

MMSE and Adaptive Prediction of Time-Varying Channels for OFDM Systems

Dieter Schafhuber, *Student Member, IEEE*, and Gerald Matz, *Member, IEEE*

Abstract—We propose decision-directed channel predictors for orthogonal frequency-division multiplexing (OFDM) communications over time-varying channels. Channel prediction is capable of yielding up-to-date channel state information even without regular transmission of pilot symbols. It is thus potentially useful for delay-free equalization, antenna combining, space-time decoding, adaptive modulation, adaptive power control, and adaptive transmit antenna diversity. We derive a minimum mean-square error channel predictor and its efficient discrete Fourier transform implementation. Furthermore, we develop adaptive predictors that do not require any statistical prior knowledge and are able to track nonstationary channel and noise statistics. Simulation results involving an OFDM receiver in which channel prediction is used for delay-free equalization demonstrate the excellent performance of the proposed techniques even for fast time-varying channels.

Index Terms—Adaptive predictors, channel prediction, equalization, multicarrier modulation, orthogonal frequency-division multiplexing (OFDM), time-varying channels.

I. INTRODUCTION

ORTHOGONAL frequency-division multiplexing (OFDM) is an efficient modulation scheme for broadband wireless communications [1]–[4]. OFDM is being commercially applied for wireless local area networks (IEEE 802.11a and HIPERLAN/2), terrestrial digital audio broadcasting (DAB-T), and terrestrial digital video broadcasting (DVB-T), and it is being considered for wireless broadband access systems by the IEEE 802.16 work group. Conventional OFDM systems use rectangular transmit and receive pulses and a cyclic prefix. Generalized *pulse-shaping* OFDM systems have been proposed in [1] and [5]–[8].

Most OFDM systems use coherent detection, which has approximately a 3-dB signal-to-noise (SNR) gain over differential techniques [9] but requires channel state information at the receiver. Channel state information at the receiver and/or transmitter is also necessary for techniques such as antenna combining, space-time decoding, adaptive modulation, adaptive power control, and adaptive transmit antenna diversity.

Manuscript received March 19, 2003; revised August 23, 2003 and November 20, 2003; accepted January 9, 2004. The editor coordinating the review of this paper and approving it for publication is I. B. Collings. This work was supported by FWF under Grant P15156 and by the EU-IST project ANTIUM (IST-2000-26222).

D. Schafhuber is with BMW Forschung und Technik, 80788 Munich, Germany and also with the Institute for Communications and Radio-Frequency Engineering, Vienna University of Technology, Vienna 1040, Austria (e-mail: Dieter.Schafhuber@ieee.org).

G. Matz is with the Institute for Communications and Radio-Frequency Engineering, Vienna University of Technology, Vienna 1040, Austria (e-mail: gmatz@aurora.nt.tuwien.ac.at).

Digital Object Identifier 10.1109/TWC.2004.843055

A widely explored approach for obtaining channel state information in OFDM systems is training data-based least squares (LS) channel estimation [10], [11] or minimum mean-square error (MMSE) channel estimation [12]–[17] (the latter requires nominal or estimated second-order channel statistics). The training data consists either of pilot symbols that are continually multiplexed into the data stream (e.g., $\approx 10\%$ of the total transmitted symbols in DVB-T), or, for packet transmission, of a training data block at the beginning of each packet (e.g., two OFDM symbols in IEEE 801.11a and HIPERLAN/2). Channel estimation has the drawback of potentially yielding *outdated* channel state information, which can degrade system performance significantly and/or prevent the application of promising techniques that require up-to-date channel estimates (e.g., adaptive modulation). Furthermore, training data-based channel estimation has the disadvantage of a reduced net data rate and sometimes increased latencies.

An alternative approach is decision-directed channel estimation (e.g., [18]). Here, training data is replaced by previously detected data. Thus, virtually all transmit symbols are available for channel estimation, resulting in potentially increased performance at the cost of increased complexity. However, decision-directed channel estimation performs poorly for fast time-varying channels [18].

In this paper, we propose novel schemes for decision-directed MMSE and adaptive *channel prediction* in OFDM systems [19], [20] that feature the following advantages.

- 1) Our channel predictors yield accurate up-to-date channel state information without additional latencies, even for fast time-varying channels or large prediction horizons.
- 2) Due to the decision-directed operation mode, no continual transmission of training data is required; typically, a single known OFDM symbol suffices for initialization.
- 3) Thanks to efficient fast Fourier transform (FFT) implementations, the computational complexity of the proposed channel predictors is very low.
- 4) Our channel prediction schemes enable key techniques (such as adaptive modulation, adaptive power control, and adaptive transmit diversity) that improve the system capacity and/or link reliability of wireless OFDM systems.
- 5) The adaptive versions of our channel predictors do not require any statistical prior knowledge and are able to track nonstationary channel and noise statistics.

We note that pilot-symbol-assisted MMSE channel prediction was proposed and applied to adaptive loading for HIPERLAN/2

in [21]. Furthermore, in a non-OFDM context, the prediction of fading channels and applications were previously investigated in [22]–[26].

The rest of this paper is organized as follows. Section II describes the OFDM system model. In Section III, we develop linear MMSE channel predictors for random time-varying channels. In Section IV, we propose adaptive versions of our channel predictor. Simulation results are provided in Section V.

II. OFDM SYSTEM MODEL

We consider a pulse-shaping OFDM system with K subcarriers (e.g., $K = 64$ for IEEE 802.11a, HIPERLAN/2 and $K = 8192$ for DVB-T). At the transmitter, a bit stream D_i is coded and mapped to transmit symbols $B_{n,k}$ ($n \in \mathbb{Z}$ is the OFDM symbol index and $k \in \{0, \dots, K-1\}$ is the subcarrier index). The discrete-time baseband transmit signal is given by [5]–[8]

$$s[m] = \sum_{n=-\infty}^{\infty} \sum_{k=0}^{K-1} B_{n,k} g_{n,k}[m] \quad (1a)$$

with

$$g_{n,k}[m] \triangleq g[m - nN] e^{j2\pi k(m - nN)/K}. \quad (1b)$$

Here, $g[m]$ is a transmit pulse and N denotes the OFDM symbol duration. The received signal is

$$x[m] = (\mathbb{H}_0 s)[m] + u[m] = \sum_{l=0}^L h_0[m, l] s[m - l] + u[m] \quad (2)$$

where $h_0[m, l]$ is the impulse response of the linear time-varying channel \mathbb{H}_0 , L is the channel's maximum delay, and $u[m]$ is additive noise. From $x[m]$, the demodulator computes the K sequences

$$X_{n,k} = \langle x, f_{n,k} \rangle \triangleq \sum_{m=-\infty}^{\infty} x[m] f_{n,k}^*[m] \quad (3a)$$

with

$$f_{n,k}[m] \triangleq f[m - nN] e^{j2\pi k(m - nN)/K} \quad (3b)$$

where $f[m]$ is a receive pulse. Throughout this paper, we use the term OFDM also for the case $g[m] \neq f[m]$, which is sometimes referred to as *biorthogonal* frequency division multiplexing.

For a conventional OFDM system using rectangular transmit and receive pulses and a cyclic prefix of length L_{cp} , the symbol duration is $N = K + L_{cp}$, the transmit pulse is $g[n] = 1/\sqrt{K}$ for $n = -L_{cp}, \dots, K-1$ and zero elsewhere, and the receive pulse is $f[n] = 1/\sqrt{K}$ for $n = 0, \dots, K-1$ and zero elsewhere. Here, the modulator (1) and demodulator (3) can be efficiently implemented by means of an (I)FFT [2]. However, efficient FFT-based implementations are also available in the general (pulse-shaping) case [6].

A. System Relation

Combining (1)–(3), the system's input–output relation is obtained as [5], [6]

$$X_{n,k} = \sum_{n'=-\infty}^{\infty} \sum_{k'=0}^{K-1} \langle \mathbb{H}_0 g_{n',k'}, f_{n,k} \rangle B_{n',k'} + Z_{n,k} \quad (4)$$

with $Z_{n,k} \triangleq \langle u, f_{n,k} \rangle$. For $(n, k) \neq (n', k')$, the coefficients $\langle \mathbb{H}_0 g_{n',k'}, f_{n,k} \rangle$ describe the intersymbol interference (ISI) and intercarrier interference (ICI) introduced by the channel \mathbb{H}_0 . For an ideal channel (i.e., no delays and Doppler shifts) and in the absence of noise, we have perfect demodulation, $X_{n,k} = B_{n,k}$, if and only if the transmit pulse $g[m]$ and receive pulse $f[m]$ satisfy the (bi)orthogonality condition

$$\langle g_{n',k'}, f_{n,k} \rangle = \delta[n - n'] \delta[k - k']. \quad (5)$$

Validity of (5) requires $N/K \geq 1$ [5], [8]. In practical systems, N/K is between 1.03 and 1.25.

Real-world channels are *underspread*, which means that the product of maximum delay¹ L and maximum normalized Doppler shift ν_{\max} satisfies $L\nu_{\max} \ll 1$ [5], [9], [27]. For practical channels $L\nu_{\max}$ is about 10^{-4} to 10^{-3} . It can then be shown that for properly chosen symbol duration N , subcarrier spacing $1/K$, and pulses $g[m]$ and $f[m]$, the ISI/ICI terms in (4) are approximately zero, i.e., $\langle \mathbb{H}_0 g_{n',k'}, f_{n,k} \rangle \approx 0$ for $(n', k') \neq (n, k)$ [5], [6], [27], [28]. Therefore, the system relation (4) simplifies to

$$X_{n,k} = H_{n,k} B_{n,k} + Z_{n,k} \quad (6)$$

with the *time-varying channel coefficients* $H_{n,k} \triangleq \langle \mathbb{H}_0 g_{n,k}, f_{n,k} \rangle$ that depend on the channel and the transmit/receive pulses. The intuition behind (6) is that the channel varies considerably between OFDM symbols but negligibly within an OFDM symbol. The channel coefficients can also be written as

$$H_{n,k} = \sum_{l=0}^L h_{n,l} e^{-j2\pi kl/K} \quad (7a)$$

with

$$h_{n,l} \triangleq \sum_{m'=-\infty}^{\infty} h_0[m', l] g[nN - m' - l] \cdot f^*[nN - m']. \quad (7b)$$

Thus, $H_{n,k}$ is the (subsampling) time-dependent transfer function [29] of an *equivalent channel* \mathbb{H} with impulse response $h_{n,l}$. For a conventional cyclic-prefix OFDM system, (7) simplifies to $h_{n,l} = (1/K) \sum_{m=0}^{K-1} h_0[nN + m, l]$.

For later convenience, we rewrite (6) as

$$\mathbf{x}_n = \mathbf{B}_n \mathbf{h}_n + \mathbf{z}_n \quad (8)$$

with the $K \times 1$ vectors $\mathbf{x}_n \triangleq [X_{n,0} \dots X_{n,K-1}]^T$, $\mathbf{h}_n \triangleq [H_{n,0} \dots H_{n,K-1}]^T$, $\mathbf{z}_n \triangleq [Z_{n,0} \dots Z_{n,K-1}]^T$ and the diagonal $K \times K$ matrix $\mathbf{B}_n \triangleq \text{diag}\{B_{n,0}, \dots, B_{n,K-1}\}$.

B. Statistical Characterization

Hereafter, we model \mathbb{H}_0 as a random Rayleigh fading channel satisfying the wide-sense stationary uncorrelated scattering (WSSUS) assumption [9], [29] $E\{h_0[m, l] h_0^*[m', l']\} = T_{\mathbb{H}_0}[m - m', l] \delta[l - l']$, where $T_{\mathbb{H}_0}[m, l]$ is the channel's time-delay correlation function [9], [29]. The scattering function of \mathbb{H}_0 is defined by $C_{\mathbb{H}_0}(l, \nu) \triangleq \sum_{m=-\infty}^{\infty} T_{\mathbb{H}_0}[m, l] e^{-j2\pi \nu m}$ [9], [29]. Its support region is assumed to be contained in $[0, L] \times [-\nu_{\max}, \nu_{\max}]$, where L and ν_{\max} are the maximum

¹Typically, the maximum channel delay L is much smaller than the OFDM symbol duration N .

delay and maximum normalized Doppler shift, respectively. The reciprocal of the maximum delay and the maximum normalized Doppler shift determine the coherence bandwidth and coherence time, respectively.

The equivalent channel \mathbb{H} is also WSSUS, i.e.,

$$\mathbb{E} \{ h_{n,l} h_{n',l'}^* \} = T_{\mathbb{H}}[n - n', l] \delta[l - l'] \quad (9)$$

with

$$T_{\mathbb{H}}[n, l] = \sum_{m=-\infty}^{\infty} \sum_{m'=-\infty}^{\infty} T_{\mathbb{H}_0}[nN + m - m', l] g[m - l] \cdot g^*[m' - l] f[m'] f^*[m].$$

Thus, the channel coefficients $H_{n,k}$ in (7) constitute a two-dimensional (2-D) stationary process with correlation

$$R_{\mathbb{H}}[n, k] \triangleq \mathbb{E} \{ H_{n',k'} H_{n'-n, k'-k}^* \} = \sum_{l=0}^L T_{\mathbb{H}}[n, l] e^{-j2\pi kl/K}. \quad (10)$$

Strong correlations (large values of $R_{\mathbb{H}}[n, k]$) effectively occur within the coherence time and coherence bandwidth, i.e., for $|n| \leq 1/(N\nu_{\max})$ and $|k| \leq K/L$. These correlations between the $H_{n,k}$ will be exploited by our channel predictors.

The noise $u[m]$ in (2) is assumed zero-mean, stationary, white, and circularly symmetric complex Gaussian with variance σ_u^2 . Thus, for $Z_{n,k} = \langle u, f_{n,k} \rangle$ in (6) we obtain

$$\mathbb{E} \{ Z_{n,k} Z_{n',k'}^* \} = \sigma_u^2 \langle f, f_{n-n', k-k'} \rangle.$$

III. OFDM CHANNEL PREDICTION

We next describe the generic receiver structure (cf. Fig. 1) that we propose for channel prediction in OFDM systems. This scheme yields the channel state information that is required for coherent receiver processing (including, e.g., equalization, detection, and decoding). In contrast to the channel estimators in [10]–[17], and to the predictive adaptive loading method in [21], our scheme operates in a decision-directed mode. Furthermore, different from the decision-directed channel estimator in [18], our channel predictor is able to yield up-to-date channel state information. We note that prediction of fading channels in a non-OFDM context was considered in [22]–[26].

Basically, our decision-directed channel predictors process the demodulated receive vector \mathbf{x}_n in order to yield an estimate $\hat{\mathbf{h}}_n$ of the current channel coefficient vector. To this end, past detected symbols $\hat{\mathbf{B}}_{n-p}, \hat{\mathbf{B}}_{n-p-1}, \dots$, obtained by re-encoding previously detected bits \hat{D}_i , are exploited. Here, p denotes the *prediction horizon*, i.e., the number of OFDM symbols the channel is predicted ahead; it is determined e.g., by the latency introduced by the receiver processing and the re-encoder. Successful channel prediction relies on mostly correctly detected symbols $\hat{\mathbf{B}}_n$. In case of too many incorrectly detected symbols, error propagation may result (simulation results in Section V-E show that error propagation is avoided if the SNR is above a certain threshold). For purposes of initialization, the channel predictor can alternatively be operated in training mode where

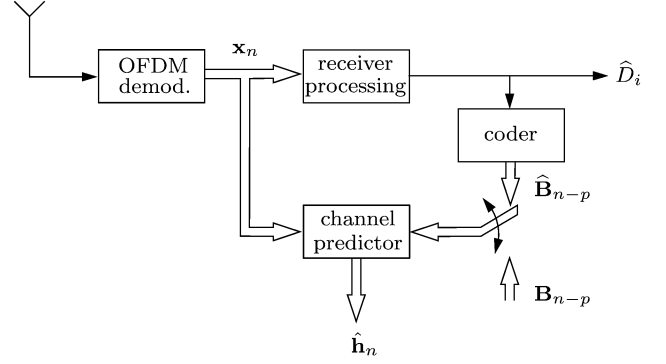


Fig. 1. Generic receiver structure for decision-directed and training-based channel prediction in coded OFDM systems. Prediction horizon is p OFDM symbols.

a few known transmit symbols \mathbf{B}_n serve as training data (numerical experiments in Section V-A suggest that already one OFDM symbol suffices for initialization).

We next develop linear MMSE predictors [19]. In Section IV, adaptive predictor implementations are proposed [20]. For tractability of the problem, we assume that $\hat{\mathbf{B}}_n = \mathbf{B}_n$, i.e., we will neglect error propagation (this is equivalent to assuming that the predictors operate in training mode).

A. MMSE Predictor

We propose to calculate the predicted channel coefficient vector² $\hat{\mathbf{h}}_{n+p} = [\hat{H}_{n+p,0} \dots \hat{H}_{n+p,K-1}]^T$ from the current and past received vectors $\mathbf{x}_n, \mathbf{x}_{n-1}, \dots, \mathbf{x}_{n-M+1}$ assuming knowledge of symbols $\mathbf{B}_n, \mathbf{B}_{n-1}, \dots, \mathbf{B}_{n-M+1}$ by means of a linear multi-input multi-output (MIMO) predictor filter of length (memory) M

$$\hat{\mathbf{h}}_{n+p} = \sum_{m=0}^{M-1} \mathbf{K}_{n,m} \mathbf{x}_{n-m}. \quad (11)$$

The predictor memory allows us to exploit the temporal channel correlations and the MIMO approach is used to exploit the spectral channel correlations (i.e., between different subcarriers). Note that the channel correlations can be fully exploited only if the predictor memory exceeds the channel's coherence time, i.e., if $M > 1/(\nu_{\max}N)$. Hence, the choice of M is a tradeoff between good prediction accuracy and low computational complexity. (The channel's coherence bandwidth is always maximally exploited because the predictor uses all subcarriers.) The predictor coefficient matrices $\mathbf{K}_{n,m}$ of size $K \times K$ are time varying to account for the nonstationarity of the received vector process \mathbf{x}_n (this nonstationarity is due to the symbols \mathbf{B}_n in (8)). The optimum $\mathbf{K}_{n,m}$ minimizing the MSE $\epsilon_{n+p} \triangleq (1/K) \mathbb{E} \{ \|\mathbf{h}_{n+p} - \hat{\mathbf{h}}_{n+p}\|^2 \}$ are derived in Appendix I. It is shown there that the first stage of the resulting MMSE channel predictor is a division of the received symbols $X_{n,k}$ by the symbols $B_{n,k}$, i.e., calculation of

$$Y_{n,k} \triangleq \frac{X_{n,k}}{B_{n,k}} = H_{n,k} + \tilde{Z}_{n,k}, \quad \text{with} \quad \tilde{Z}_{n,k} \triangleq \frac{Z_{n,k}}{B_{n,k}}. \quad (12)$$

²For notational convenience, the problem statement is shifted by p OFDM symbols.

Using $\mathbf{y}_n \triangleq [Y_{n,0} \dots Y_{n,K-1}]^T$, $\tilde{\mathbf{z}}_n \triangleq [\tilde{Z}_{n,0} \dots \tilde{Z}_{n,K-1}]^T$, this can be rewritten as

$$\mathbf{y}_n = \mathbf{B}_n^{-1} \mathbf{x}_n = \mathbf{h}_n + \tilde{\mathbf{z}}_n. \quad (13)$$

Note that in contrast to existing channel estimators [10]–[17], [21] that use the division by the transmit symbols in a rather *ad hoc* manner, we have explicitly derived it as a part of our MMSE predictor. The second stage of the MMSE predictor is a MIMO filter similar to (11) using (13) as input, i.e., $\hat{\mathbf{h}}_{n+p} = \sum_{m=0}^{M-1} \mathbf{W}_{n,m} \mathbf{y}_{n-m}$. The filter matrices $\mathbf{W}_{n,m}$ are also provided in Appendix I. In general, they are time varying since $\tilde{\mathbf{z}}_n$ is nonstationary. Their implementation would require an online inversion of a $KM \times KM$ matrix in each symbol interval. Because this is clearly impractical, we will next develop a simplified MMSE predictor with significantly reduced complexity.

B. Reduced-Complexity MMSE Predictor

A first step toward simplification of the MMSE channel predictor is as follows. We retain the division stage (12), but for the subsequent processing we model the symbols $B_{n,k}$ as random. More specifically, we assume that the $B_{n,k}$ are zero mean and independent identically distributed (i.i.d.). Then, the noise term $\tilde{Z}_{n,k}$ in (12) has mean $E\{\tilde{Z}_{n,k}\} = E\{Z_{n,k}\}E\{1/B_{n,k}\} = 0$ and correlation

$$\begin{aligned} E\{\tilde{Z}_{n,k} \tilde{Z}_{n',k'}^*\} &= E\{Z_{n,k} Z_{n',k'}^*\} E\left\{\frac{1}{B_{n,k} B_{n',k'}^*}\right\} \\ &= \beta^2 \delta[n - n'] \delta[k - k'] \end{aligned}$$

with the “equivalent noise variance” $\beta^2 \triangleq \sigma_u^2 \|f\|^2 E\{1/B_{n,k}\}^2$. It is seen that $\tilde{Z}_{n,k}$ is zero mean, wide-sense stationary, and white. Therefore, the vector $\mathbf{y}_n = \mathbf{h}_n + \tilde{\mathbf{z}}_n$ is stationary and thus can be processed without loss of optimality by means of a *time-invariant* MIMO predictor filter

$$\hat{\mathbf{h}}_{n+p} = \sum_{m=0}^{M-1} \mathbf{W}_m \mathbf{y}_{n-m}. \quad (14)$$

For $p = 0$, the channel predictor (14) becomes a channel estimator that is similar to [14].

The coefficient matrices \mathbf{W}_m are again chosen such that the MSE is minimized. According to the orthogonality principle [30], the optimum \mathbf{W}_m must be such that $E\{(\mathbf{h}_{n+p} - \hat{\mathbf{h}}_{n+p}) \mathbf{y}_{n-m}^H\} = \mathbf{0}$ for $m = 0, \dots, M-1$. With (14) and (13), this yields the Wiener–Hopf equations (for $m = 0, \dots, M-1$)

$$\sum_{m'=0}^{M-1} \mathbf{W}_{m'} (\mathbf{R}_{\mathbb{H}}[m - m'] + \beta^2 \delta[m - m'] \mathbf{I}) = \mathbf{R}_{\mathbb{H}}[m + p] \quad (15)$$

with the $K \times K$ Toeplitz correlation matrix $\mathbf{R}_{\mathbb{H}}[m] \triangleq E\{\mathbf{h}_n \mathbf{h}_{n-m}^H\}$ whose first row is $[R_{\mathbb{H}}[m, 0] \dots R_{\mathbb{H}}[m, -K + 1]]$ [cf. (10)]. Introducing the $K \times KM$ matrix $\mathbf{W} \triangleq [\mathbf{W}_0 \dots \mathbf{W}_{M-1}]$, the $K \times KM$ matrix $\mathbf{V}_{\mathbb{H}} \triangleq [\mathbf{R}_{\mathbb{H}}[p] \dots \mathbf{R}_{\mathbb{H}}[p + M - 1]]$, and the $KM \times KM$ block Toeplitz matrix $\mathbf{R}_{\mathbb{H}}$ with first block row $[\mathbf{R}_{\mathbb{H}}[0] \dots \mathbf{R}_{\mathbb{H}}[M - 1]]$, (15) can be compactly written as

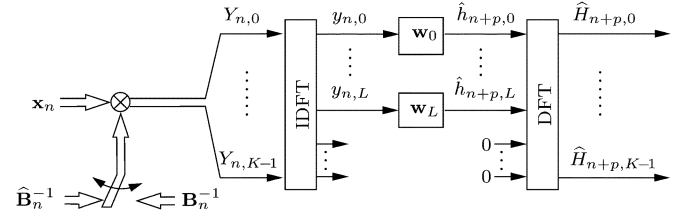


Fig. 2. Efficient DFT implementation of the reduced-complexity MMSE channel predictor.

$\mathbf{W}(\mathbf{R}_{\mathbb{H}} + \beta^2 \mathbf{I}) = \mathbf{V}_{\mathbb{H}}$. Thus, the MMSE-optimum predictor coefficient matrices $\mathbf{W}_{m,\text{opt}}$ are given by

$$\mathbf{W}_{\text{opt}} = [\mathbf{W}_{0,\text{opt}} \dots \mathbf{W}_{M-1,\text{opt}}] = \mathbf{V}_{\mathbb{H}} (\mathbf{R}_{\mathbb{H}} + \beta^2 \mathbf{I})^{-1}. \quad (16)$$

The computation of \mathbf{W}_{opt} is numerically stable (due to the term $\beta^2 \mathbf{I}$) and can efficiently be performed using the Wax–Kailath algorithm [31] (because $\mathbf{R}_{\mathbb{H}} + \beta^2 \mathbf{I}$ is Hermitian Toeplitz-block-Toeplitz). We note that the reduced complexity predictor is equivalent to the full-blown MMSE predictor of Appendix I if the symbol alphabet is phase-shift keying (PSK) and if the noise $Z_{n,k}$ is white ($Z_{n,k}$ is white only if $f[m] = g[m]$).

The minimum MSE obtained with the optimum coefficients $\mathbf{W}_{m,\text{opt}}$ can be shown to be

$$\begin{aligned} \epsilon_{\min} &= R_{\mathbb{H}}[0, 0] - \frac{1}{K} \text{tr} \left\{ \sum_{m=0}^{M-1} \mathbf{W}_{m,\text{opt}} \mathbf{R}_{\mathbb{H}}[-m - p] \right\} \\ &= R_{\mathbb{H}}[0, 0] - \frac{1}{K} \text{tr} \left\{ \mathbf{V}_{\mathbb{H}} (\mathbf{R}_{\mathbb{H}} + \beta^2 \mathbf{I})^{-1} \mathbf{V}_{\mathbb{H}}^H \right\}. \quad (17) \end{aligned}$$

This depends on the channel correlation $\mathbf{R}_{\mathbb{H}}[m]$ and the equivalent noise variance β^2 . While (17) is difficult to interpret, we experimentally verify in Section V-B that ϵ_{\min} tends to decrease for channel coefficients $H_{n,k}$ that are more strongly correlated, i.e., for channels with larger coherence time and/or larger coherence bandwidth. Equivalently, ϵ_{\min} will be smaller for channels that are more underspread [5], [9], [27] in the sense of a smaller maximum normalized Doppler frequency ν_{\max} (slower channel time variations) and/or a smaller maximum delay L . The prediction accuracy also improves for lower noise variance β^2 . These observations agree with results for MMSE channel estimation in [13], [14], and [16], [17].

C. Efficient DFT Implementation

We next propose an efficient implementation of our reduced-complexity MMSE predictor (cf. Fig. 2) that is similar to the pilot-based channel estimators in [10], [14], and [15]. In particular, a K -point IDFT is applied to the $Y_{n,k} = X_{n,k}/B_{n,k} = H_{n,k} + \tilde{Z}_{n,k}$, which yields [cf. (7)]

$$\begin{aligned} y_{n,l} &= \frac{1}{K} \sum_{k=0}^{K-1} Y_{n,k} e^{j2\pi lk/K} \\ &= \begin{cases} h_{n,l} + \tilde{z}_{n,l}, & l = 0, \dots, L, \\ \tilde{z}_{n,l}, & l = L + 1, \dots, K - 1. \end{cases} \quad (18) \end{aligned}$$

Subsequently, a length- M MMSE predictor predicts the channel impulse response $h_{n+p,l}$ for each delay $l = 0, \dots, L$. Due to the WSSUS property (9), $h_{n,l}$ and $h_{n,l'}$ are uncorrelated for $l \neq l'$.

TABLE I
COMPUTATIONAL COMPLEXITY OF THE VARIOUS CHANNEL PREDICTORS

Predictor type	Design	Prediction	Update
MMSE predictor	—	$\mathcal{O}(M^3K^3)$	—
Reduced-complexity MMSE predictor	$\mathcal{O}(M^2K^2)$	$\mathcal{O}(MK^2)$	—
DFT implementation	$\mathcal{O}(M^2L)$	$\mathcal{O}(K \log_2 K + ML)$	—
Adaptive predictor using NLMS algorithm	—	$\mathcal{O}(K \log_2 K + ML)$	$\mathcal{O}(ML)$
Adaptive predictor using RLS algorithm	—	$\mathcal{O}(K \log_2 K + ML)$	$\mathcal{O}(M^2L)$

Therefore, the MMSE predictor decomposes into $L + 1$ parallel single-input single-output (SISO) predictors given by

$$\hat{h}_{n+p,l} = \mathbf{w}_l^H \mathbf{y}_{n,l}, \quad l = 0, \dots, L \quad (19)$$

with predictor coefficients $\mathbf{w}_l \triangleq [w_{0,l} \dots w_{M-1,l}]^T$ (specified below) and $\mathbf{y}_{n,l} \triangleq [y_{n,l} \dots y_{n-M+1,l}]^T$. Finally, the predicted channel coefficients $\hat{H}_{n+p,k}$ are obtained from the predicted impulse response samples $\hat{h}_{n+p,l}$ via a K -point DFT

$$\hat{H}_{n+p,k} = \sum_{l=0}^L \hat{h}_{n+p,l} e^{-j2\pi kl/K}. \quad (20)$$

This implementation has significantly reduced complexity because $L + 1 \ll K$ SISO predictors are used instead of a $K \times K$ MIMO predictor (see Section III-D).

The Wiener–Hopf equations for the MMSE predictor coefficients \mathbf{w}_l can be obtained by diagonalization of (15), corresponding to a transformation into the eigenspace of the correlation matrices $\mathbf{R}_{\mathbb{H}}[m]$. This diagonalizing transformation is a K -point DFT since the $K \times K$ matrices $\mathbf{R}_{\mathbb{H}}[m]$ can be shown to be circulant. Indeed, the predictor coefficients \mathbf{W}_m in (14) and \mathbf{w}_l in (19) are related as $\mathbf{W}_m = \mathbf{U} \mathbf{\Lambda}_m \mathbf{U}^H$, where $\mathbf{\Lambda}_m \triangleq \text{diag}\{w_{m,0}, \dots, w_{m,L}, 0, \dots, 0\}$ and \mathbf{U} is the $K \times K$ DFT matrix. The resulting diagonalized normal equations can be rearranged as $(\mathbf{T}_{\mathbb{H}}[l] + (\beta^2/K)\mathbf{I})\mathbf{w}_l = \mathbf{t}_{\mathbb{H}}[l]$, $l = 0, \dots, L$, with the $M \times M$ Hermitian Toeplitz correlation matrix $\mathbf{T}_{\mathbb{H}}[l]$ with first row $[T_{\mathbb{H}}[0, l], \dots, T_{\mathbb{H}}[M-1, l]]$ and the cross-correlation vector $\mathbf{t}_{\mathbb{H}}[l] = [T_{\mathbb{H}}[-p, l] \dots T_{\mathbb{H}}[-p-M+1, l]]^T$ [cf. (9)]. Thus, the predictor coefficients are given by

$$\mathbf{w}_{l,\text{opt}} = \left(\mathbf{T}_{\mathbb{H}}[l] + \frac{\beta^2}{K} \mathbf{I} \right)^{-1} \mathbf{t}_{\mathbb{H}}[l], \quad l = 0, \dots, L. \quad (21)$$

The computation of (21) can be done stably and efficiently using the Levinson algorithm [32].

The minimum MSE in (17) can correspondingly be rewritten as

$$\begin{aligned} \epsilon_{\min} &= \sum_{l=0}^L [T_{\mathbb{H}}[0, l] - \mathbf{w}_{l,\text{opt}}^H \mathbf{t}_{\mathbb{H}}[l]] \\ &= R_{\mathbb{H}}[0, 0] - \sum_{l=0}^L \mathbf{t}_{\mathbb{H}}^H[l] \left(\mathbf{T}_{\mathbb{H}}[l] + \frac{\beta^2}{K} \mathbf{I} \right)^{-1} \mathbf{t}_{\mathbb{H}}[l]. \end{aligned}$$

As analyzed in Appendix II, the DFT implementation deviates from the reduced-complexity predictor if some subcarriers are not used (as is done in practical systems to reduce out-of-band emissions).

D. Computational Complexity of MMSE Predictors

The computational complexity of the predictors discussed so far is compared in Table I. We specify the complexity both for the design (performed in advance only once) and for the actual channel prediction (performed in each symbol interval). For the full-blown MMSE predictor, the design has to be performed anew for each symbol and thus it is considered as part of the prediction. It is assumed that the design is based on the Wax–Kailath algorithm [31] in the case of the reduced-complexity MMSE predictor and on the Levinson algorithm [32] in the case of the DFT implementation.

It is seen that for typical values of the parameters K , M , and L , the DFT implementation is significantly less complex than the original implementation of the reduced-complexity MMSE predictor, which in turn is significantly less complex than the full-blown MMSE predictor.

IV. ADAPTIVE OFDM CHANNEL PREDICTORS

In practice, calculation of the MMSE channel predictors in (16) or (21) presupposes knowledge of the channel correlations $\mathbf{R}_{\mathbb{H}}[m]$, respectively, $\mathbf{T}_{\mathbb{H}}[l]$ and of the noise variance β^2 . Furthermore, the statistics of real-world channels change slowly over time (cf. [33]), thus necessitating reestimation of $\mathbf{R}_{\mathbb{H}}[m]$ ($\mathbf{T}_{\mathbb{H}}[l]$) and recalculation of the channel predictor once in a while. To avoid these problems, we propose novel *adaptive* channel predictors that perform a continual update of the predictor coefficients, do not require knowledge of the channel and noise statistics, and are capable of tracking nonstationary statistics.

The adaptive predictors will be based on the same structure as the DFT implementation of Fig. 2 because it is computationally efficient and involves only $M(L + 1)$ scalar coefficients, instead of the MK^2 scalar coefficients required by the reduced-complexity MMSE predictor (note that adaptive algorithms generally perform better when fewer coefficients have to be adapted [34]). Replacing the MMSE predictor filters in $\hat{h}_{n+p,l} = \mathbf{w}_{l,\text{opt}}^H \mathbf{y}_{n,l}$ [cf. (19) and (21)] with adaptive filters

$\mathbf{w}_l[n]$, the adaptively predicted channel taps (assuming for simplicity that the adaptation starts at $n = 0$) are³ (cf. [20])

$$\hat{h}_{n+p,l} = \mathbf{w}_l^H[n] \mathbf{y}_{n,l}, \quad n \geq 0, \quad l = 0, \dots, L. \quad (22)$$

For the adaptation, we will consider the normalized least mean-square (NLMS) algorithm and the recursive least squares (RLS) algorithm [34], designed such that the prediction error $h_{n+p,l} - \hat{h}_{n+p,l}$ is minimized. Rewriting the channel impulse response p OFDM symbols ahead as $h_{n+p,l} = \mathbf{w}_{l,\text{opt}}^H \mathbf{y}_{n,l} + \varepsilon_{n,l}$, where $\varepsilon_{n,l}$ is the prediction error achieved by the MMSE predictor filter $\mathbf{w}_{l,\text{opt}}$, the error of the adaptive predictor equals $h_{n+p,l} - \hat{h}_{n+p,l} = (\mathbf{w}_{l,\text{opt}} - \mathbf{w}_l[n])^H \mathbf{y}_{n,l} + \varepsilon_{n,l}$. Obviously, this error is minimized for $\mathbf{w}_l[n] = \mathbf{w}_{l,\text{opt}}$. Hence, the adaptive algorithm will tune $\mathbf{w}_l[n]$ to approximate the (time-invariant) MMSE predictor filters $\mathbf{w}_{l,\text{opt}}$. Thus, our approach differs from adaptive channel tracking schemes (e.g., [11] and [35]), where the adaptive algorithms directly track/approximate the (time-varying) channel impulse response.

A. NLMS Algorithm

The NLMS algorithm belongs to the family of stochastic gradient algorithms that iteratively estimate the MMSE predictor filters. We use the NLMS algorithm rather than the LMS algorithm because the selection of the adaptation constant is simpler. The predictor filters $\mathbf{w}_l[n]$ are updated according to [34]

$$\mathbf{w}_l[n] = \mathbf{w}_l[n-1] + \frac{\mu}{\|\mathbf{y}_{n-p,l}\|^2} e_{n,l}^* \mathbf{y}_{n-p,l}, \quad n \geq p$$

where μ is the adaptation constant, $\|\mathbf{y}_{n-p,l}\|^2 = \sum_{i=0}^{M-1} |y_{n-p-i,l}|^2$, and $e_{n,l}$ is the prediction error that would ideally be given by $h_{n,l} - \mathbf{w}_l^H[n-1] \mathbf{y}_{n-p,l}$. However, since we perform noisy prediction, $h_{n,l}$ is not available. Thus, we approximate $h_{n,l}$ by $y_{n,l} = h_{n,l} + \tilde{z}_{n,l}$ in (18), i.e.,

$$e_{n,l} \triangleq y_{n,l} - \mathbf{w}_l^H[n-1] \mathbf{y}_{n-p,l}, \quad n \geq p. \quad (23)$$

The error introduced by this approximation will be small for most practical SNRs. For $n = 0, \dots, p-1$, the prediction filters are initialized as

$$\mathbf{w}_l[n] = [1 \quad 0 \quad \dots \quad 0]^T, \quad n = 0, \dots, p-1. \quad (24)$$

Thus, $\hat{h}_{n+p,l} = y_{n,l}$ for $n = 0, \dots, p-1$.

Stable operation of the NLMS algorithm requires $0 < \mu < 2$ [34]. The selection of μ determines the convergence speed, the excess MSE, and the ability to track nonstationary channel statistics. We obtained good results with $\mu \approx 0.5$; however, the exact value is rather uncritical.

B. RLS Algorithm

With the RLS algorithm, the l th predictor filter $\mathbf{w}_l[n]$ is calculated such that it minimizes the error [34]

$$\epsilon_{\text{RLS},l}[n] \triangleq \sum_{n'=p}^n \lambda^{n-n'} |h_{n',l} - \mathbf{w}_l^H[n] \mathbf{y}_{n'-p,l}|^2.$$

³Here, we require at least a reasonable guess of the maximum channel delay L .

Here, λ with $0 < \lambda \leq 1$ is a forgetting factor that accounts for possible nonstationarity of the input $\mathbf{y}_{n,l}$. Similar to the NLMS algorithm, the selection of λ is not critical. We obtained good results for $\lambda = 0.99$.

The resulting update equation for $\mathbf{w}_l[n]$ is

$$\mathbf{w}_l[n] = \mathbf{w}_l[n-1] + \mathbf{k}_{n-p,l} e_{n,l}^*, \quad n \geq p$$

where $e_{n,l}$ is as in (23) and $\mathbf{k}_{n,l}$ is the RLS gain vector given by

$$\mathbf{k}_{n,l} = \frac{\mathbf{P}_{n-1,l} \mathbf{y}_{n,l}}{\lambda + \mathbf{y}_{n,l}^H \mathbf{P}_{n-1,l} \mathbf{y}_{n,l}}, \quad n \geq 1.$$

The matrix $\mathbf{P}_{n,l}$ is the inverse of the $M \times M$ sample covariance matrix $\sum_{i=0}^n \lambda^{n-i} \mathbf{y}_{i,l} \mathbf{y}_{i,l}^H$; it can be calculated recursively as

$$\mathbf{P}_{n,l} = \frac{1}{\lambda} (\mathbf{I} - \mathbf{k}_{n,l} \mathbf{y}_{n,l}^H) \mathbf{P}_{n-1,l}, \quad n \geq 1.$$

The RLS recursion is initialized as in (24). Furthermore, we set

$$\mathbf{k}_{0,l} = \mathbf{P}_{0,l} \mathbf{y}_{0,l} = \frac{1}{\|\mathbf{y}_{0,l}\|^2 + \delta} \mathbf{y}_{0,l}$$

and

$$\mathbf{P}_{0,l} = (\mathbf{y}_{0,l} \mathbf{y}_{0,l}^H + \delta \mathbf{I})^{-1} = \frac{1}{\delta} \left[\mathbf{I} - \frac{\mathbf{y}_{0,l} \mathbf{y}_{0,l}^H}{\|\mathbf{y}_{0,l}\|^2 + \delta} \right]$$

where the stabilization factor δ is in the range $0 < \delta \ll 1$ (we chose $\delta = 0.1$).

Compared to the NLMS algorithm, the RLS algorithm converges faster with smaller excess MSE. Its convergence rate is independent of the eigenvalue spread of the input process [34].

C. Computational Complexity of Adaptive Predictors

The last two rows of Table I show the computational complexity of the channel prediction (filter operation) and coefficient update for the adaptive predictors. Both operations have to be performed in each symbol interval. While the complexity of the channel prediction for NLMS and RLS is equal (and furthermore the same as the complexity of the DFT implementation), the coefficient update is significantly more costly for RLS than for NLMS, in particular for large filter lengths M .

V. SIMULATION RESULTS

We simulated a conventional OFDM system with $K = 120$ subcarriers, cyclic prefix length $L_{\text{cp}} = 20$, and 16-QAM modulation with $E\{|B_{n,k}|^2\} = 1$. A (15, 7) Reed–Solomon code over $\text{GF}(2^4)$ was used with each code word corresponding to one 16-QAM symbol. Furthermore, symbol interleaving across subcarriers was used. At the receiver, the system shown in Fig. 1 was implemented. The receiver processing consisted of zero-forcing equalization, hard symbol decisions, and decoding. The zero-forcing equalizer calculates $\tilde{X}_{n,k} \triangleq X_{n,k} / \hat{H}_{n,k}$ [cf. (6)] with channel coefficient estimates $\hat{H}_{n,k}$ provided by the channel predictor. The channel predictor consisted of 20 SISO prediction filters (see Fig. 2), each of length $M = 10$. The prediction horizon was $p = 1$ unless indicated otherwise. The parameters of the NLMS and RLS algorithms were chosen as $\mu = 0.5$ and $\lambda = 0.99$, respectively. The channel predictor was operated in decision-directed mode (where error propagation may occur) or

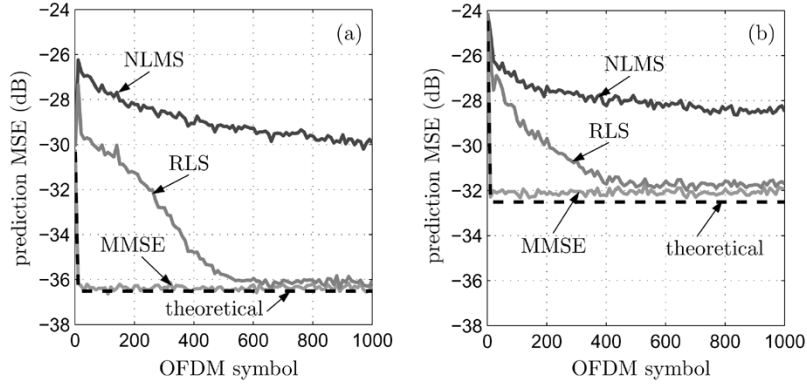


Fig. 3. Convergence behavior of the adaptive predictors in decision-directed mode for (a) slow channel and (b) fast channel.

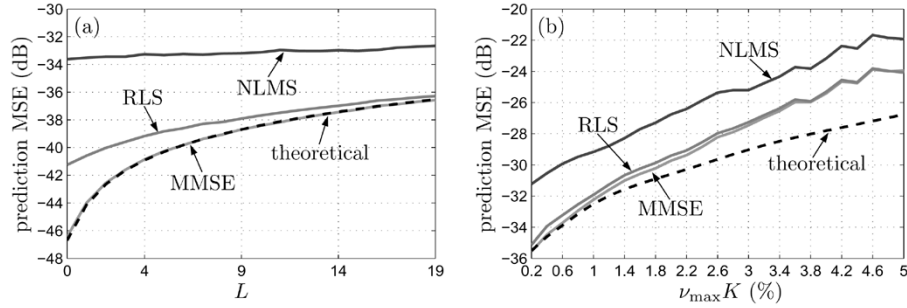


Fig. 4. Dependence of the prediction MSE on (a) channel's maximum delay L (with fixed $\nu_{\max}K = 0.001$) and (b) maximum Doppler frequency ν_{\max} (with fixed $L = 19$).

in training mode (no error propagation). The channel was simulated by means of the technique described in [36], using a scattering function with exponentially decaying delay profile and Jakes Doppler profile [37], i.e.,

$$C_{H_0}(l, \nu) = \begin{cases} \frac{\exp(-\frac{l}{\tau_0})}{\sqrt{\nu_{\max}^2 - \nu^2}}, & \text{for } l = 0, \dots, L, |\nu| < \nu_{\max} \\ 0, & \text{elsewhere.} \end{cases} \quad (25)$$

Unless stated otherwise, we chose $L = 19$ and $\tau_0 = L / \log_e(2L)$. We considered both a “slow” (slowly time-varying) channel with $\nu_{\max}K = 0.001$ and a “fast” (rapidly time-varying) channel with $\nu_{\max}K = 0.01$ (recall that ν_{\max} is the maximum normalized Doppler shift). The quantity $\nu_{\max}K$ can be interpreted as the percentage of channel variation within one OFDM symbol. The additive noise was zero mean and white with variance σ_u^2 . The SNR $\sigma_{H_0}^2 / \sigma_u^2$ (here, $\sigma_{H_0}^2 \triangleq \sum_{l=0}^L \int_{-\nu_{\max}}^{\nu_{\max}} C_{H_0}(l, \nu) d\nu$ is the channel's path loss) was 25 dB unless indicated otherwise.

A. Convergence of the Adaptive Predictors

We assume packet transmission with a packet length of 1000 OFDM symbols and a packet preamble of only one known OFDM symbol. At the beginning of each packet, the channel predictors are initialized using the preamble. Afterwards, the predictors are operated in decision-directed mode. With the adaptive channel predictors, the coefficient adaptation is started when $M = 10$ OFDM symbols have been received. For the first M OFDM symbols, the length of the MMSE channel predictor was increased from 1 to M , equal to the number of OFDM symbols received.

Fig. 3(a) and (b) illustrates the convergence of the adaptive channel predictors for the slow and the fast channel, respectively. The (normalized) prediction MSE was averaged over 100 experiments. For comparison, the MSE of the reduced-complexity MMSE predictor from Section III-B and the theoretical MMSE according to (17) are also plotted. It is seen that the MSE is below -25 dB for all predictors. However, the MSE obtained after convergence is about 2–4 dB higher for the fast channel than for the slow channel. The RLS-based predictor converges faster than the NLMS-based predictor and has almost no excess MSE relative to the reduced-complexity MMSE predictor. The excess MSE of the NLMS-based predictor is 4–6 dB. We note that in this simulation, the predictors always converged, i.e., a single OFDM training symbol suffices for startup at SNR 25 dB.

B. Dependence of the Prediction MSE on the Maximum Delay and Doppler

In the following simulation, for better clarity the predictor was operated in training mode to exclude effects resulting from decision-directed operation. Fig. 4 shows how the prediction MSE (after convergence in the case of adaptive predictors) depends on the channel's maximum delay L and maximum normalized Doppler ν_{\max} [cf. (25)]. Note that larger L (ν_{\max}) implies a smaller coherence bandwidth (coherence time). The prediction MSE was obtained by averaging over 100 experiments with 10^4 symbols each.

In Fig. 4(a), L is varied between 0 and 19 while $\nu_{\max}K$ is fixed at 0.001 (slow channel). In Fig. 4(b), $\nu_{\max}K$ is varied between 0.002 and 0.05 at a fixed L of 19. It can be seen that the prediction MSE increases with increasing L and ν_{\max} . This is an

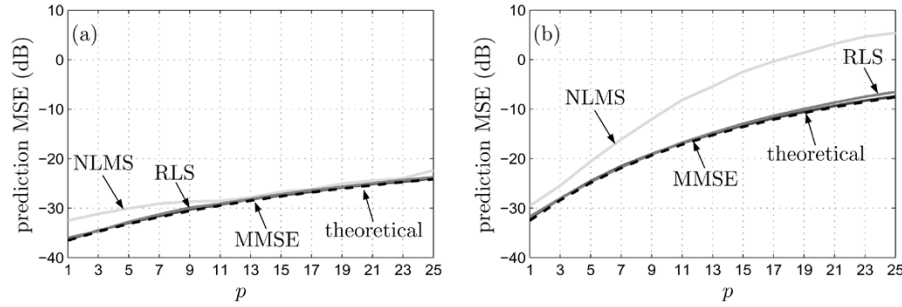


Fig. 5. Prediction MSE versus prediction horizon p for (a) slow channel and (b) fast channel.

expected behavior since prediction is more difficult for weaker correlations, i.e., for a smaller coherence bandwidth and coherence time. We also see that the RLS-based predictor performs nearly as well as the MMSE predictor. According to Fig. 4(b), for high ν_{\max} all predictors deviate from the theoretical MMSE. This is because the approximate system relation (6) becomes less accurate with increasing ν_{\max} [6] due to increasing ICI.

C. Prediction MSE for Different Prediction Horizons

Fig. 5(a) and (b) shows the prediction MSE (averaged over ten experiments with 10^4 symbols each) versus the prediction horizon p . The predictor was again operated in training mode. For the slow channel considered in Fig. 5(a), channel prediction is accurate even for large prediction horizons. The excess MSE of the RLS predictor is negligible; that of the NLMS predictor is between 1 and 4 dB. For the fast channel in Fig. 5(b), reliable channel estimates can be obtained only for significantly smaller prediction horizons than for the slow channel (e.g., for an MSE of -20 dB, p can be at most 5...9 for the fast channel, while for the slow channel p can be beyond 25). The RLS predictor again performs as well as the MMSE predictor. However, the performance of the NLMS predictor is substantially worse.

D. Tracking of Nonstationary Channel Statistics

We next study the ability of the adaptive channel predictors to track nonstationary channel statistics [33]. The nonstationary channel was obtained as a superposition of two different WSSUS channels. Channel 1 had a flat scattering function (i.e., rectangular delay and Doppler profiles) with maximum delay $L = 14$, maximum relative Doppler $\nu_{\max}K = 0.004$, and a path loss such that the SNR was 25 dB. Channel 2 had the Jakes exponential scattering function (25) with $L = 19$, $\nu_{\max}K = 0.01$, and a path loss such that the SNR was 35 dB. Channel 1 was “active” during the first 1000 OFDM symbols (phase 1), and channel 2 was active during the last 1000 OFDM symbols (phase 2). During the intermediate transition phase, the actual channel was obtained by gradually blending channel 1 into channel 2.

Fig. 6 shows the prediction MSE versus the OFDM symbol index n for the adaptive channel predictors operated in decision-directed mode (initial convergence is not shown.) The prediction MSE was averaged over 100 experiments. It is seen that both adaptive algorithms succeed in tracking the nonstationary statistics. During phases 1 and 3, the RLS-based predictor performs practically as well as the respective MMSE predictor. The NLMS-based predictor has an excess MSE of about 3 dB. As expected, the MMSE channel predictors perform rather poorly if they are

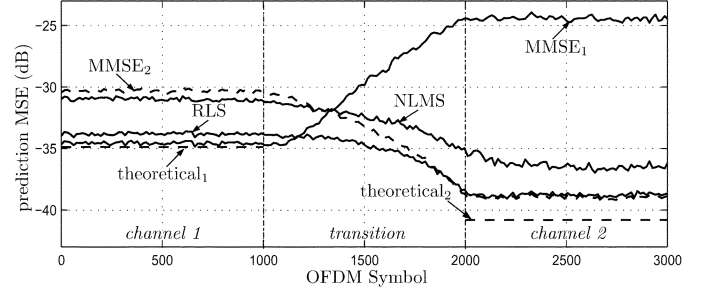


Fig. 6. Tracking of a channel with nonstationary statistics. Curves labeled MMSE_i and theoretical_i show the estimated and theoretical MMSE, respectively, for channel i , $i = 1, 2$.

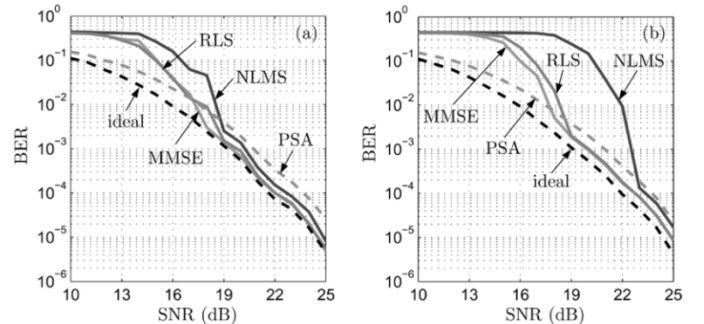


Fig. 7. BER versus SNR for various OFDM receivers: (a) slow channel and (b) fast channel.

not matched to the current channel conditions. During phase 3, the MMSE predictor deviates from the theoretical MMSE by about 2 dB. This can be attributed to large ICI that is caused by the fast channel and that violates the system relation (6).

E. BER Performance of Predictive Equalization

We finally assess the bit-error rate (BER) performance of the overall OFDM system with decision-directed predictive equalization shown in Fig. 1. In Fig. 7(a) and (b), we show the BER (after decoding) versus SNR obtained with the various receivers for the slow and fast channel, respectively. The BERs were obtained by averaging over 25 experiments with 10^4 OFDM symbols each. All channel predictors are operated in decision-directed mode. For comparison, we show the performance of an idealized receiver that has perfect channel state information. Furthermore, we simulated an OFDM system using the 2-D pilot symbol assisted (PSA) MMSE channel estimator from [16] (here, approximately 10% of the transmit symbols were pilots and the estimator length was equal to the prediction filter length). This MMSE estimator uses a flat

scattering function resulting from a robust design approach and requires knowledge of the maximum delay L , the maximum normalized Doppler ν_{\max} , and the SNR.

Due to the decision-directed operation, good performance of the predictors requires the SNR to be above a certain threshold. This threshold SNR is approximately equal for the RLS-based predictor and the MMSE predictor and about 2–4 dB higher for the NLMS-based predictor. For SNR values above the threshold, the BER of the receivers with predictive channel equalization is nearly equal to the BER of the ideal receiver. Moreover, our decision-directed predictive receiver outperforms the PSA receiver for high SNR. This is due to the fact that our channel predictors use symbol decisions from all K subcarriers (resulting in a computational complexity that is larger than that of the PSA estimator). Finally, we note that the SNR threshold strongly depends on the specific channel code that is used. A theoretical assessment of this threshold is beyond the scope of this paper.

VI. CONCLUSION

We presented decision-directed channel predictors for OFDM communications over time-varying channels. Channel prediction is practically relevant because it yields up-to-date channel state information by compensating for delays that are due to decoding, feedback, etc. Pilot symbols are not required for our channel predictors but can be incorporated by replacing the corresponding detected symbols. In this paper, we focused on delay-free equalization; however, channel prediction also enables other techniques that hold the promise of improved system capacity and link reliability (e.g., link adaptation).

In addition to a reduced-complexity minimum MSE predictor and its efficient DFT-based implementation, we developed adaptive predictors using the NLMS and RLS algorithms. These adaptive predictors avoid an explicit design, do not require any statistical prior knowledge, and are able to track nonstationary channel and noise statistics. Our simulation results demonstrated that in a decision-directed mode, using a single OFDM symbol for initialization, adaptive prediction features excellent performance even in the case of fast time-varying channels. However, to avoid error propagation, a certain minimum SNR is required. Above this SNR threshold, prediction of time-varying channels over large prediction horizons is feasible.

APPENDIX I

DERIVATION OF THE MMSE CHANNEL PREDICTOR

The input–output relation of the MIMO predictor (11) can be compactly written as

$$\hat{\mathbf{h}}_{n+p} = \mathcal{K}_n \mathcal{X}_n \quad (26)$$

with the predictor coefficient block matrix $\mathcal{K}_n \triangleq [\mathbf{K}_{n,0} \dots \mathbf{K}_{n,M-1}]$ and the stacked vector $\mathcal{X}_n \triangleq [\mathbf{x}_n^T \dots \mathbf{x}_{n-M+1}^T]^T$. The optimum \mathcal{K}_n minimize the MSE $\epsilon_{n+p} \triangleq (1/K)E\{\|\mathbf{h}_{n+p} - \hat{\mathbf{h}}_{n+p}\|^2\}$. According to the orthogonality principle [30], they are such that $E\{(\mathbf{h}_{n+p} - \hat{\mathbf{h}}_{n+p})\mathcal{X}_n^T\} = \mathbf{0}$. Upon insertion of (8) and (26), this yields the normal equations $\mathcal{K}_n \mathbf{B}_n (\mathcal{R}_{\mathbb{H}} + \mathbf{B}_n^{-1} \mathcal{R}_Z \mathbf{B}_n^{-H}) = \mathcal{V}_{\mathbb{H}}$. Here, $\mathbf{B}_n \triangleq \text{diag}\{\mathbf{B}_n, \dots, \mathbf{B}_{n-M+1}\}$ is a diagonal matrix with

the transmit symbols $B_{n-M+1,k}, \dots, B_{n,k}$ as diagonal elements, and $\mathcal{V}_{\mathbb{H}} \triangleq [\mathbf{R}_{\mathbb{H}}[p] \dots \mathbf{R}_{\mathbb{H}}[p+M-1]]$. The correlation matrices $\mathcal{R}_{\mathbb{H}}$ and \mathcal{R}_Z are block-Toeplitz with first block rows $[\mathbf{R}_{\mathbb{H}}[0,0] \dots \mathbf{R}_{\mathbb{H}}[M-1]]$ and $[\mathbf{R}_Z[0,0] \dots \mathbf{R}_Z[M-1]]$, respectively, where $\mathbf{R}_{\mathbb{H}}[m] \triangleq E\{\mathbf{h}_n \mathbf{h}_{n-m}^H\}$ and $\mathbf{R}_Z[m] \triangleq E\{\mathbf{z}_n \mathbf{z}_{n-m}^H\}$. Thus, the optimum predictor coefficients are

$$\mathcal{K}_{\text{opt},n} = \mathbf{W}_n \mathbf{B}_n^{-1} \quad (27)$$

with

$$\begin{aligned} \mathbf{W}_n &= [\mathbf{W}_{n,0} \dots \mathbf{W}_{n,M-1}] \\ &= \mathcal{V}_{\mathbb{H}} (\mathcal{R}_{\mathbb{H}} + \mathbf{B}_n^{-1} \mathcal{R}_Z \mathbf{B}_n^{-H})^{-1}. \end{aligned} \quad (28)$$

The MMSE achieved by $\mathcal{K}_{\text{opt},n}$ can be shown to be $\epsilon_{\min,n} = R_{\mathbb{H}}[0,0] - \text{tr}\{\mathbf{W}_n \mathcal{V}_{\mathbb{H}}^H\}$.

Inserting (27) in (26) yields $\hat{\mathbf{h}}_{n+p} = \mathbf{W}_n \mathbf{B}_n^{-1} \mathcal{X}_n$, which shows that the predictor implements two operations. The first operation is a division of the $X_{n,k}$ by the data symbols $B_{n,k}$, $\mathcal{Y}_n = \mathbf{B}_n^{-1} \mathcal{X}_n$, or, equivalently, $Y_{n,k} = X_{n,k}/B_{n,k}$. The second operation $\hat{\mathbf{h}}_{n+p} = \mathbf{W}_n \mathcal{Y}_n$ corresponds to a MIMO system that is time varying because it depends on the transmit symbols via \mathbf{B}_n [cf. (28)].

APPENDIX II

SYSTEMATIC ERROR CAUSED BY UNUSED SUBCARRIERS

We analyze the systematic error resulting from the DFT implementation in Section III-C (see Fig. 2) when only the subcarriers $k \in \mathcal{K} \triangleq \{K_{\min}, K_{\min} + 1, \dots, K_{\max}\}$ are used for data transmission and the symbols of the remaining subcarriers are set to zero. Thus, only $K_{\text{used}} \triangleq K_{\max} - K_{\min} + 1$ subcarriers transmit data symbols. For simplicity, we assume zero noise ($Z_{n,k} = 0$) and a time-invariant channel ($H_{n,k} = H_k$ and $h_{n,l} = h_l$). Furthermore, we assume that (6) is satisfied exactly, so that $X_{n,k} = H_k B_{n,k}$.

According to Fig. 2, the first step of the DFT implementation is a division of the received sequences by the data symbols, which yields $Y_{n,k} = X_{n,k}/B_{n,k} = H_k = \sum_{l=0}^L h_l e^{-j2\pi kl/K}$ for $k \in \mathcal{K}$. The last equation can be written as $\mathcal{H} = \mathbf{U}\mathbf{h}$, where $\mathcal{H} \triangleq [H_{K_{\min}} \dots H_{K_{\max}}]^T$, $\mathbf{h} \triangleq [h_0 \dots h_L]^T$, and the columns of the $K_{\text{used}} \times (L+1)$ matrix \mathbf{U} are $[e^{-j2\pi K_{\min} l/K} \dots e^{-j2\pi K_{\max} l/K}]^T$, $l = 0, \dots, L$. Note that typically $K_{\text{used}} \gg L$. Applying to $\mathcal{H} = \mathbf{U}\mathbf{h}$ an IDFT as in (18) followed by a DFT as in (20) yields

$$\hat{\mathcal{H}} = \frac{1}{K} \mathbf{U} \mathbf{U}^H \mathcal{H} = \frac{1}{K} \mathbf{U} \mathbf{U}^H \mathbf{U} \mathbf{h}.$$

Because $\mathbf{U} \mathbf{U}^H \neq K \mathbf{I}$ for $K_{\text{used}} < K$, we have $\hat{\mathcal{H}} \neq \mathcal{H}$. Using results from [38], the systematic MSE $\epsilon_{\text{unused}} \triangleq (1/K_{\text{used}})E\{\|\mathcal{H} - \hat{\mathcal{H}}\|_F^2\}$ ($\|\cdot\|_F$ denotes the Froebenius norm) can be shown to be

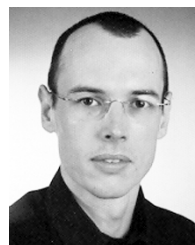
$$\begin{aligned} \epsilon_{\text{unused}} &= \frac{1}{K_{\text{used}}} \text{tr} \left\{ \mathbf{U} \left(\mathbf{I} - \frac{1}{K} \mathbf{U}^H \mathbf{U} \right) \right. \\ &\quad \left. \cdot \mathbf{T}_{\mathbb{H}} \left(\mathbf{I} - \frac{1}{K} \mathbf{U}^H \mathbf{U} \right) \mathbf{U}^H \right\} \\ &= \frac{1}{K_{\text{used}}} \sum_{l=0}^L T_{\mathbb{H}}[l] |\gamma_l|^2 \end{aligned} \quad (29)$$

where $\mathbf{T}_H \triangleq E\{\mathbf{h}\mathbf{h}^H\} = \text{diag}\{T_H[0,0], \dots, T_H[L]\}$ and $\boldsymbol{\gamma}_l$ is the l th column of the matrix $\boldsymbol{\Gamma} \triangleq \mathbf{U}(\mathbf{I} - (1/K)\mathbf{U}^H\mathbf{U})$. Note that \mathbf{T}_H is diagonal due the WSSUS assumption (9).

Assuming $K_{\text{used}}/K = 0.9$ and $\mathbf{T}_H = \mathbf{I}/(L+1)$, i.e., all $L+1$ paths are equally strong and the path loss is normalized to $\text{tr}\{\mathbf{T}_H\} = 1$, ϵ_{unused} calculated from (29) is about -16 dB for $K = 100$ subcarriers and between -20 dB and -25 dB for $K = 1000$ subcarriers.

REFERENCES

- [1] R. W. Chang, "Synthesis of band-limited orthogonal signals for multi-channel data transmission," *Bell Syst. Tech. J.*, vol. 45, pp. 1775–1796, Dec. 1966.
- [2] A. Peled and A. Ruiz, "Frequency domain data transmission using reduced computational complexity algorithms," in *Proc. IEEE ICASSP-80*, Denver, CO, 1980, pp. 964–967.
- [3] L. J. Cimini, "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. Commun.*, vol. 33, no. 7, pp. 665–675, Jul. 1985.
- [4] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, pp. 5–14, May 1990.
- [5] W. Kozek and A. F. Molisch, "Nonorthogonal pulses for multicarrier communications in doubly dispersive channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 10, pp. 1579–1589, Oct. 1998.
- [6] D. Schafhuber, G. Matz, and F. Hlawatsch, "Pulse-shaping OFDM/BFDM systems for time-varying channels: ISI/ICI analysis, optimal pulse design, and efficient implementation," in *Proc. IEEE PIMRC-02*, Lisbon, Portugal, Sept. 2002, pp. 1012–1016.
- [7] A. N. Akansu, P. Duhamel, X. Lin, and M. de Courville, "Orthogonal transmultiplexers in communication: A review," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 979–995, Apr. 1998.
- [8] H. Bölcskei, "Efficient design of pulse shaping filters for OFDM systems," in *Proc. SPIE Wavelet Applications Signal Image Processing VII*, Denver, CO, Jul. 1999, pp. 625–636.
- [9] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.
- [10] Y. Li, N. Seshadri, and S. Ariyavisitakul, "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 3, pp. 461–471, Mar. 1999.
- [11] I. Barhum, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615–1624, Jun. 2003.
- [12] P. Hoeher, "TCM on frequency-selective land-mobile fading channels," in *Proc. 5th Tirrenia Int. Workshop*, Tirrenia, Italy, Sep. 1991, pp. 317–328.
- [13] P. Hoeher, S. Kaiser, and P. Robertson, "Two-dimensional pilot-symbol-aided channel estimation by Wiener filtering," in *Proc. IEEE ICASSP-97*, Munich, Germany, Apr. 1997, pp. 1845–1848.
- [14] Y. Li, L. Cimini, and N. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 902–915, Jul. 1998.
- [15] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Börjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 931–939, Jul. 1998.
- [16] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 7, pp. 1207–1215, Jul. 2000.
- [17] D. Schafhuber, G. Matz, F. Hlawatsch, and P. Loubaton, "MMSE estimation of time-varying channels for DVB-T systems with strong co-channel interference," in *Proc. EUSIPCO*, vol. III, Toulouse, France, Sep. 2002, pp. 25–28.
- [18] V. Mignone and A. Morello, "CD3-OFDM: A novel demodulation scheme for fixed and mobile receivers," *IEEE Trans. Commun.*, vol. 44, no. 9, pp. 1144–1151, Sep. 1996.
- [19] D. Schafhuber, G. Matz, and F. Hlawatsch, "Predictive equalization of time-varying channels for coded OFDM/BFDM systems," in *Proc. IEEE GLOBECOM*, San Francisco, CA, Nov./Dec. 2000, pp. 721–725.
- [20] —, "Adaptive prediction of time-varying channels for coded OFDM systems," in *Proc. IEEE ICASSP*, Orlando, FL, May 2002, pp. 2549–2552.
- [21] S. Thoen, L. Van der Perre, B. Gyselinckx, M. Engels, and H. De Man, "Predictive adaptive loading for HIPERLAN II," in *Proc. IEEE VTC-00 Fall*, Boston, MA, Sep. 2000, pp. 2166–2172.
- [22] F. Duel-Hallen, S. Hu, and H. Hallen, "Long-range prediction of fading signals," *IEEE Signal Process. Mag.*, vol. 17, pp. 62–75, May 2000.
- [23] Y. Liu and S. D. Blostein, "Identification of frequency nonselective fading channels using decision feedback and adaptive linear prediction," *IEEE Trans. Commun.*, vol. 43, pp. 1484–1492, Feb.–Apr. 1995.
- [24] T. Ekman and G. Kubin, "Nonlinear prediction of mobile radio channels: Measurements and MARS model designs," in *Proc. IEEE ICASSP*, Phoenix, AZ, Mar. 1999, pp. 2667–70.
- [25] T.-S. Yang and A. Duel-Hallen, "Adaptive modulation using outdated samples of another fading channel," in *Proc. IEEE WCNC*, Orlando, FL, Mar. 2002, pp. 477–481.
- [26] T. Ekman, "Prediction of mobile radio channels," Ph.D. dissertation, Uppsala Univ., Uppsala, Sweden, 2002.
- [27] G. Matz and F. Hlawatsch, "Time-frequency characterization of random time-varying channels," in *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, B. Boashash, Ed. Oxford, U.K.: Elsevier, 2003, ch. 9.5, pp. 410–419.
- [28] Y. Li and L. Cimini, "Bounds on the interchannel interference of OFDM in time-varying impairments," *IEEE Trans. Commun.*, vol. 49, pp. 401–404, Mar. 2001.
- [29] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. 11, pp. 360–393, 1963.
- [30] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [31] M. Wax and T. Kailath, "Efficient inversion of Toeplitz-block Toeplitz matrix," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, pp. 1218–1221, Oct. 1983.
- [32] S. M. Kay, *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
- [33] G. Matz, "Characterization of non-WSSUS fading dispersive channels," in *Proc. IEEE ICC*, Anchorage, AK, May 2003, pp. 2480–2484.
- [34] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1991.
- [35] D. Schafhuber, M. Rupp, G. Matz, and F. Hlawatsch, "Adaptive identification and tracking of doubly selective fading channels for wireless MIMO-OFDM systems," in *Proc. IEEE SPAWC*, Rome, Italy, Jun. 2003.
- [36] D. Schafhuber, G. Matz, and F. Hlawatsch, "Simulation of wideband mobile radio channels using subsampled ARMA models and multistage interpolation," in *Proc. 11th IEEE Workshop Statistical Signal Processing*, Singapore, Aug. 2001, pp. 571–574.
- [37] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.
- [38] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. CAS-25, no. 4, pp. 772–781, Sep. 1978.



Dieter Schafhuber (S'02) received the Dipl.-Ing. and Dr. Techn. degrees in electrical engineering from Vienna University of Technology, Vienna, Austria, in 1999 and 2004, respectively.

From 2001 to 2004, he was with the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology. Since 2004, he has been with BMW, Munich, Germany. His research interests include the areas of wireless communications and statistical signal processing.



Gerald Matz (M'01) received the Dipl.-Ing., Dr. Techn., and Habilitation degrees in electrical engineering from Vienna University of Technology, Vienna, Austria, in 1994, 2000, and 2004, respectively.

Since 1995, he has been with the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology, currently as a Tenured Associate Professor. His research interests include the areas of wireless communications, statistical signal processing, and time-frequency methods.