requirement for the capacity lower bound to become tight is that the signal and noise terms in the equivalent system model of (12) are uncorrelated [7]. Unfortunately, for arbitrary linear estimators $\mathbf{w}$ this may not be the case, as the interpolation error $H_{n,\ell}^{(\mu,\nu)} - \breve{H}_{n,\ell}^{(\mu,\nu)}$, with $\breve{H}_{n,\ell}^{(\mu,\nu)} = \mathbf{w}^H\tilde{\boldsymbol{h}}^{(\mu,\nu)}$, introduces correlations between the effective noise term $\eta_{n,\ell}^{(\mu,\nu)}$ and the CTF $H_{n,\ell}^{(\mu,\nu)}$. On the other hand, the noise part of the estimation error $\mathbf{w}^H\tilde{\boldsymbol{z}}^{(\nu)}$ is statistically independent of both $H_{n,\ell}^{(\mu,\nu)}$ and $\breve{H}_{n,\ell}^{(\mu,\nu)}$. Therefore, for perfect interpolation ($\breve{H}_{n,\ell}^{(\mu,\nu)} = H_{n,\ell}^{(\mu,\nu)}$ and $\sigma_{\mathrm{i}}^2 = 0$), the signal part $\breve{H}_{n,\ell}^{(\mu,\nu)}$ and the effective noise term $\eta_{n,\ell}^{(\mu,\nu)}$ in (12) become statistically independent. To this end, one condition for a tight capacity bound is $\sigma_{\mathrm{n}}^2 \gg \sigma_{\mathrm{i}}^2$, so to ensure that $\breve{H}_{n,\ell}^{(\mu,\nu)}$ and $\eta_{n,\ell}^{(\mu,\nu)}$ are sufficiently decorrelated.

In the following, we focus on the problem of capacity maximization. By doing so, we consider

(i) pilot boost S_{p},
(ii) pilot overhead Ω_{p},
(iii) number of transmit antennas N_{T}

as optimization parameters such that the capacity is maximized.

4.1. Optimum pilot boost

The effect of a pilot boost is twofold: first, the estimation error decreases; second, the energy dedicated to pilot symbols increases. So, there clearly exists an optimum pilot boost S_{p} which minimizes the loss in SNR due to channel estimation, $\Delta\gamma$ in (16), and thus maximizes the system capacity in (19).

The optimum pilot boost for the parametrized estimation error model is obtained by differentiating $\Delta\gamma$ from (16) or C from (19) with respect to S_{p} and setting the result to zero. This results in

$$S_{\mathrm{p,opt}} = \sqrt{\frac{1 - N_{\mathrm{T}}\Omega_{\mathrm{p}}}{N_{\mathrm{T}}\Omega_{\mathrm{p}}G_{\mathrm{n}}}}. \quad (20)$$

The optimum pilot boost $S_{\mathrm{p,opt}}$ is seen to increase if less pilots are used and/or less transmit antennas are used, but decreases with growing estimator gain, G_{n}. In any case, a pilot boost is only effective to reduce the noise part of the MSE σ_{n}^2 in (10), while the interpolation error σ_{i}^2 in (11) remains unaffected. Hence, the attainable gains of a pilot boost diminish with growing σ_{i}^2 and SNR γ_0, as deduced from $\Delta\gamma$ in (16) or C in (19), although the optimum pilot boost $S_{\mathrm{p,opt}}$ in (20) is independent of both SNR and σ_{i}^2.

The loss in SNR for the optimally chosen pilot boost (20) yields

$$\Delta\gamma\big|_{S_{\mathrm{p}}=S_{\mathrm{p,opt}}} = \frac{1}{\sigma_{\breve{H}}^2}\left(\sqrt{1-N_{\mathrm{T}}\Omega_{\mathrm{p}}}+\sqrt{\frac{N_{\mathrm{T}}\Omega_{\mathrm{p}}}{G_{\mathrm{n}}}}\right)^2 + \frac{\sigma_{\mathrm{i}}^2}{\sigma_{\breve{H}}^2}\gamma_0, \quad (21)$$

whereas the capacity becomes

$$C\big|_{S_{\mathrm{p}}=S_{\mathrm{p,opt}}} = (1-N_{\mathrm{T}}\Omega_{\mathrm{p}})E\left[\log_2 \det\left(\mathbf{I}_{N_{\mathrm{R}}}+\frac{\mathbf{H}_{n,\ell}\mathbf{H}_{n,\ell}^H}{N_{\mathrm{T}}} \cdot \frac{\gamma_0\sigma_{\breve{H}}^2}{\left(\sqrt{1-N_{\mathrm{T}}\Omega_{\mathrm{p}}}+\sqrt{N_{\mathrm{T}}\Omega_{\mathrm{p}}/G_{\mathrm{n}}}\right)^2+\sigma_{\mathrm{i}}^2\gamma_0}\right)\right]. \quad (22)$$

As both G_{n} and σ_{i}^2 depend on Ω_{p}, an analytical solution for Ω_{p} that maximizes (22) is a task of formidable complexity which is not pursued here. Instead, for the special case of perfect interpolation ($\sigma_{\mathrm{i}}^2 = 0$), we derive the optimum pilot overhead Ω_{p} in closed form, whereas we propose a semianalytical procedure for the general case.

4.2. Ideal lowpass interpolation filter (LPIF)

Motivated by previous work where it was shown that equidistant placement of pilot symbols minimizes the MSE [13],

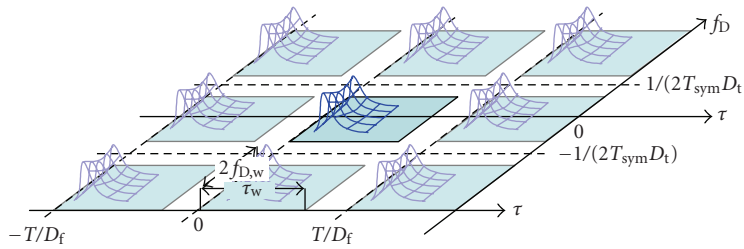

Figure 4: Filter transfer function of an ideal 2D low-pass interpolation filter.

as well as maximizes the capacity [8], we focus on channel estimation by interpolation with equispaced pilots in time and frequency. By using a scattered pilot grid the received OFDM frame is sampled in two dimensions, with rate D_f/T and $D_t T_{sym}$ in frequency and time, respectively.

An ideal lowpass interpolation filter (LPIF) is characterized by the 2D rectangular shaped filter transfer function

$$W(\tau, f_D) = \sum_{n=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} w_{n,\ell}\, e^{-j2\pi(n\tau/T + \ell f_D T_{sym})} = \begin{cases} 1, & 0 \le \tau \le \tau_w,\ |f_D| \le f_{D,w}, \\ 0, & \tau_w < \tau \le \dfrac{T}{D_f},\ f_{D,w} < |f_D| \le \dfrac{1}{2T_{sym}D_t}, \end{cases} \tag{23}$$

where $w_{n,\ell}$ denotes the filter coefficient of pilot subcarrier n and OFDM symbol ℓ. The filter parameters τ_w and $f_{D,w}$ specify the cut-off region of the filter. The transformation to the time (delay) and Doppler domains is described by a discrete time Fourier transform (DTFT) [14], between the variable pairs $n \to \tau$ and $\ell \to f_D$. Due to sampling, $W(\tau, f_D)$ is periodically repeated at intervals $[0, T/D_f]$ and $[-1/(2T_{sym}D_t), 1/(2T_{sym}D_t)]$. This is illustrated in Figure 4 where the filter transfer function of an ideal LPIF is drawn in the 2D plane.

Applied to PACE the LPIF is to be designed so that the spectral components of the CTF $H_{n,\ell}^{(\mu,\nu)}$ which are nonzero within the range $[0, \tau_{max}]$ and $[-f_{D,max}, f_{D,max}]$ pass the filter undistorted, while spectral components outside this range are blocked. Furthermore, the pilot spacings D_f and D_t must be sufficiently small to prevent spectral overlap between the filter passband and its aliases. Hence, in order to reconstruct the signal the sampling theorem requires that [18]

$$\tau_{max} \le \tau_w < \frac{T}{D_f}, \qquad f_{D,max} \le f_{D,w} < \frac{1}{2D_t T_{sym}}. \tag{24}$$

The filter parameters τ_w and $f_{D,w}$ represent the maximum assumed delay of the channel and the maximum assumed Doppler frequency, according to worst case channel conditions, as indicated in Figure 4.

Applied to the MSE analysis in Section 3.1, an ideal LPIF has some appealing properties as follows.

(i) Provided that (24) is satisfied, the interpolation error diminishes, $\sigma_i^2 = 0$; that is, the LPIF resembles perfect interpolation. Consequently, the MSE is equivalent to $\sigma_\varepsilon^2 = \sigma_n^2 = \mathbf{w}^H\mathbf{w}/\gamma_p$ in (10). Furthermore, the MSE becomes independent of the subcarrier and OFDM symbol indices n and ℓ. Hence, no deviation over n and ℓ is observed, $\sigma_\varepsilon^2[n,\ell] = \sigma_\varepsilon^2$. As opposed to the general case of linear estimators in Section 3.1, averaging over n and ℓ is not required.

(ii) Due to perfect interpolation, the useful signal power in (13) becomes $\sigma_{\underset{H}{\sim}}^2 = \sigma_H^2 = 1$, that is, the LPIF produces unbiased estimates.

(iii) As the number of filter coefficients in frequency and time approach infinity, $\{M_f, M_t\} \to \infty$, the pilot overhead becomes $\Omega_p = 1/(D_f D_t)$.

(iv) In the high SNR regime, the performance of an ideal LPIF asymptotically approaches the MMSE [28]. In general, however, the MSE of the ideal LPIF is strictly larger than the MMSE.

By invoking Parseval's theorem [14], the MSE can be transformed to

$$\sigma_\varepsilon^2 = \frac{\mathbf{w}^H\mathbf{w}}{\gamma_p} = \frac{D_f D_t T_{sym}}{\gamma_p T} \int_0^{T/D_f} \int_{-1/(2D_t T_{sym})}^{1/(2D_t T_{sym})} |W(\tau, f_D)|^2\, d\tau df_D. \tag{25}$$

The MSE, σ_ε^2, is determined by the fraction of the AWGN suppressed by the filter. Inserting (23) and solving (25) yields

$$\sigma_\varepsilon^2 = \frac{1}{\gamma_p G_n} = \frac{1}{\gamma_p \beta_f \beta_t} = \frac{c_w^2}{\gamma_p \Omega_p} \tag{26}$$

with $c_w = \sqrt{2\tau_w f_{D,w} T_{sym}/T}$. The factors $\beta_f = T/(D_f\tau_w)$ and $\beta_t = 1/(2D_t f_{D,w} T_{sym})$ are a measure for the amount of oversampling in frequency and time, with respect to minimum sampling rates T/D_f and $1/(2D_t T_{sym})$, required by the sampling theorem in (24). The MSE is inversely proportional to the oversampling factors β_f and β_t, as well as the pilot overhead $\Omega_p = 1/(D_f D_t)$. Hence, increasing the pilot overhead directly improves the MSE.

4.2.1. Capacity of PACE with perfect 2D interpolation

The expression for the MSE in (26) establishes a relation between estimator gain and pilot overhead, $G_n = \Omega_p/c_w^2$, that allows to maximize the channel capacity in closed form. Moreover, the effective signal and noise terms in the equivalent system model (12) become statistically independent, ensuring a tight capacity bound in (22).

For an ideal LPIF, the optimum pilot boost (20) can be conveniently expressed as

$$S_{\mathrm{p,opt}} = c_{\mathrm{w}} \frac{\sqrt{1 - N_{\mathrm{T}}\Omega_{\mathrm{p}}}}{\sqrt{N_{\mathrm{T}}\Omega_{\mathrm{p}}}}. \tag{27}$$

Inserting $G_{\mathrm{n}} = \Omega_{\mathrm{p}}/c_{\mathrm{w}}^2$ into (21) and after some algebraic transformations the SNR loss for $S_{\mathrm{p}} = S_{\mathrm{p,opt}}$ becomes

$$\Delta\gamma|_{S_{\mathrm{p}}=S_{\mathrm{p,opt}}} = \left(\sqrt{1 - N_{\mathrm{T}}\Omega_{\mathrm{p}}} + c_{\mathrm{w}}\sqrt{N_{\mathrm{T}}}\right)^2. \tag{28}$$

This means that $\Delta\gamma|_{S_{\mathrm{p}}=S_{\mathrm{p,opt}}}$ is minimized by the maximum pilot overhead Ω_{p}, that is, all transmitted symbols are dedicated to pilots $\Omega_{\mathrm{p}} = 1/N_{\mathrm{T}}$. However, in this case the capacity becomes zero. In fact, the capacity is maximized by selecting the smallest pilot overhead $\Omega_{\mathrm{p,min}}$ which still satisfies the sampling theorem

$$C_{\max} = (1 - N_{\mathrm{T}}\Omega_{\mathrm{p,min}}) E\left[\log_2 \det\left(\mathbf{I}_{N_{\mathrm{R}}} + \frac{\mathbf{H}_{n,\ell}\mathbf{H}_{n,\ell}^H}{N_{\mathrm{T}}} \cdot \frac{\gamma_0}{\left(\sqrt{1 - N_{\mathrm{T}}\Omega_{\mathrm{p,min}}} + c_{\mathrm{w}}\sqrt{N_{\mathrm{T}}}\right)^2}\right)\right], \tag{29}$$

where $\Omega_{\mathrm{p,min}} = 1/(D_{\mathrm{f,max}}D_{\mathrm{t,max}})$ is attained by the maximum pilot spacings which satisfy (24), $D_{\mathrm{f,max}} = \lfloor T/\tau_{\mathrm{w}} \rfloor$ and $D_{\mathrm{t,max}} = \lfloor 1/(2f_{\mathrm{D,w}}T_{\mathrm{sym}}) \rfloor$, where $\lfloor x \rfloor$ is the largest integer equal or smaller than x. To prove (29) it can be easily checked that $C_{\max}$ is a monotonically decreasing function with respect to Ω_{p}, with the global maximum at $\Omega_{\mathrm{p}} = 0$. Hence, (29) is maximized by $\Omega_{\mathrm{p,min}}$, since $\Delta\gamma|_{S_{\mathrm{p}}=S_{\mathrm{p,opt}}}$ is only valid for pilot grids which satisfy the sampling theorem in (24).

By ignoring the rounding effects and thus approximating $D_{\mathrm{f,max}}D_{\mathrm{t,max}} \approx c_{\mathrm{w}}^2$, we obtain $\Omega_{\mathrm{p,min}} \approx c_{\mathrm{w}}^2$, that is, the effects of channel estimation errors for PACE are completely described by $\Omega_{\mathrm{p,min}}$. Interestingly, the estimator gain now approaches unity, $G_{\mathrm{n}} = 1$ and the SNR loss becomes

$$\Delta\gamma_{\min} = \left(\sqrt{1 - N_{\mathrm{T}}\Omega_{\mathrm{p,min}}} + \sqrt{N_{\mathrm{T}}\Omega_{\mathrm{p,min}}}\right)^2. \tag{30}$$

Furthermore, (27) and (29) can be approximated by

$$S_{\mathrm{p,opt}} = \frac{\sqrt{1 - N_{\mathrm{T}}\Omega_{\mathrm{p}}}}{\sqrt{N_{\mathrm{T}}\Omega_{\mathrm{p}}}}, \tag{31}$$

$$C_{\max} \approx (1 - N_{\mathrm{T}}\Omega_{\mathrm{p,min}}) E\left[\log_2 \det\left(\mathbf{I}_{N_{\mathrm{R}}} + \frac{\mathbf{H}_{n,\ell}\mathbf{H}_{n,\ell}^H}{N_{\mathrm{T}}} \cdot \frac{\gamma_0}{\left(\sqrt{1 - N_{\mathrm{T}}\Omega_{\mathrm{p,min}}} + \sqrt{N_{\mathrm{T}}\Omega_{\mathrm{p,min}}}\right)^2}\right)\right]. \tag{32}$$

Finally, it turns out that the fraction of energy dedicated to data relative to the overall transmit energy becomes

$$\frac{E_{\mathrm{d}}N_{\mathrm{d}}}{E_0 N_{\mathrm{c}} L} = \frac{E_{\mathrm{d}}}{E_0}(1 - N_{\mathrm{T}}\Omega_{\mathrm{p,min}}) \approx \frac{\sqrt{1 - N_{\mathrm{T}}\Omega_{\mathrm{p,min}}}}{\sqrt{1 - N_{\mathrm{T}}\Omega_{\mathrm{p,min}}} + \sqrt{N_{\mathrm{T}}\Omega_{\mathrm{p,min}}}}, \tag{33}$$

where the ratio E_{d}/E_0 is defined in (6). Interestingly, (30) and (33) are equivalent to the results obtained for the block fading channel (see [7, equation (34)]). Applied to MIMO-OFDM, the block fading assumption translates to a time/frequency area the channel is assumed constant. Then, the same results apply given that the *interval, the channel is constant* for the block fading assumption in [7], is replaced by the *maximum pilot spacing that satisfies the sampling theorem*, $D_{\mathrm{f,max}}D_{\mathrm{t,max}} = 1/\Omega_{\mathrm{p,min}}$. This means that the capacity lower bound of [7] is extended to the more general case of time-variant frequency selective channels.

An interesting observation can be made by setting

$$N_{\mathrm{T}}\Omega_{\mathrm{p,min}} = \frac{1}{2}. \tag{34}$$

By devoting half of the resources to pilot symbols we obtain from (31) and (33), respectively,

$$S_{\mathrm{p,opt}} = 1, \qquad \frac{E_{\mathrm{d}}N_{\mathrm{d}}}{E_0 N_{\mathrm{c}} L} = \frac{1}{2}. \tag{35}$$

In case half of the frame is devoted to training purposes, in order to maximize capacity, pilots should not be boosted, and consequently half of the transmit energy is invested on pilots. A similar conclusion is also provided in [7] assuming a block fading channel.

4.2.2. Number of transmit antennas

Suppose that N_{T}' out of the N_{T} transmit antennas are used for communication. Inserting N_{T}' for N_{T} in the capacity expression for MIMO-OFDM with optimum pilot grid, $C_{\max}$ in (32), the number of transmit antennas that maximizes channel capacity $C_{\max}$ is given by [7, 29]

$$N_{\mathrm{T}}' = \min\left\{N_{\mathrm{T}}, N_{\mathrm{R}}, \frac{1}{2\Omega_{\mathrm{p,min}}}\right\}. \tag{36}$$

Several important conclusions with respect to the capacity maximization in MIMO-OFDM can be drawn from (36).

(i) If $N_{\mathrm{T}} = N_{\mathrm{R}} = 1/(2\Omega_{\mathrm{p,min}})$, from (34) and (35) it follows that pilots should not be boosted, that is, they should be of equal energy as the data symbols..
(ii) The amount of training should not exceed half of the OFDM frame.
(iii) The number of transmit and receive antennas should be equal.

4.3. Semianalytical approach for pilot grid design

The optimal pilot grid that maximizes the channel capacity is derived for an ideal LPIF in Section 4.2. For realizable

estimators, we propose the following procedure to obtain the optimum pilot grid.

(i) Specify the filter parameters τ_w and $f_{D,w}$ so that the relation in (24) is satisfied.
(ii) Choose maximum possible pilot spacings and estimator dimensions M_f and M_t, that maintain a certain interpolation error σ_i^2. This determines the minimum pilot overhead $\Omega_{p,min}$, and the estimator gain $G_n = 1/(\mathbf{w}^H\mathbf{w})$.
(iii) Determine the optimum pilot boost $S_{p,opt}$ using (20).
(iv) Calculate the optimum number of transmit antennas using (36).

Considering (i), in a well-designed OFDM system the maximum channel delay τ_{max} should not exceed the cyclic prefix. Therefore, it is reasonable to assume $\tau_w = T_{CP}$. In addition, $f_{D,w}$ is set according to the maximum Doppler frequency expected in a certain propagation scenario.

Considering (ii), this condition is imposed to keep σ_i^2 sufficiently low. The impact of σ_i^2 on the SNR penalty in (21) becomes negligible if $\sigma_i^2 < \epsilon_{th}/\gamma_w$, where ϵ_{th} is a small positive constant and γ_w denotes the largest expected SNR. This condition effectively enforces a sufficient degree of oversampling. That is, $\Omega_{p,min}$ is required to be larger than the theoretical minimum.

The condition on the interpolation error σ_i^2 in step (ii) serves another important requirement. The fact that σ_i^2 is negligible in the SNR region of interest ensures that the useful signal part and the estimation error in the equivalent signal model in (12) become uncorrelated, so that the capacity bound in (22) becomes tight.

Considering (iii), this step ensures that the capacity in (22) is maximized, given that in step (ii) $\Omega_{p,min}$ is appropriately chosen, so that the interpolation error σ_i^2 is sufficiently small. As a formal proof appears difficult, we verify through a numerical example in Section 5 that the proposed semianalytical procedure indeed maximizes capacity.

5. NUMERICAL RESULTS

An OFDM system with $N_c = 512$ subcarriers, and a cyclic prefix of duration $T_{CP} = 64 \cdot T_{spl}$, is employed. One frame consists of $L = 65$ OFDM symbols, and it is assumed that channel estimation is carried out, after all pilots of one frame have been received. The signal bandwidth is 20 MHz, which translates to a sampling duration of $T_{spl} = 50$ ns. This results in an OFDM symbol duration of $T_{sym} = 35.97\,\mu$s of which the cyclic prefix is $T_{CP} = 3.2\,\mu$s. A high mobility scenario is considered with velocities up to 300 km/h. At 5 GHz carrier frequency this translates to a normalized maximum Doppler frequency of $f_{D,max}T_{sym} \leq 0.04$.

Channel estimation unit

Since pilots belonging to different transmit antennas are orthogonally separated in time and/or frequency, the problem of MIMO-OFDM channel estimation breaks down to estimating the channel of a single antenna OFDM system. A cascaded channel estimator consisting of two one-dimensional (1D) estimators termed 2×1D PACE is implemented. 2×1D PACE performs only slightly worse than optimal 2D PACE, while being significantly less complex [16].

Table 1: Power delay profile of WINNER channel model C2 [31].

Delay (ns)	0	5	135	160	215	260	385
Power (dB)	−0.5	0	−3.4	−2.8	−4.6	−0.9	−6.7
Delay (ns)	400	530	540	650	670	720	750
Power (dB)	−4.5	−9.0	−7.8	−7.4	−8.4	−11	−9.0
Delay (ns)	800	945	1035	1185	1390	1470	
Power (dB)	−5.1	−6.7	−12.1	−13.2	−13.7	−19.8	

The estimator was implemented by a Wiener interpolation filter (WIF) with model mismatch [16]. The filter coefficients in frequency and time are generated assuming a uniformly distributed power delay profile and Doppler power spectrum, nonzero within the range $[0, \tau_w]$ and $[-f_{D,w}, f_{D,w}]$. Furthermore, the average SNR at the filter input, γ_w, is required, which should be equal or larger than actual average SNR, so $\gamma_w \geq \gamma_0$. To generate the filter coefficients, we set $\tau_w = T_{CP}$, $f_{D,w} = 0.04T_{spl}$ and $\gamma_w = 30$ dB. With these parameters, the sampling theorem in (24) requires for the pilot spacings in frequency and time $D_f < 8$ and $D_t \leq 12$.

The WIF with model mismatch is closely related to an LPIF, and therefore inherits many of its properties—signals with spectral components within the range $[0, \tau_w]$ and $[0, f_{D,max}]$ pass the filter undistorted, while spectral components outside this range are blocked. In fact, it was shown in [30] that for an infinite number of coefficients, $\{M_f, M_t\} \to \infty$, the mismatched WIF approaches an ideal LPIF.

Results

The performance of a channel estimation unit generally depends on the chosen channel model. On the other hand, the optimum pilot grid and the associated channel estimation unit is expected to operated in a wide variety of channel conditions. Hence, it is important to test the performance of the considered estimators for various channel models. In Figure 5 the channel estimation MSE determined by (8) is plotted for the pilot grid $D_f = 6$, $D_t = 8$, and filter orders $M_f = 16$, $M_t = 9$. The following channel models are considered:

Chn A: IST-WINNER channel model C2 for typical urban propagation environments [31]; the power delay profile (PDP) is shown in Table 1;
Chn B: flat fading channel with PDP $\rho(\tau) = \delta(\tau - T_{CP}/2)$;
Chn C: 2-tap channel with PDP $\rho(\tau) = (\delta(\tau) + \delta(\tau - T_{CP}))/2$;
Chn D: uniformly distributed PDP nonzero within the range $[0, T_{CP}]$. This is the channel used to generate the WIF coefficients, that is, the WIF is matched to *Chn D*.

For all models, the independent fading taps are generated using Jakes' model [32] with $f_{D,max}T_{sym} = 0.033$, corresponding to a velocity of 250 km/h at 5 GHz carrier frequency. It is seen in Figure 5 that the MSE is virtually independent

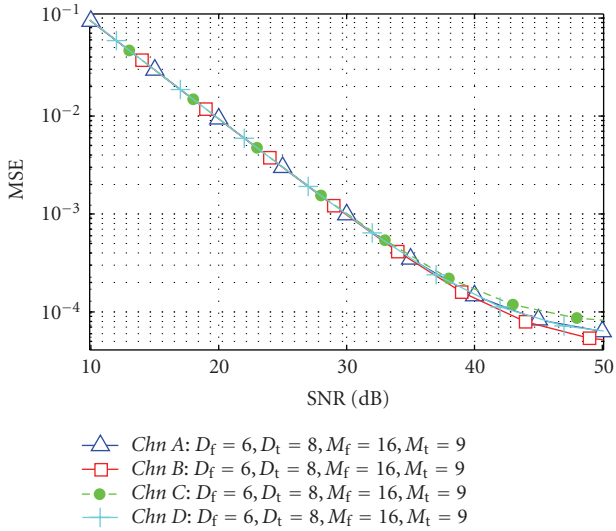

FIGURE 5: MSE versus SNR of 2 × 1D-PACE for various channel models, $D_f = 6$, $D_t = 8$, $M_f = 16$, $M_t = 9$.

of the particular channel model, although the PDPs of the considered channels cover an extensive range of possible propagation scenarios. Note that *Chn C* is the worst-case channel, since its two taps are placed at the closest position with respect to the cutoff regions of the WIF (compare with Figure 4). Likewise, *Chn B* is the best-case channel, as its single tap is located right in the center of the filter passband.

The optimum pilot grid for the considered MIMO-OFDM system is assembled in the following as discussed in Section 4.3. All results are plotted for *Chn A* with normalized maximum Doppler $f_{D,\max} T_{sym} = 0.033$. We note that results in Figure 5 suggest that the identified optimum pilot grid is valid for any channel model with $\tau_{\max} \leq \tau_w$, $f_{D,\max} \leq f_{D,w}$ and $\gamma_0 \leq \gamma_w$.

From the set of allowable D_f and D_t, the following candidate grids are selected: $D_f = \{4, 6\}$ and $D_t = \{4, 8\}$, which translates to the oversampling factors $\beta_f = \{2, 1.25\}$ and $\beta_t = \{3, 1.5\}$. The filter order in time direction was set equal to the number of pilots per frame, so $M_t = \{17, 9\}$. In frequency direction, on the other hand, the number of pilots is $N_c/D_f = 85$ and 128, respectively, allowing for much higher filter orders M_f.

Figure 6 shows the interpolation error σ_i^2 (in Figure 6(a)) by computing (11) and the estimator gain $G_n = 1/(\mathbf{w}^H\mathbf{w})$ (in Figure 6(b)) against the filter order in frequency M_f for $M_t = \{9, 17\}$ and various pilot grids (parameters D_f and D_t). Provided that $M_f \geq 12$ we observe that for all pilot grids, $\gamma_w \sigma_i^2 < 0.1$ with $\gamma_0 \leq \gamma_w = 30$ dB. Therefore, the impact of σ_i^2 on $\Delta\gamma$ is negligible (less than 0.5 dB). By setting $M_f = 16$ in the following, none of the considered grids can be ruled out at this point.

In Figure 7, the channel capacity versus pilot boost S_p of an 8 × 8 MIMO-OFDM system with different pilot grids is depicted at SNR $\gamma_0 = 10$ dB. The plots are obtained by inserting σ_i^2 and $G_n = 1/(\mathbf{w}^H\mathbf{w})$ obtained in Figure 6 into the capacity expression (19) assuming different pilot grids. It is

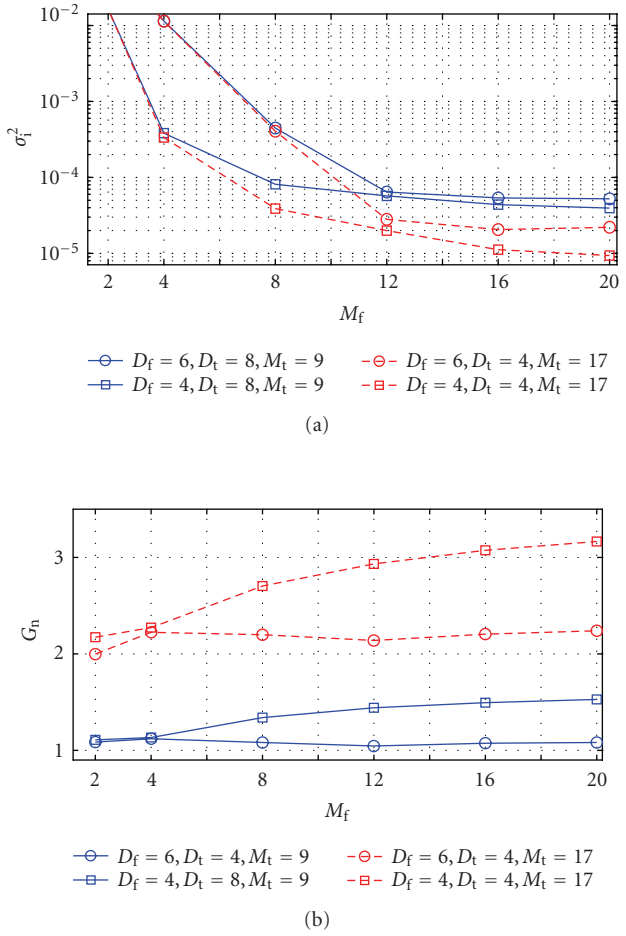

FIGURE 6: Interpolation error σ_i^2 and estimator gain G_n at an SNR of $\gamma_0 = 30$ dB against the filter order in frequency M_f.

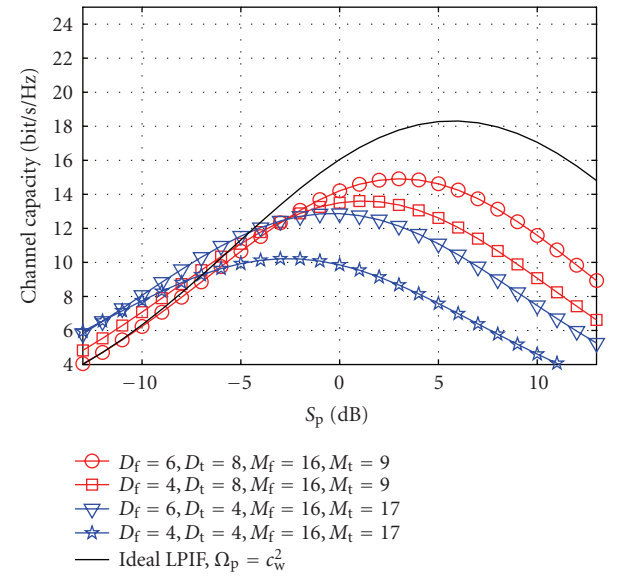

FIGURE 7: Capacity versus pilot boost for 8 × 8 MIMO-OFDM system with different pilot grids at an SNR of $\gamma_0 = 10$ dB.

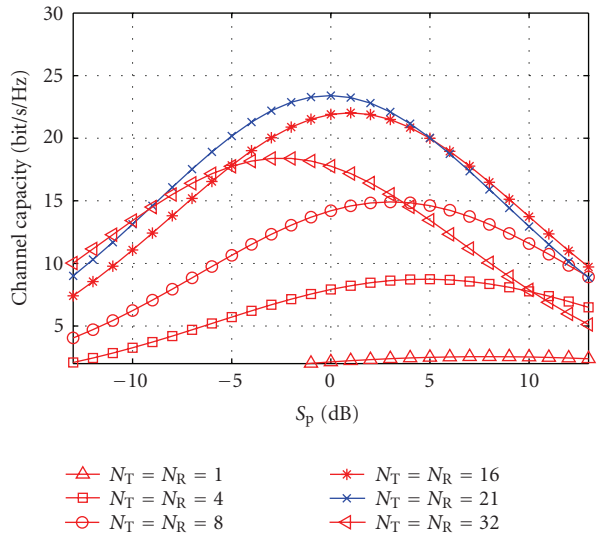

FIGURE 8: Capacity versus pilot boost for $N \times N$ MIMO-OFDM system for different number of antennas N, SNR $\gamma_0 = 10$ dB, $D_f = 6$, $D_t = 8$, $M_f = 16$, $M_t = 9$.

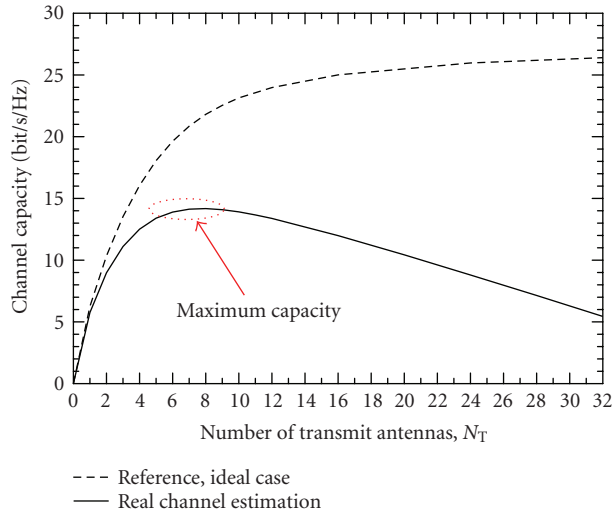

FIGURE 9: Capacity versus number of transmit antennas N_T for MIMO-OFDM system with $N_R = 8$ receive antennas, SNR $\gamma_0 = 10$ dB, $D_f = 6$, $D_t = 8$, $M_f = 16$, $M_t = 9$.

seen that the most bandwidth efficient grid ($D_f = 6, D_t = 8$) maximizes capacity. Furthermore, maximum capacity C_{max} for all grids is achieved for those S_p that satisfy (20). This plot confirms the proposed semianalytical procedure described in Section 4.3. As reference, the capacity assuming an ideal LPIF is also plotted in Figure 7. A significant gap in capacity between the ideal LPIF relative to the realizable estimators is visible. This is mainly due to the fact that a realizable filter does not exhibit a rectangular filter transfer function. This inevitably requires a higher pilot overhead and also reduces the attainable estimator gain.

The channel capacity versus pilot boost, S_p, of an $N \times N$ MIMO-OFDM system with pilot grid $D_f = 6$, $D_t = 8$, filter orders $M_f = 16$, $M_t = 9$ and for different number of transmit/receive antennas, $N = N_T = N_R$, is depicted in Figure 8. Again, the SNR is set to $\gamma_0 = 10$ dB. Maximum capacity is achieved for $N_T = 21$ yielding $N_T\Omega_p \approx 1/2$ and $S_p = 0$ dB. This confirms the analytical results from Section 4.2.2 as well as the proposed semianalytical procedure from Section 4.3. The significance of these results lie in the fact that results of [7] are generalized to MIMO-OFDM operating in dynamic channels and to arbitrary linear estimators.

The channel capacity versus the number of transmit antennas N_T of an $N_T \times 8$ MIMO-OFDM system for grid $D_f = 6$, $D_t = 8$, filter orders $M_f = 16$, $M_t = 9$, at SNR $\gamma_0 = 10$ dB is shown in Figure 9. The plots were generated using the capacity expression in (22) including the optimum pilot boost $S_{p,opt}$ according to (20). As a reference, capacity of the corresponding system assuming perfect channel estimation and no loss due to pilots is shown. It can be observed that for $N_T \approx 8$ maximum capacity is achieved, corresponding to the conclusion from Section 4.2.2. For higher values the reduction in available bandwidth due to the pilot insertion dominates, lowering the achievable capacity.

6. CONCLUSIONS

In this paper, a framework for pilot grid design in MIMO-OFDM was developed and used to determine the pilot spacing and boost, so to maximize the capacity of the target MIMO-OFDM system, including channel estimation errors and pilot overhead. The analysis show that the previously derived capacity lower bound for a block fading channel is also valid for MIMO-OFDM over time-varying frequency-selective channels. The derived bound applies to perfect interpolation, which essentially requires infinitely long pilot sequences and filter coefficients. Furthermore, a semianalytical procedure was proposed to maximize the capacity for realizable and possibly suboptimum channel estimation schemes.

ACKNOWLEDGMENTS

This work has been performed in the framework of the IST project IST-4-027756 WINNER (World Wireless Initiative New Radio), which is partly funded by the European Union. This paper was presented in part at the IEEE International Conference on Communications (ICC-07), Glasgow, UK, June 2007.

REFERENCES

[1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[3] H. Bolcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Transactions on Communications*, vol. 50, no. 2, pp. 225–234, 2002.

[4] R. Van Nee, V. K. Jones, G. Awater, A. Van Zelst, J. Gardner, and G. Steele, "The 802.11n MIMO-OFDM standard for wireless LAN and beyond," *Wireless Personal Communications*, vol. 37, no. 3-4, pp. 445–453, 2006.

[5] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, "Physical layer aspects for evolved Universal Terrestrial Radio Access (UTRA)," June 2006.

[6] G. Auer, "Analysis of pilot-symbol aided channel estimation for OFDM systems with multiple transmit antennas," in *Proceedings of IEEE International Conference on Communications (ICC '04)*, vol. 6, pp. 3221–3225, Paris, France, June 2004.

[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, 2003.

[8] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of training for frequency-selective block-fading channels," *IEEE Transactions on Information Theory*, vol. 48, no. 8, pp. 2338–2353, 2002.

[9] S. Adireddy and L. Tong, "Optimal placement of known symbols for slowly varying frequency-selective channels," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1292–1296, 2005.

[10] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "On the capacity of frequency-selective channels in training-based transmission schemes," *IEEE Transactions on Signal Processing*, vol. 52, no. 9, pp. 2572–2583, 2004.

[11] O. Simeone and U. Spagnolini, "Lower bound on training-based channel estimation error for frequency-selective block-fading Rayleigh MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 11, pp. 3265–3277, 2004.

[12] S. Ohno and G. B. Giannakis, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 2138–2145, 2004.

[13] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Transactions on Consumer Electronics*, vol. 44, no. 3, pp. 1122–1128, 1998.

[14] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, USA, 2nd edition, 1999.

[15] J.-J. van de Beek, O. Edfors, M. Sandell, S. Wilson, and P. Börjesson, "On channel estimation in OFDM systems," in *Proceedings of the 45th IEEE Vehicular Technology Conference (VTC '95)*, vol. 2, pp. 815–819, Chicago, Ill, USA, July 1995.

[16] P. Höher, S. Kaiser, and P. Robertson, "Pilot-symbol-aided channel estimation in time and frequency," in *Proceedings of the IEEE Global Telecommunications Conference (Globecom '97)*, vol. 4, pp. 90–96, Phoenix, Ariz, USA, November 1997.

[17] P. Höher, "TCM on frequency selective land-mobile radio channels," in *Proceedings of the 5th Tirrenia International Workshop on Digital Communications*, pp. 317–328, Tirrenia, Italy, September 1991.

[18] F. Sanzi and J. Speidel, "An adaptive two-dimensional channel estimator for wireless OFDM with application to mobile DVB-T," *IEEE Transactions on Broadcasting*, vol. 46, no. 2, pp. 128–133, 2000.

[19] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 4, pp. 1207–1215, 2000.

[20] Y.-H. Yeh and S.-G. Chen, "DCT-based channel estimation for OFDM systems," in *Proceedings of IEEE International Conference on Communications (ICC '04)*, vol. 4, pp. 2442–2446, Paris, France, June 2004.

[21] M.-H. Hsieh and C.-H. Wei, "Channel estimation for OFDM systems based on comb-type pilot arrangement in frequency selective fading channels," *IEEE Transactions on Consumer Electronics*, vol. 44, no. 1, pp. 217–225, 1998.

[22] Y. Li, N. Seshadri, and S. Ariyavisitakul, "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 3, pp. 461–471, 1999.

[23] Y. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Transactions on Wireless Communications*, vol. 1, no. 1, pp. 67–75, 2002.

[24] W. G. Jeon, K. H. Paik, and Y. S. Cho, "Two-dimensional MMSE channel estimation for OFDM systems with transmitter diversity," in *Proceedings of the 54th IEEE Vehicular Technology Conference (VTC '01 fall)*, vol. 3, pp. 1682–1685, Atlantic City, NJ, USA, October 2001.

[25] M. Speth, "LMMSE channel estimation for MIMO OFDM," in *Proceedings of the 8th International OFDM Workshop (InOWo '03)*, Hamburg, Germany, September 2003.

[26] J.-W. Choi and Y.-H. Lee, "Optimum pilot pattern for channel estimation in OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2083–2088, 2005.

[27] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, Englewood Cliffs, NJ, USA, 1993.

[28] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers*, John Wiley & Sons, NewYork, NY, USA, 2nd edition, 1998.

[29] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, 2002.

[30] G. Auer and E. Karipidis, "Pilot aided channel estimation for OFDM: a separated approach for smoothing and interpolation," in *Proceedings of IEEE International Conference on Communications (ICC '05)*, vol. 4, pp. 2173–2178, Seoul, Korea, May 2005.

[31] IST-4-027756 WINNER II, "D1.1.2 WINNER II channel models," September 2007.

[32] W. C. Jakes, *Microwave Mobile Communications*, John Wiley & Sons, New York, NY, USA, 1974.