Pilot-Assisted Channel Estimation Based on Second-Order Statistics

Frank A. Dietrich, *Student Member, IEEE,* and Wolfgang Utschick

***Abstract*—A survey about linear channel estimation exploiting slowly varying channel properties is given for receivers with multiple antenna elements. The slowly varying channel properties are described by the channel's second-order statistics. First, a detailed comparison of classical linear pilot-based channel estimators, a novel matched filter, and reduced-rank (RR) approaches is made in a common framework with respect to the model they assume for the channel, their performance (MSE), and complexity. The matched filter channel estimator is introduced, which exploits second-order statistics with quadratic order of complexity. For flat correlated Rayleigh fading channels, an analytical performance comparison of all estimators in terms of the uncoded bit error probability is provided. It is a generalization of previous results and shows that the matched filter is a low complexity alternative with good performance at an interesting signal-to-noise ratio (SNR) range. The effects of these approaches on linear equalization are briefly discussed in the context of direct-sequence code division multiple access (DS-CDMA), where the focus is on a generalization of the rake receiver in space and time, which reduces the channel rank based on its second-order statistics, resulting in a complexity reduction of the equalizer. For this discussion, a new notation for the generalized rake is presented, which allows for its interpretation in the context of channel estimation and reveals alternatives for implementation. We conclude that exploiting second-order channel statistics results in significant performance gains, and RR channel estimation should only be used together with equalization in the reduced signal subspace.**

***Index Terms*—CDMA, channel estimation, multiple antennas, rake receiver, rank reduction.**

I. Introduction

FOR the Universal Mobile Telecommunications System (UMTS), a direct-sequence code division multiple access (DS-CDMA) based system, base stations with multiple antenna elements (antenna array) are envisioned for the future to increase system capacity, reduce the number of base stations in cities, or decrease the transmitted power [1]. With multiple antennas, more degrees of freedom are available to improve separation of the users (equalization) as the spatial channel structure can be exploited by adaptive spatial-temporal signal processing. To utilize the additional degrees of freedom in the receiver design, i.e., more parameters in the design criterion, more parameters have to be estimated. Moreover, the design is more sensitive to errors in the parameter estimates. This problem is particularly severe in a cellular system with mobile users, which may travel at rather high speed. In this scenario, only pilot symbols from one slot can be used to estimate the channel due to the fast-varying channel. Although the coefficients of the channel model vary on a high rate, this is not true for all parameters in a more detailed model: Angles of arrivals, delays, and average power of the arriving waves change much more slowly (*long-term parameters*). These slowly varying channel properties can be described using *second-order statistics* of the channel coefficients, i.e., with a stochastic channel model. This property needs to be exploited for channel estimation.

TABLE I
OVERVIEW OF LINEAR CHANNEL ESTIMATORS INTRODUCED IN SECTION III AND THEIR UNDERLYING CHANNEL MODEL (RR = REDUCED RANK)

Channel estimator	Deterministic channel model	Stochastic channel model	Section
Wiener Filter		X	III-A.1
Maximum Likelihood	X		III-A.2
Matched Filter		X	III-A.3
Correlator	X		III-A.4
RR Max. Likelihood	X	(X)	III-B.1
RR Correlator	X	(X)	III-B.2

First, we give a survey of well-known and new channel estimation algorithms and discuss to what extent they are able to exploit slowly varying channel parameters (see Table I and Section III). Classical approaches are the *Wiener filter* (WF), *maximum likelihood* (ML) [2], and *correlator*. Starting with the ML principle, Nicoli *et al.* recently derived a *reduced-rank (RR) ML* (RML) estimator, which relies on the time invariance of long-term channel parameters [3]–[5]. Their work is based on the RML from Stoica *et al.* [6] and presents a generalization of results from [7]. We give a derivation of their approach fitting a common framework to compare it with the *RR correlator* and classical estimators. Jelitto *et al.* [8], [9] also use the principle of rank reduction for equalization. Their arguments follow those of Scharf [10] about RR representation of a random vector. This point of view has been completed for RR channel estimation in [11]. It is applied to multi-input multi-output (MIMO) systems in [12]. From Section III, it will become clear that these RR approaches have a worse performance *and* higher computational complexity (due to rank optimization and RR approximation) than the Wiener estimator for known second-order statistics. This comparison with its important conclusions has not been made in the literature. Other RR estimators are known for orthogoanl frequency division multiplexing (OFDM) [13] or as

Manuscript received August 25, 2003; revised March 31, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Martin Haardt.

The authors are with the Institute for Circuit Theory and Signal Processing, Munich University of Technology, 80290 Munich, Germany (e-mail: Dietrich@nws.ei.tum.de).

Digital Object Identifier 10.1109/TSP.2004.842176

a low-rank Wiener filter [10], whose primary goal is to reduce complexity of the estimator.

Note that the RR estimators are based on a *deterministic channel model* in their core, as they rely on the ML principle (RML—Section III-B1) or use the correlator as a first stage (RR correlator—Section III-B2). However, implicitly, they exploit the *long-term channel properties* given by the channel correlation matrix, i.e., they exploit the stationary *stochastic nature of the channel* from its second-order statistics (see Table I).

The known estimation approaches exploiting second-order channel statistics are rather complex (cubic order of complexity). We derive a *novel channel estimator* from the *matched filter* (MF) principle (Section III-A), which takes advantage of the long-term properties with quadratic order of complexity.

For now, the focus is on pilot sequence-based estimators, which estimate a block-wise constant channel. They may be used as a basis or a good initialization for the following valuable extension, which are beyond the scope of this paper: interpolation of channel estimates [14], tracking of the channel coefficients, blind techniques [15], decision-directed channel estimation, joint channel estimation and equalization [16], and applications of the turbo principle to estimation and equalization [17].

To investigate the impact of channel estimation on equalization, RR equalizers are particularly interesting as they are known to be less sensitive to channel estimation errors [18]. Most of them exploit properties of the currently available channel realization for equalization, e.g., the multistage Wiener filter [18]. Our focus is on a class of receivers, which reduces complexity and sensitivity based on the slowly varying channel parameters mentioned above. The most famous representative is the *temporal rake* [1], [19], [20], which selects temporal channel taps based on estimates of the average power-delay profile. It was extended to space and time by Brunner *et al.* [21]–[24] based on results from Naguib [25]. Some aspects related to the number of fingers (optimum channel rank) relevant for the *generalized space-time rake receiver* are discussed in [11] and [26]. It was shown to be an efficient and simple single-user detector based on a linear minimum mean square error (LMMSE) or matched filter equalizer using channel estimates from a correlator. To reduce computations and avoid an eigenvalue decomposition, which is necessary for generalizing the rake to space and time, a low-cost approximation of eigenvectors was proposed in [27]. The architecture was further explored together with a nonlinear equalizer in [28]. These rake receiver structures have been described rather intuitively in the references and treated in special cases only. We further *generalize the rake concept* and present a *novel notation* as a finite impulse response (FIR) structure, revealing *alternatives for implementation*. Furthermore, we work out the relation of the (generalized) rake concept to general channel estimation methods (the RR correlator in particular) that exploit second-order statistics (see Section V). Here, we follow the classical approach of a separate optimization of channel estimation and equalization with a discussion of the impact of RR channel estimation on equalization.

In Section II, the channel and DS-CDMA signal model for data and pilot channel are introduced in matrix vector notation, where the focus is on a physical interpretation of the second-order channel statistics. The full-rank and RR channel estimators are described and compared w.r.t. *mean square error* (MSE), complexity, and sensitivity to estimation of the second-order statistics in Section III (see Table I for a summary). A detailed analysis of their *bit error probability* (BEP) and MSE in an equivalent flat fading wireless channel is presented in Section IV, for which the estimators and the MSE of the channel estimates have a simple and transparent form. The *temporal rake* is *generalized* to spatial-temporal processing in Section V. The LMMSE equalizer and generalized rake are compared in terms of uncoded bit error rate (BER) using different channel estimates in Section VI. Our survey is concluded in Section VII. More detailed derivations for the matched filter, RML channel estimator, RR correlator, and the BEP for imperfect channel knowledge are given in the Appendixes.

The following notation is used.

1) Special matrices: We define the $M \times M$ identity matrix $\boldsymbol{I}_M$, the canonical basis vector $\boldsymbol{e}_n$ as nth column of an identity matrix, an $M \times N$ matrix $\boldsymbol{0}_{M\times N}$ of zeros, the $M \times (M+N)$ selection matrix $\boldsymbol{J}_{(\ell,M,N)} = [\boldsymbol{0}_{M\times\ell}, \boldsymbol{I}_M, \boldsymbol{0}_{M\times(N-\ell)}]$, a (block) diagonal matrix $\operatorname{diag}(\boldsymbol{a})(\operatorname{diag}(\{\boldsymbol{A}_1,\ldots,\boldsymbol{A}_N\}))$ with elements of $\boldsymbol{a}$ (matrices $\boldsymbol{A}_n$) on its diagonal, and the vector of N ones $\boldsymbol{o}_N$. Vectors are always arranged in a column.
2) Matrix operations: We define the Kronecker product $\otimes$, vec-operator $\operatorname{vec}(\boldsymbol{A})$ stacking the columns of matrix $\boldsymbol{A}$, trace $\operatorname{tr}(\boldsymbol{A})$, and the complex conjugate transpose $\boldsymbol{A}^{\mathrm{H}}$, transpose $\boldsymbol{A}^{\mathrm{T}}$ and complex conjugate $\boldsymbol{A}^*$ of a matrix. The relation [29]
$$\operatorname{vec}(\boldsymbol{A}\boldsymbol{D}\boldsymbol{B}^{\mathrm{T}}) = (\boldsymbol{B}\otimes\boldsymbol{A})\operatorname{vec}(\boldsymbol{D}) = (\boldsymbol{B}\odot\boldsymbol{A})\boldsymbol{d} \qquad (1)$$
is needed for some derivations, where the second equality is true if $\boldsymbol{D}$ is a diagonal matrix with diagonal $\boldsymbol{d}$. The Khatri–Rao product is defined as $\boldsymbol{A}\odot\boldsymbol{B} = [\boldsymbol{a}_1\otimes\boldsymbol{b}_1,\ldots,\boldsymbol{a}_n\otimes\boldsymbol{b}_n]$ ($\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ are the ith column of $\boldsymbol{A}$ and $\boldsymbol{B}$).
3) Others: We define convolution $*$, Dirac distribution $\delta(t)$, Kronecker function $\delta[n]$, expectation operator $\mathrm{E}[A]$, a circular symmetric complex Gauss distribution of a complex random variable $\mathcal{N}_{\mathrm{c}}(\boldsymbol{\mu},\boldsymbol{R})$, $\lceil a\rceil$ the smallest integer greater or equal than a, and the weighted vector norm $\|\boldsymbol{a}\|_{\boldsymbol{A}}^2 = \boldsymbol{a}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{a}$ with $\boldsymbol{A}$ positive semidefinite.

II. Signal and Channel Model

We consider a wireless DS-CDMA[1] link with a single antenna at the transmitter and M antennas at the receiver. All signals and channel impulse responses are given in their equivalent baseband representation.

[1]As the generalized rake (Section V) is generally associated with DS-CDMA systems and its generalization from a system without spreading is not straightforward, we introduce the DS-CDMA signal model. Note that it is not a necessary assumption for the derivations and comparison of channel estimation concepts.

1) Transmit Signal: Pilot and data symbols are time-multiplexed and separated by a prefix (see below). The data symbols $s_\mathrm{d}[i]$ of the user of interest with power $P_\mathrm{d} = \mathrm{E}[|s_\mathrm{d}[i]|^2]$ are spread with a short spreading sequence $c[q]$ of length Q (spreading factor or processing gain) and chip period T_c. The resulting signal

$$x_\mathrm{d}(t) = \sum_i s_\mathrm{d}[i] \sum_{q=0}^{Q-1} c[q] p'(t - iQT_\mathrm{c} - qT_\mathrm{c}) \quad (2)$$

is bandlimited by an impulse shape $p'(t)$, e.g., a root-raised cosine. A discrete-time representation with sampling-period T_c, as needed later, is $x_\mathrm{d}[n] = x_\mathrm{d}(nT_\mathrm{c})$. Equivalently, N_p pilot symbols at chip level are transmitted as[2]

$$x_\mathrm{p}(t) = \sum_{q=1}^{N_\mathrm{p}} s_\mathrm{p}[q] p'(t - qT_\mathrm{c}) \quad (3)$$

with symbol power $P_\mathrm{p} = \mathrm{E}[|s_\mathrm{p}[i]|^2]$.

2) Channel Model: The time-invariant channel impulse response is modeled by a tapped delay line with L temporal taps and delays τ_ℓ:

$$\boldsymbol{h}'(t) = \sum_{\ell=1}^{L} \boldsymbol{A}_\ell \boldsymbol{\xi}_\ell \delta(t - \tau_\ell) \quad \in \mathbb{C}^M. \quad (4)$$

The spatial channel structure of each tap is represented by the array steering vectors of W discrete wavefronts impinging from different azimuth directions $\phi_{w,\ell}$, which are collected in $\boldsymbol{A}_\ell = [\boldsymbol{a}_{1,\ell}(\phi_{1,\ell}), \ldots, \boldsymbol{a}_{W,\ell}(\phi_{W,\ell})] \in \mathbb{C}^{M \times W}$. The random vector $\boldsymbol{\xi}_\ell \in \mathbb{C}^W \sim \mathcal{N}_\mathrm{c}(\boldsymbol{0}, P_\tau(\tau_\ell)\boldsymbol{I}_W)$ describes a slot-wise independent fading (Rayleigh fading, i.e., no line of sight, is assumed for simplicity). $P_\tau(\tau_\ell)$ is the power-delay profile of the channel [20]. Asynchronous signals can be modeled by distinct minimum channel delays $\min_\ell\{\tau_\ell\}$ among the users.

3) Receive Signal: The received data signal from 1 user is

$$\boldsymbol{y}_\mathrm{d}(t) = x_\mathrm{d}(t) * \boldsymbol{h}'(t) + \boldsymbol{\eta}_\mathrm{d}(t) \in \mathbb{C}^M \quad (5)$$

disturbed by additive noise $\boldsymbol{\eta}_\mathrm{d}(t) \sim \mathcal{N}_\mathrm{c}(\boldsymbol{0}, \boldsymbol{R}_{\boldsymbol{\eta}_\mathrm{s}})$—equivalently, $\boldsymbol{\eta}_\mathrm{p}(t) \sim \mathcal{N}_\mathrm{c}(\boldsymbol{0}, \boldsymbol{R}_{\boldsymbol{\eta}_\mathrm{s}})$ for the pilot channel—with a regular spatial covariance matrix $\boldsymbol{R}_{\boldsymbol{\eta}_\mathrm{s}} \in \mathbb{C}^{M \times M}$.

The receive filter matched to the chip waveform and the impulse shape $p'(t)$ are included in $p(t)$, which is assumed to be a raised cosine impulse. Sampling at the chip rate $1/T_\mathrm{c}$ yields discrete-time approximations for the receive data and pilot signals

$$\boldsymbol{y}_\mathrm{d}[n] = \sum_i s_\mathrm{d}[i] \sum_{q=0}^{Q-1} c[q] \sum_{\ell=1}^{L} \boldsymbol{A}_\ell \boldsymbol{\xi}_\ell \times p(nT_\mathrm{c} - iQT_\mathrm{c} - qT_\mathrm{c} - \tau_\ell) + \boldsymbol{\eta}_\mathrm{d}[n] \quad (6)$$

$$\boldsymbol{y}_\mathrm{p}[n] = \sum_{q=1}^{N_\mathrm{p}} s_\mathrm{p}[q] \sum_{\ell=1}^{L} \boldsymbol{A}_\ell \boldsymbol{\xi}_\ell \times p(nT_\mathrm{c} - qT_\mathrm{c} - \tau_\ell) + \boldsymbol{\eta}_\mathrm{p}[n]. \quad (7)$$

4) Channel Characteristics of Two Time Scales: The impulse shape and receive filter $p(t)$ are now considered as part of the channel resulting in the *discrete-time channel model* of $\boldsymbol{h}(t) = \boldsymbol{h}'(t) * p(t)$

$$\boldsymbol{h}[n] = \boldsymbol{h}(nT_\mathrm{c}) = \sum_{\ell=1}^{L} \boldsymbol{A}_\ell \boldsymbol{\xi}_\ell p(nT_\mathrm{c} - \tau_\ell) \quad \in \mathbb{C}^M. \quad (8)$$

The largest channel delay is $\tau_\mathrm{max} = \max_\ell\{\tau_\ell\}$. The noncausal raised-cosine waveform $p(t)$ with infinite support is modeled by the receiver using a truncated (windowed) FIR approximation $\boldsymbol{p}_\ell = [p(-\beta T_\mathrm{c} - \tau_\ell), p(T_\mathrm{c} - \beta T_\mathrm{c} - \tau_\ell), \ldots, p(\lceil(\tau_\mathrm{max})/(T_\mathrm{c})\rceil T_\mathrm{c} + \beta T_\mathrm{c} - \tau_\ell)]^\mathrm{T} \in \mathbb{R}^{L_\mathrm{c}}$ of length $L_\mathrm{c} = 1 + \lceil(\tau_\mathrm{max})/(T_\mathrm{c})\rceil + 2\beta$, where β describes the window size. Now, the channel matrix is given as

$$\begin{aligned}\boldsymbol{H} &= [\boldsymbol{h}[0], \ldots, \boldsymbol{h}[L_\mathrm{c} - 1]] \\ &= \sum_{\ell=1}^{L} \sum_{w=1}^{W} \boldsymbol{a}_{w,\ell} \xi_{w,\ell} \boldsymbol{p}_\ell^\mathrm{T} \\ &= \boldsymbol{A} \boldsymbol{\Xi} \boldsymbol{P} \in \mathbb{C}^{M \times L_\mathrm{c}}\end{aligned} \quad (9)$$

where $\boldsymbol{\Xi} = \mathrm{diag}(\mathrm{vec}([\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_L]))$, the matrix of array steering vectors $\boldsymbol{A} = [\boldsymbol{A}_1, \ldots, \boldsymbol{A}_L]$, and

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{o}_W \otimes \boldsymbol{p}_1^\mathrm{T} \\ \vdots \\ \boldsymbol{o}_W \otimes \boldsymbol{p}_L^\mathrm{T} \end{bmatrix} \in \mathbb{C}^{WL \times L_\mathrm{c}}. \quad (10)$$

We distinguish *two time scales*: Amplitudes and phases in $\boldsymbol{\xi}_\ell$ vary on the scale of a carrier-wavelength (*short-term* or small-scale), whereas delays $\tau_{w,\ell}$, angles of arrival $\phi_{w,\ell}$, and average receive power change on a much larger scale (*long-term* or large-scale channel properties) [3], [21], [30]–[32]. Whether the notion of time or spatial scale is preferred depends on the context as both are linked via the velocity of the movement, e.g., of the transmitter or receiver. The first- and second-order moments of complex Gaussian-distributed channel coefficients describe all long-term, i.e., slowly varying, channel properties. This dependency is also true for other channel distributions, where these moments do not contain all information. We see this fact explicitly for the correlation matrix of

$$\boldsymbol{h} = \mathrm{vec}(\boldsymbol{H}) = (\boldsymbol{P}^\mathrm{T} \otimes \boldsymbol{A})\mathrm{vec}(\boldsymbol{\Xi}) = (\boldsymbol{P}^\mathrm{T} \odot \boldsymbol{A})\boldsymbol{\xi} \in \mathbb{C}^{ML_\mathrm{c}} \quad (11)$$

which is

$$\boldsymbol{R}_{\boldsymbol{h}} = \mathrm{E}[\mathrm{vec}(\boldsymbol{H})\mathrm{vec}(\boldsymbol{H})^\mathrm{H}] = (\boldsymbol{P}^\mathrm{T} \odot \boldsymbol{A})\mathrm{E}[\boldsymbol{\xi}\boldsymbol{\xi}^\mathrm{H}](\boldsymbol{P}^\mathrm{T} \odot \boldsymbol{A})^\mathrm{H}$$

applying (1) with $\boldsymbol{\xi} = [\boldsymbol{\xi}_1^\mathrm{T}, \ldots, \boldsymbol{\xi}_L^\mathrm{T}]^\mathrm{T} \in \mathbb{C}^{LW}$ and $\mathrm{E}[\boldsymbol{\xi}\boldsymbol{\xi}^\mathrm{H}] = \mathrm{diag}(\{P_\tau(\tau_\ell)\}_{\ell \in \{1,\ldots,L\}}) \otimes \boldsymbol{I}_W$ [3]. Its eigenvalue decomposition (EVD) is $\boldsymbol{R}_{\boldsymbol{h}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\mathrm{H}$, where $\boldsymbol{\Lambda}$ is a diagonal matrix with eigenvalues $\lambda_m, m \in \{1, \ldots, ML_\mathrm{c}\}$ and $\lambda_m \geq \lambda_{m+1}$. The

[2]In practice, pilot symbols are also spread for reasons of implementation in a multiuser system. In order to simplify notation, we consider pilot symbols ("chips") at the chip level as used for channel estimation, i.e., they include the spreading sequence.

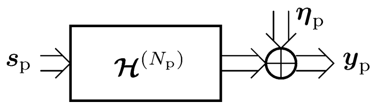


Fig. 1. System model for the pilot channel (Matrix-vector notation).

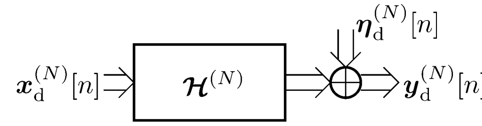


Fig. 2. System model for the data channel (Matrix-vector notation).

correlations are determined only by the power-delay distribution in the current environment, the spatial and temporal channel structure, and the impulse shape. Thus, channel estimation and receiver processing based on long-term channel properties have to be based on the channel's first- and second-order moments. In the sequel, we assume wide-sense stationarity of the channel $\boldsymbol{h}$, i.e., within the observation time for channel estimation and data detection, the long-term channel properties do not change. Generally, $\boldsymbol{R_h}$ is of full algebraic rank but with eigenvalues λ_m decaying quickly [31].

5) Matrix-Vector Notation: For the description and derivation of channel estimation algorithms, we introduce a more compact notation. As we focus on single-user channel estimation and detection algorithms,[3] only one user of interest is considered in the following models, and the multiuser interference is modeled as part of the noise $\boldsymbol{\eta}_{\mathrm{p}}[n]$ and $\boldsymbol{\eta}_{\mathrm{d}}[n]$. Note that all described channel estimators and linear equalizers can be generalized to multiple users except for the generalized rake, which is intended as a single-user receiver only.

The *pilot signal* can be written as (Fig. 1)

$$\boldsymbol{y}_{\mathrm{p}} = \boldsymbol{\mathcal{H}}^{(N_{\mathrm{p}})}\boldsymbol{s}_{\mathrm{p}} + \boldsymbol{\eta}_{\mathrm{p}} = \boldsymbol{S}_{\mathrm{p}}\boldsymbol{h} + \boldsymbol{\eta}_{\mathrm{p}} \in \mathbb{C}^{MN_{\mathrm{p}}} \tag{12}$$

with channel coefficients $\boldsymbol{h} = \mathrm{vec}(\boldsymbol{H})$, pilot symbols $\boldsymbol{s}_{\mathrm{p}} = [s_{\mathrm{p}}[2-L_{\mathrm{c}}], \ldots, s_{\mathrm{p}}[N_{\mathrm{p}}]]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{p}}+L_{\mathrm{c}}-1}$ in $\boldsymbol{S}_{\mathrm{p}} = \boldsymbol{S}'_{\mathrm{p}} \otimes \boldsymbol{I}_M \in \mathbb{C}^{MN_{\mathrm{p}} \times ML_{\mathrm{c}}}$, and Toeplitz matrix

$$\boldsymbol{S}'_{\mathrm{p}} = \begin{bmatrix} s_{\mathrm{p}}[1] & \cdots & s_{\mathrm{p}}[2-L_{\mathrm{c}}] \\ \vdots & & \vdots \\ s_{\mathrm{p}}[N_{\mathrm{p}}] & \cdots & s_{\mathrm{p}}[N_{\mathrm{p}}-L_{\mathrm{c}}+1] \end{bmatrix} \in \mathbb{C}^{N_{\mathrm{p}} \times L_{\mathrm{c}}}. \tag{13}$$

The block Toeplitz channel matrix is

$$\begin{aligned} \boldsymbol{\mathcal{H}}^{(N)} &= \sum_{\ell=0}^{L_{\mathrm{c}}-1} \boldsymbol{J}_{(\ell,N,L_{\mathrm{c}}-1)} \otimes \boldsymbol{h}[L_{\mathrm{c}}-\ell-1] \in \mathbb{C}^{MN \times N+L_{\mathrm{c}}-1} \\ &= \begin{bmatrix} \boldsymbol{h}[L_{\mathrm{c}}-1] & \cdots & \boldsymbol{h}[0] & & \\ & \ddots & & \ddots & \\ & & \boldsymbol{h}[L_{\mathrm{c}}-1] & \cdots & \boldsymbol{h}[0] \end{bmatrix}. \end{aligned} \tag{14}$$

Here, a prefix of $L_{\mathrm{c}} - 1$ additional pilot symbols is inserted to avoid interference between pilot and data signal. The noise and receive signal vector are $\boldsymbol{\eta}_{\mathrm{p}} = [\boldsymbol{\eta}_{\mathrm{p}}[1]^{\mathrm{T}}, \ldots, \boldsymbol{\eta}_{\mathrm{p}}[N_{\mathrm{p}}]^{\mathrm{T}}]^{\mathrm{T}}$ and $\boldsymbol{y}_{\mathrm{p}} = [\boldsymbol{y}_{\mathrm{p}}[1]^{\mathrm{T}}, \ldots, \boldsymbol{y}_{\mathrm{p}}[N_{\mathrm{p}}]^{\mathrm{T}}]^{\mathrm{T}}$, respectively. The noise is distributed as $\boldsymbol{\eta}_{\mathrm{p}} \sim \mathcal{N}_{\mathrm{c}}(\boldsymbol{0}, \boldsymbol{R_\eta})$. For the simplifying assumption of temporally white noise processes, the covariance matrix is $\boldsymbol{R_\eta} = \boldsymbol{I}_{N_{\mathrm{p}}} \otimes \boldsymbol{R}_{\boldsymbol{\eta}_{\mathrm{s}}}$ with variance $\sigma_{\mathrm{n}}^2 = \mathrm{trace}(\boldsymbol{R_\eta})/(MN_{\mathrm{p}})$.

[3]It is well known that single-user detection approaches are suboptimum for multiuser CDMA systems but aim at solutions of low-complexity typically considering each user separately.

The *receive data signal* is (see Fig. 2)

$$\boldsymbol{y}_{\mathrm{d}}^{(N)}[n] = \boldsymbol{\mathcal{H}}^{(N)}\boldsymbol{x}_{\mathrm{d}}^{(N)}[n] + \boldsymbol{\eta}_{\mathrm{d}}^{(N)}[n] \in \mathbb{C}^{MN} \tag{15}$$

where $\boldsymbol{y}_{\mathrm{d}}^{(N)} = [\boldsymbol{y}_{\mathrm{d}}[n]^{\mathrm{T}}, \ldots, \boldsymbol{y}_{\mathrm{d}}[n+N-1]^{\mathrm{T}}]^{\mathrm{T}}$. Spreading the symbols $s_{\mathrm{d}}[i]$ by $\boldsymbol{c} = [c[0], \ldots, c[Q-1]]^{\mathrm{T}}$ yields the chip-level signal at the transmitter $\boldsymbol{x}_{\mathrm{d}}^{(N)}[n] = [x_{\mathrm{d}}[n-L_{\mathrm{c}}+1], \ldots, x_{\mathrm{d}}[n], \ldots, x_{\mathrm{d}}[n+N-1]]^{\mathrm{T}} \in \mathbb{C}^{N+L_{\mathrm{c}}-1}$. Index N is introduced to adjust notation to the equalizer length (see Section V).

III. Channel Estimation Methods

We restrict our survey to linear channel estimators based on pilot symbols, estimating the block-wise constant channel coefficients $\boldsymbol{h}$ by

$$\hat{\boldsymbol{h}} = \boldsymbol{W}\boldsymbol{y}_{\mathrm{p}} \in \mathbb{C}^{ML_{\mathrm{c}}}. \tag{16}$$

Linear estimators can be *classified* according to the channel model they assume, i.e., *stochastic* or *deterministic*, and whether they consider the channel to be full rank or low rank.[4] *Reduced-rank* or low-rank estimators assume a more detailed model of $\boldsymbol{h} = \mathrm{vec}(\boldsymbol{H}) \in \mathbb{C}^{ML_{\mathrm{c}}}$ or $\boldsymbol{R_h}$, e.g., its eigenvalues or rank. If the model is more accurate, it results in a performance benefit compared with full-rank estimators, which are based on the same signal model as the corresponding low-rank approach. An overview of the approaches introduced in this section and the underlying channel models is given in Table I.

A. Full-Rank Channel Estimation

1) Bayesian Approach: The linear full rank estimator minimizing the mean square error is[5]

$$\boldsymbol{W}_{\mathrm{WF}} = \underset{\boldsymbol{W}}{\mathrm{argmin}}\, \mathrm{E}\left[\|\hat{\boldsymbol{h}} - \boldsymbol{h}\|_2^2\right]. \tag{17}$$

This LMMSE estimator is often referred to as the Wiener filter. It is a Bayesian approach with a quadratic risk function, i.e., a conditional mean estimator, if channel coefficients and noise are jointly Gaussian distributed [33], [34]. The well-known solution (for zero mean channels) is

$$\boldsymbol{W}_{\mathrm{WF}} = \left(\boldsymbol{I}_{ML_{\mathrm{c}}} + \boldsymbol{R_h}\boldsymbol{S}_{\mathrm{p}}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\eta}}^{-1}\boldsymbol{S}_{\mathrm{p}}\right)^{-1}\boldsymbol{R_h}\boldsymbol{S}_{\mathrm{p}}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\eta}}^{-1}. \tag{18}$$

It depends on the noise covariance matrix, the pilot sequence, and the channel correlation matrix, i.e., *long-term channel properties* (see Fig. 3). The Wiener filter finds the best tradeoff between bias and variance of the estimate for every signal to noise

[4]If $\boldsymbol{R_h}$ is of rank R, every realization of the channel $\boldsymbol{h}$ is an element of the R-dimensional subspace spanned by the R eigenvectors of $\boldsymbol{R_h}$ belonging to nonzero eigenvalues.

[5]The expectation is taken with respect to the channel $\boldsymbol{h}$ and noise $\boldsymbol{\eta}_{\mathrm{p}}$.

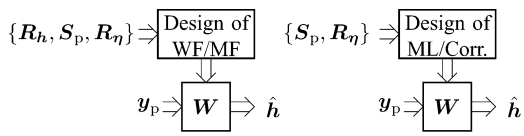


Fig. 3. Full-rank channel estimators rely on different amount of information about the channel parameters.

ratio (SNR). If the channel covariance matrix $\boldsymbol{R_h}$ is of rank R, the Wiener filter (18) is of rank R as well. The following channel estimation concepts can be viewed as suboptimum approximations of the Wiener filter.

2) Maximum Likelihood (Zero Forcing): For complex Gaussian noise and interference the maximum likelihood principle [2] yields the following cost function and solution:

$$\hat{\boldsymbol{h}}_{\mathrm{ML}} = \underset{\boldsymbol{h}}{\operatorname{argmin}} \|\boldsymbol{y}_{\mathrm{p}} - \boldsymbol{S}_{\mathrm{p}}\boldsymbol{h}\|^2_{\boldsymbol{R}_{\boldsymbol{\eta}}^{-1}}. \tag{19}$$

The estimate is unbiased and minimum variance and, thus, achieves the Cramér–Rao bound [34]. The Wiener filter (18) converges to the ML solution for $\sigma_{\mathrm{n}}^2 \to 0$, which is also known as zero forcing or weighted least squares. It is given by

$$\hat{\boldsymbol{h}}_{\mathrm{ML}} = \left(\boldsymbol{S}_{\mathrm{p}}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\eta}}^{-1}\boldsymbol{S}_{\mathrm{p}}\right)^{-1}\boldsymbol{S}_{\mathrm{p}}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\eta}}^{-1}\boldsymbol{y}_{\mathrm{p}} = \boldsymbol{h} + \boldsymbol{\varepsilon} \tag{20}$$

with covariance matrix of the estimation error $\boldsymbol{R_\varepsilon} = \mathrm{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{H}}] = (\boldsymbol{S}_{\mathrm{p}}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\eta}}^{-1}\boldsymbol{S}_{\mathrm{p}})^{-1}$. The corresponding ML estimator is (see Fig. 3)

$$\boldsymbol{W}_{\mathrm{ML}} = \left(\boldsymbol{S}_{\mathrm{p}}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\eta}}^{-1}\boldsymbol{S}_{\mathrm{p}}\right)^{-1}\boldsymbol{S}_{\mathrm{p}}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\eta}}^{-1}. \tag{21}$$

3) Matched Filter: The simplest estimation approach is to optimize for the filter with maximum cross-correlation and small noise amplification (see Fig. 3) [35]

$$\boldsymbol{W}_{\mathrm{MF}} = \underset{\boldsymbol{W}}{\operatorname{argmax}} \frac{|\mathrm{E}[\boldsymbol{h}^{\mathrm{H}}\hat{\boldsymbol{h}}]|^2}{\mathrm{E}\left[\|\boldsymbol{W}\boldsymbol{\eta}\|_2^2\right]} = \frac{\sigma_{\mathrm{n}}^2}{N_{\mathrm{p}}P_{\mathrm{p}}}\boldsymbol{R_h}\boldsymbol{S}_{\mathrm{p}}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\eta}}^{-1}. \tag{22}$$

The criterion presents the generalization of the standard matched filter criterion used for estimating scalar random variables, e.g., data detection [20] to random vectors [35]. Its solution is derived in Appendix A. The arbitrary scalar factor in the solution is chosen as $\alpha = (\sigma_{\mathrm{n}}^2)/(N_{\mathrm{p}}P_{\mathrm{p}})$ to be consistent with the ML approach for flat uncorrelated channels. Although the MF principle is often used for channel equalization or signal detection, it has apparently been ignored for channel estimation. In Sections IV and VI, we show that it offers a significant performance gain for low SNR compared to maximum likelihood.

The estimator does not cancel intersymbol interference in the channel estimates. After noise whitening, it correlates the received signal with the pilot sequence. This estimate is then weighted with the channel correlation matrix. The Wiener filter (18) scaled by σ_{n}^2 converges to the matched filter for low SNR.

4) Correlator: For uncorrelated channel coefficients $\boldsymbol{R_h} = \boldsymbol{I}_{ML_{\mathrm{c}}}$ and white noise $\boldsymbol{R_\eta} = \sigma_{\mathrm{n}}^2\boldsymbol{I}_{MN_{\mathrm{p}}}$, the matched filter (22) yields the correlator

$$\boldsymbol{W}_{\mathrm{C}} = \frac{1}{N_{\mathrm{p}}P_{\mathrm{p}}}\boldsymbol{S}_{\mathrm{p}}^{\mathrm{H}} \tag{23}$$

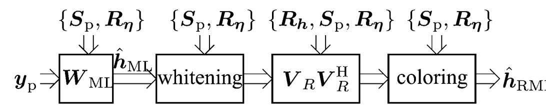


Fig. 4. RML channel estimation.

which is the standard channel estimator in most CDMA systems due to its simplicity. It does not take into account channel correlation properties (see Fig. 3).

B. Reduced-Rank Channel Estimation

Channel estimators based on a deterministic signal model like the ML-estimator or correlator do not exploit correlations in the channel coefficients. Particularly, their variance tends to infinity for low SNR (high noise) [2]. The principle of *reducing the number of parameters* to be estimated using a new signal model is common to the following two methods. Thus, *estimation variance* [2] is reduced at the expense of introducing a *bias*. It is also known from the mathematical literature as "ridge regression" (see [36] and references therein).

1) Reduced-Rank Maximum Likelihood: The ML estimate provides a sufficient statistic for the channel coefficients [34] with respect to model (12), i.e., no information about the channel is lost. Therefore, it can be used to improve the channel estimate by means of *post-processing* based on a more *detailed model* [3], [5]. It is assumed that the channel correlation matrix $\boldsymbol{R_h}$ has rank R or can be approximated by its best rank R approximation truncating its $(ML_{\mathrm{c}} - R)$ smallest eigenvalues [30]. The channel is modeled by a slot-dependent (time-variant or short-term) parameter vector $\boldsymbol{\zeta}^f \in \mathbb{C}^R$ of slot[6] f and a constant matrix $\boldsymbol{U}_{\mathrm{ST}} \in \mathbb{C}^{ML_{\mathrm{c}}\times R}$ of rank R describing the R-dimensional subspace, which contains the channel coefficients

$$\boldsymbol{h}^f = \boldsymbol{U}_{\mathrm{ST}}\boldsymbol{\zeta}^f, \quad \boldsymbol{U}_{\mathrm{ST}} \in \mathbb{C}^{ML_{\mathrm{c}}\times R} \text{ and } \boldsymbol{\zeta}^f \in \mathbb{C}^R. \tag{24}$$

This model is identical to (11) if the rank of $\boldsymbol{P}^{\mathrm{T}} \odot \boldsymbol{A}$ is R. Only R slot-dependent parameters have to be estimated compared with ML_{c} in (9) or LW elements of $\boldsymbol{\zeta}$ in (11). We will see below that $\boldsymbol{U}_{\mathrm{ST}}$ depends only on channel and noise statistics, which are assumed perfectly known asymptotically.

The estimates $[\hat{\boldsymbol{U}}_{\mathrm{ST}}, \hat{\boldsymbol{\zeta}}^f]$ are obtained from

$$\min_{\boldsymbol{U}_{\mathrm{ST}},\boldsymbol{\zeta}^f} \lim_{F\to\infty} \frac{1}{F}\sum_{f=1}^{F} \left\| \hat{\boldsymbol{h}}_{\mathrm{ML}}^f - \boldsymbol{U}_{\mathrm{ST}}\boldsymbol{\zeta}^f \right\|^2_{\boldsymbol{R}_{\boldsymbol{\varepsilon}^{-1}}} \tag{25}$$

which is equivalent to maximizing the *likelihood function* for estimation of the parameters in (24) based on the full-rank ML estimate in (20) with error covariance matrix $\boldsymbol{R_\varepsilon} = (\boldsymbol{S}_{\mathrm{p}}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\eta}}^{-1}\boldsymbol{S}_{\mathrm{p}})^{-1}$. We assume that the spatial-temporal channel structure in $\boldsymbol{P}^{\mathrm{T}} \odot \boldsymbol{A}$ (11) is time-invariant, and an infinite number of slots $F \to \infty$ are available for estimating $\boldsymbol{U}_{\mathrm{ST}}$.

The RML estimate is (see Fig. 4)

$$\hat{\boldsymbol{h}}_{\mathrm{RML}}^f = \hat{\boldsymbol{U}}_{\mathrm{ST}}\hat{\boldsymbol{\zeta}}^f = \boldsymbol{P}_{\mathrm{RML}}\hat{\boldsymbol{h}}_{ML}^f. \tag{26}$$

[6]For the derivations in this section, we denote the slot dependency of the channel parameters, e.g., $\boldsymbol{h}$, explicitly as $\boldsymbol{h}^f$.

It is obtained by *post-processing* with

$$\boldsymbol{P}_{\mathrm{RML}} = \boldsymbol{R}_{\boldsymbol{\varepsilon}}^{+1/2}\boldsymbol{V}_R\boldsymbol{V}_R^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\varepsilon}}^{-1/2} \tag{27}$$

where $\boldsymbol{V}_R = \boldsymbol{R}_{\boldsymbol{\varepsilon}}^{-1/2}\hat{\boldsymbol{U}}_{\mathrm{ST}}$ are the R eigenvectors associated with $\boldsymbol{R}_{\boldsymbol{\varepsilon}}^{-1/2}\boldsymbol{R}_{\boldsymbol{h}}\boldsymbol{R}_{\boldsymbol{\varepsilon}}^{-1/2}$, i.e., the correlation matrix of the whitened channel estimates $\hat{\boldsymbol{h}}_{\mathrm{ML}}$. The transformation $\boldsymbol{V}_R^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\varepsilon}}^{-1/2}$ into the R-dimensional subspace decorrelates the whitened ML channel estimates. The derivation is similar to [3] and given in Appendix B.

The *RML* estimator $\boldsymbol{W}_{\mathrm{RML}}$ can be written as (see Fig. 4)

$$\boldsymbol{W}_{\mathrm{RML}} = \boldsymbol{P}_{\mathrm{RML}}\boldsymbol{W}_{\mathrm{ML}}. \tag{28}$$

Simeone *et al.* [5] also derived spatial, temporal, and separate spatial/temporal RML estimators for complexity reduction, as well as this joint spatial-temporal RR modeling and estimation approach.

The *performance advantage* of the RML is due to a *bias-variance tradeoff*, which is controlled by its rank R (see Sections III-D and IV). It is important to note that the *optimal* (MSE minimizing) *post-processing* of the full-rank ML estimate exploiting second-order statistics, i.e., long-term channel properties, is a Wiener filter $\boldsymbol{P}_{\mathrm{WF}}$

$$\boldsymbol{P}_{\mathrm{WF}} = \underset{\boldsymbol{P}}{\operatorname{argmin}}\, \mathrm{E}\left[\|\boldsymbol{P}\hat{\boldsymbol{h}}_{\mathrm{ML}} - \boldsymbol{h}\|_2^2\right]. \tag{29}$$

It achieves the *optimum bias-variance tradeoff*. The cascade of an ML estimator and a Wiener post-processing yields the Wiener estimator $\boldsymbol{W}_{\mathrm{WF}}$ (18), as the ML estimator provides a sufficient statistic

$$\boldsymbol{W}_{\mathrm{WF}} = \boldsymbol{P}_{\mathrm{WF}}\boldsymbol{W}_{\mathrm{ML}}. \tag{30}$$

With the eigenvalue decomposition $\boldsymbol{R}_{\boldsymbol{h}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}$ (29) can be written as

$$\boldsymbol{P}_{\mathrm{WF}} = \boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{H}}\left(\boldsymbol{S}_{\mathrm{p}}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\eta}}^{-1}\boldsymbol{S}_{\mathrm{p}}\right)^{-1}\boldsymbol{U}\boldsymbol{\Lambda}^{-1} + \boldsymbol{I}_{ML_{\mathrm{c}}}\right)^{-1}\boldsymbol{U}^{\mathrm{H}} \tag{31}$$

assuming that $\boldsymbol{R}_{\boldsymbol{h}}$ is nonsingular. It performs the *optimum weighting* of the decorrelated channel coefficients in the basis of the eigenvectors $\boldsymbol{U}$ of $\boldsymbol{R}_{\boldsymbol{h}}$ and can be considered a "soft" weighting compared with the RR approaches, which are constrained to be of rank R and simply "switch off" subspaces of the channel.

The RML estimator also requires the determination of the *optimum channel rank* R, as explained in Subsection III-C, which is not needed for the Wiener filter inherently determining the best bias-variance tradeoff.

Remarks on the Relation to Other RR Estimation Approaches: The ML criterion (25) with the RR model (24) is a special case of the criterion for RR estimation in [37] at high SNR with weighting matrix $\boldsymbol{R}_{\boldsymbol{\varepsilon}}$ given by the ML problem (25) for finite as well as for infinite F, which corresponds to time and ensemble averages for ergodic channel processes. It can also be recast as a weighted low-rank approximation problem [38] of the matrix $[\hat{\boldsymbol{h}}^1_{\mathrm{ML}}, \ldots, \hat{\boldsymbol{h}}^F_{\mathrm{ML}}]$ by the rank R matrix $[\hat{\boldsymbol{h}}^1_{\mathrm{RML}}, \ldots, \hat{\boldsymbol{h}}^F_{\mathrm{RML}}]$ with weighting matrix $\boldsymbol{I}_F \otimes \boldsymbol{R}_{\boldsymbol{\varepsilon}}^{-1}$. The advantage of the derivation in [38] based on the concept of Grassman manifolds is that a factorization as in (24) is not needed. This would also allow the use of the iterative algorithm presented in [38] to track the solution (another approach for tracking is [39] and references therein).

2) Reduced-Rank Correlator: The RR correlator is part of the generalized rake receiver architecture, which was originally motivated as an efficient extension to standard DS-CDMA receivers with a simple despreading and correlator (cf. Section V), [21], [22]. The *improvement* of the channel estimates from a *correlator by pre- or post-processing* the received pilot signal based on long-term channel properties is described in the sequel. Its impact on equalizer design is discussed in Section V.

The key idea is to provide an RR approximation $\boldsymbol{h}_R \in \mathbb{C}^R$ for the channel $\boldsymbol{h}$ from its estimate $\hat{\boldsymbol{h}}$:

$$\hat{\boldsymbol{h}}_R = \boldsymbol{M}_R^{\mathrm{H}}\hat{\boldsymbol{h}} \in \mathbb{C}^R \text{ and } \boldsymbol{M}_R \in \mathbb{C}^{ML_{\mathrm{c}} \times R}. \tag{32}$$

For an appropriate choice of $\boldsymbol{M}_R$ and rank R (see Section III-C) post-processing of $\hat{\boldsymbol{h}}$ improves the MSE of the estimate, e.g., if $\hat{\boldsymbol{h}}$ is obtained from a correlator (23). The vector $\boldsymbol{h}_R \in \mathbb{C}^R$ is an approximation of $\boldsymbol{h} \in \mathbb{C}^{ML_{\mathrm{c}}}$ in an R-dimensional subspace. This notation is used for the rake implementations in Section V-B, whereas $\boldsymbol{M}_R\boldsymbol{h}_R$ is the RR approximation in the original space.

With (12), (23), and (32), the RR estimate *based on the correlator* is

$$\hat{\boldsymbol{h}}_R = \frac{1}{N_{\mathrm{p}}P_{\mathrm{p}}}\boldsymbol{M}_R^{\mathrm{H}}\boldsymbol{S}_{\mathrm{p}}^{\mathrm{H}}\boldsymbol{y}_{\mathrm{p}}. \tag{33}$$

Equivalently to (post-) processing the estimate itself, the received signal $\boldsymbol{y}_{\mathrm{p}}$ can be reduced in rank followed by channel estimation with a correlator ($\boldsymbol{Y}_{\mathrm{P}} = [\boldsymbol{y}_{\mathrm{p}}[1], \ldots, \boldsymbol{y}_{\mathrm{p}}[N_{\mathrm{p}}]] \in \mathbb{C}^{M \times N_{\mathrm{P}}}$):

$$\hat{\boldsymbol{h}}_R = \underbrace{\boldsymbol{M}_R^{\mathrm{H}}\left(\boldsymbol{I}_{L_{\mathrm{c}}} \otimes \boldsymbol{Y}_{\mathrm{p}}\right)}_{\text{preprocessing}}\underbrace{\frac{1}{N_{\mathrm{p}}P_{\mathrm{p}}}\mathrm{vec}(\boldsymbol{S}_{\mathrm{p}}'^{,*})}_{\text{correlator}}. \tag{34}$$

Which version is preferable depends on implementation aspects [21]. The equivalence of (32) and (34) is shown in Appendix C.

We still must find criteria for designing the block FIR filter coefficients $\boldsymbol{M}_R[\ell]$

$$\boldsymbol{M}_R = [\boldsymbol{M}_R[0]^{\mathrm{T}}, \ldots, \boldsymbol{M}_R[L_{\mathrm{c}}-1]^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{ML_{\mathrm{c}} \times R}. \tag{35}$$

For a given rank R, we search for the best *low-rank approximation* of the channel. The spatial-temporal wireless channel in typical urban and rural scenarios consists of only a few main directions of arrival. Thus, the channel correlation matrix $\boldsymbol{R}_{\boldsymbol{h}} = \mathrm{E}[\boldsymbol{h}\boldsymbol{h}^{\mathrm{H}}]$ has a fast-decreasing eigenvalue spectrum $\{\lambda_m\}$. There exist subspaces of the channel containing only a small amount of information about the signal. Reduced-rank processing is designed based on second-order moments of the channel representing the slowly varying channel properties. The second-order statistics are exploited to minimize the loss in signal power, possibly reduce interference, and maximize the gain in channel estimation by reduction of the number of channel coefficients to be estimated.

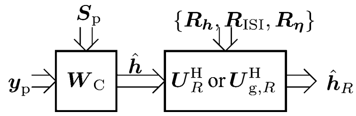


Fig. 5. Reduced-rank estimation with correlator for $M_R = U_R$ (36)—alternatively $M_R = U_{g,R}$ (38).

The *generalized Rayleigh quotient* [38], [40] is the appropriate optimization criterion for finding a good approximation of the channel

$$U_R = \underset{U'_R}{\operatorname{argmax}}\, \mathrm{trace}\left(R_h U'_R \left(U'^{\mathrm{H}}_R U'_R\right)^{-1} U'^{\mathrm{H}}_R\right)$$
$$\text{s.t.: } U'_R \in \mathbb{C}^{ML_c \times R}, \quad \mathcal{S}^R = \mathrm{span}(U'_R) \subseteq \mathbb{C}^{ML_c}. \quad (36)$$

It measures the channel contribution in an R-dimensional subspace $\mathcal{S}^R$. Its solutions are the R eigenvectors associated with largest eigenvalues of R_h

$$U_R = U[e_1, \ldots, e_R]. \quad (37)$$

They can be used as filter coefficients $M_R = U_R$ (see Fig. 5).

An alternative approach is to find the filter for best channel approximation and interference reduction maximizing the *signal-to-interference-plus-noise ratio* (SINR):

$$U_{g,R} = \underset{U'_R}{\operatorname{argmax}} \frac{\mathrm{trace}\left(U'^{\mathrm{H}}_R R_h U'_R\right)}{\mathrm{trace}\left(U'^{\mathrm{H}}_R (R_{\mathrm{ISI}} + R_\eta) U'_R\right)}$$
$$\text{s.t.: } U'_R \in \mathbb{C}^{ML_c \times R}, \quad \mathcal{S}^R = \mathrm{span}(U'_R) \subset \mathbb{C}^{ML_c}. \quad (38)$$

R_{ISI} is the channel correlation matrix of the ISI. It can also be formulated as the min–max problem [41]

$$\mathcal{S}^R_{\mathrm{g}} = \underset{\mathcal{S}^R \subset \mathbb{C}^{ML_c}}{\operatorname{argmax}} \min_{u \in \mathcal{S}^R} \frac{u^{\mathrm{H}} R_h u}{u^{\mathrm{H}} (R_{\mathrm{ISI}} + R_\eta) u}. \quad (39)$$

The solution is given as the subspace $\mathcal{S}^R_{\mathrm{g}} = \mathrm{span}(U_{g,R})$ spanned by the R generalized eigenvectors $U_{g,R} = U_{\mathrm{g}}[e_1, \ldots, e_R]$ associated with the largest generalized eigenvalues of the generalized eigenvalue problem

$$R_h U_{g,R} = (R_{\mathrm{ISI}} + R_\eta) U_{g,R} \Lambda_{\mathrm{g}}. \quad (40)$$

Λ_{g} is a diagonal matrix with decreasing SINR values on its diagonal. In multiuser scenarios, the multiaccess interference (MAI) is reduced by applying the correlation matrix of the MAI instead of R_{ISI} [21].

Application of the eigenvectors (or Karhunen–Loève decomposition) for efficient signal representation is well known and widely used in other fields such as feature extraction, image compression [42], or classification. The application of both approaches to signal processing is covered in [41, p. 209] with the concept of oriented energy and oriented SNRs.

Instead of considering signal/channel dimensions jointly in space and time based on the spatial-temporal correlations in R_h ($\Rightarrow M^{(\mathrm{ST})}_R = U_R$), other *less complex variations* of the same principle have been developed [21].

a) Space-Time (Eigen-) Rake: If the columns of H are uncorrelated,[7] then R_h is block diagonal with $R_{h[\ell]} = \mathrm{E}[h[\ell]h[\ell]^{\mathrm{H}}], \ell \in \{0, \ldots, L_c - 1\}$ on its diagonal. The conditions for uncorrelated columns in H will generally not be satisfied, but it can serve as a good approximation of R_h (see Section III-F). The advantage is a reduced complexity due to only L_c EVDs of $M \times M$ matrices.

b) Spatial Rake: Rank reduction is performed based on the spatial channel properties captured in $R^{(\mathrm{S})} = \mathrm{E}[HH^{\mathrm{H}}]$, which can be written as $R^{(\mathrm{S})} = \sum_{\ell=0}^{L_c-1} \mathrm{E}[h[\ell]h[\ell]^{\mathrm{H}}]$. The corresponding transformation is

$$M^{(\mathrm{S})}_R = I_{L_c} \otimes U^{(\mathrm{S})}_R \in \mathbb{C}^{ML_c \times RL_c} \quad (41)$$

with $U^{(\mathrm{S})}_R \in \mathbb{C}^{M \times R}$, defined as in (37), performing a pure (spatial) beamforming. It applies R beams to capture a significant portion of the signal.

c) Temporal Rake (Conventional Rake): To reduce temporal channel dimensions, we consider the channel correlations in the delay domain $R^{(\mathrm{T})} = \mathrm{E}[H^{\mathrm{H}}H]$. Generally, we obtain the transformation

$$M^{(T)}_R = U^{(\mathrm{T})}_R \otimes I_M \in \mathbb{C}^{ML_c \times MR} \quad (42)$$

with the eigenvectors $U^{(\mathrm{T})}_R \in \mathbb{C}^{L_c \times R}$. The channel is uncorrelated in delay/time for the same assumptions as before, i.e., $R^{(\mathrm{T})} = \Lambda^{(\mathrm{T})} = M\ \mathrm{diag}(\{P_\tau(\tau_1), \ldots, P_\tau(\tau_L)\})$, where the last equality is true if the power delay profile is the same for all antenna elements (i.e., a sufficiently small antenna spacing). In this case, the eigenvectors $U^{(T)} = I_{L_c}$ are canonical basis vectors, and the power delay profile is given as $P_\tau(\tau_\ell) = (1/M)\mathrm{E}[\|h[\tau_\ell/T_c]\|_2^2]$. For $U^{(\mathrm{T})}_R = [e_1, \ldots, e_R]$, the temporal rake simply selects the strongest R delayed signals, i.e., places R temporal fingers on the most significant temporal diversity paths.

C. Optimum Rank

The *MSE* of the channel estimates for RR approaches *varies significantly with the rank* R. Usually, the channel correlation matrix R_h is not of low rank algebraically, but the dimensions associated with small eigenvalues λ_m may be neglected based on an optimum bias variance tradeoff [5], [11]. Traditionally, information-theoretic criteria are used for estimating the rank, e.g., the Akaike information criterion (AIC) or minimum description length (MDL) approach [43], but tackle a problem different from this one: They estimate the rank of algebraically rank-deficient channels based on an estimate of their correlation matrix. Here, we assume that the correlation matrix is known perfectly and of full rank. Still, these criteria may be good enough in some cases [44].

Determination of the rank can be based on the MSE, which can be computed explicitly and describes the bias-variance tradeoff (see Section IV). The *optimum rank* for the RML (equivalently for the RR correlator) is defined as

$$R_{\mathrm{opt}} = \underset{R}{\operatorname{argmin}}\, \mathrm{MSE}_{\mathrm{RML}}(R) = \underset{R}{\operatorname{argmin}}\, \mathrm{E}\left[\|\hat{h}_{\mathrm{RML}} - h\|_2^2\right]. \quad (43)$$

[7]Sufficient assumptions are uncorrelated scattering and channel delays on the sampling grid of period T_c.

TABLE II
OPTIMUM RANK w.r.t. MSE FOR RR CHANNEL ESTIMATORS $N_{\rm p} \in \{10, 50\}$

SNR	-10 dB		0 dB		10 dB		20 dB	
$N_{\rm p}$	10	50	10	50	10	50	10	50
RML	5	5	7	10	11	16	16	19
RR Corr. (EVD)	5	6	7	8	11	15	14	17

Most of the terms in the MSE of the RML have to be computed for the estimator anyway, i.e., additional computations are matrix multiplications and additions based on a full EVD. An alternative definition based on a more relevant criterion for a communication link, e.g., BEP, would be advantageous but is difficult to compute in general.

The optimum rank also depends largely on the definition of the MSE: either as in (43) or alternatively defined w.r.t. the estimate scaled by a scalar Wiener filter. The second definition seems to be preferable since it ensures a fair comparison. The optimum rank for different SNR in the scenario from Section III-D is summarized in Table II for this definition. In general, a small rank is selected for low SNR, as the variance needs to be decreased at the price of an increased bias, and for high SNR, full-rank estimation is optimum. Furthermore, the ISI is decreased in case of the RR correlator. The optimum rank increases with the number of pilot symbols and the SNR. The correlator based on the EVD chooses a smaller rank to further suppress interference.

The simulation results for the BER in Section VI show that the *optimum rank* is directly *depending on the type of equalizer* used and the sensitivity of the equalizer to errors in the channel estimates. Thus, optimum rank determination must always be performed for a specific equalizer and receiver architecture in frequency-selective channels.

D. MSE Performance

To discuss channel estimation performance independently from the equalizer, we use the MSE of the channel estimates. The estimates are scaled by a scalar Wiener filter to allow for a comparison of the different approaches.

The channel correlation matrix is determined for an *urban environment* with channels of length $L = 5$ and exponential power delay profile with delay spread 1 μs. Waves on one delay tap arrive from a uniformly distributed angle of arrival with Laplace distributed angular spread (azimuth spread: 10°) [45]. The receiver is equipped with $M = 8$ antennas in a uniform linear array. For the MSE computation, we assume uncorrelated scattering and channel delays on the sampling grid of the receiver (chip period is $T_{\rm c} = 1/3.86 \cdot 10^{-6}$ s, as in UMTS), i.e., the correlation matrix $\boldsymbol{R_h}$ is block diagonal. The noise is white in space and time. Pilot symbols (including the prefix) are chosen from a quaternary (QPSK) pseudo random sequence.

As known from its optimization criterion, the WF achieves the minimum MSE. The gains of the WF are twofold: On the one hand, it does not enhance the noise as the ML approach, which is characteristic for zero forcing, and on the other hand, it exploits correlations of the channel coefficients (long-term properties).

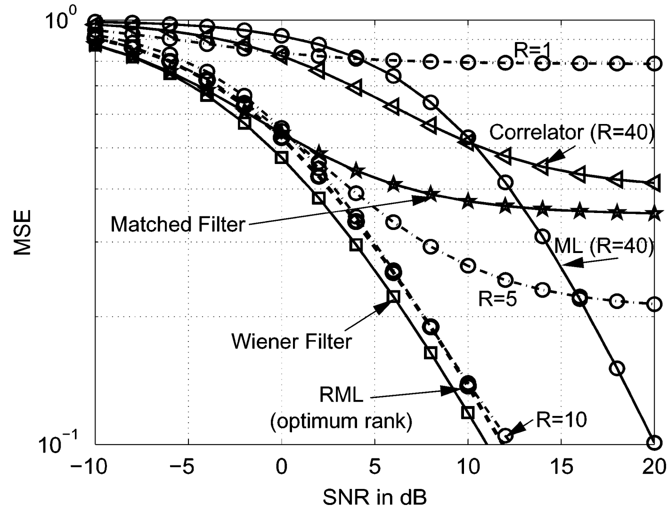


Fig. 6. MSE of channel estimators for a frequency selective channel of (full) rank equal to 40 for $R \in \{1, 5, 10, 40\}$ and $N_{\rm p} = 10$. All estimates are scaled by a scalar Wiener filter.

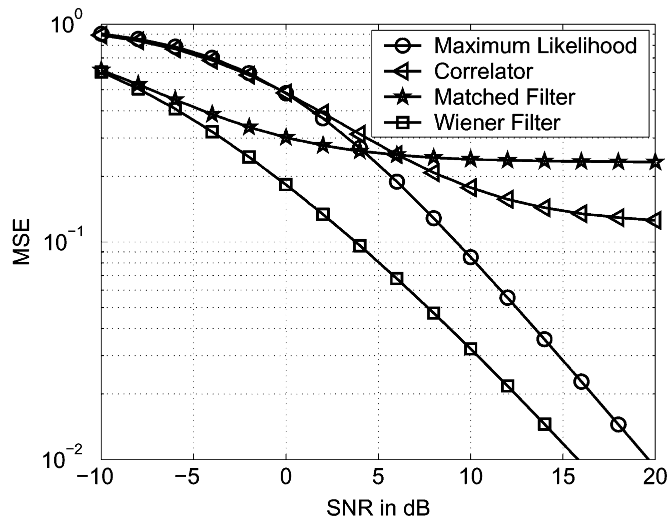


Fig. 7. MSE of full rank estimators for a frequency-selected channel ($N_{\rm p} = 50$).

For example, in a fully correlated channel with only one direction of arrival and one tap $(\mathrm{Rank}(\boldsymbol{R_h}) = 1)$, it uses this knowledge and inherently estimates just one coefficient knowing the eigenvector of the associated subspace. For $N_{\rm p} = 10$ pilot symbols, the loss of the ML method is significant, and it is clearly outperformed by the MF for low SNR, which saturates at high level due to its bias (see Fig. 6). The correlator is worse than the MF as it does not exploit channel correlations. It outperforms the ML for low SNR and $N_{\rm p} = 10$ (noise enhancement of the ML) but not for $N_{\rm p} = 50$ (see Fig. 7). For high SNR, the WF converges to the ML, as shown in Fig. 7. Moreover, MF and the correlator have a crossover point as the bias of the correlator is generally smaller at high SNR.

The MSE of the RML is decreased considerably for $R \in \{5, 10\}$ compared to the full rank ML $(R = 40)$ (Fig. 6). For a given rank the MSE saturates with the bias of the RML estimate. The RML with optimum rank, i.e., the rank is optimized w.r.t. the MSE, merges in the full rank ML estimator for high SNR.

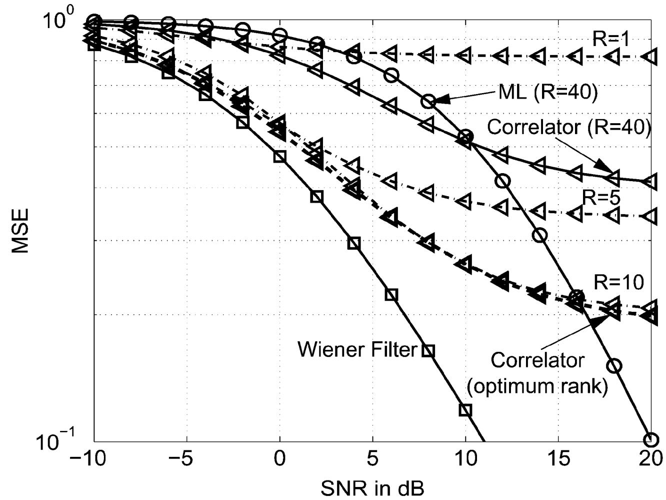


Fig. 8. MSE of RR correlator for a frequency selective channel of (full) rank equal to 40 for $R \in \{1, 5, 10, 40\}$ and $N_\mathrm{p} = 10$.

TABLE III
COMPUTATIONAL COMPLEXITY OF CHANNEL ESTIMATORS

Channel Estimator	Matrix inversion $O(M^3 L_\mathrm{c}^3)$	EVD $O(M^3 L_\mathrm{c}^3)$	$\boldsymbol{R}_{\boldsymbol{\eta}}^{-1}$
Wiener Filter	1×	—	1×
Max. Likelihood	1×	—	1×
Reduced rank ML	1×	1×	1×
Matched Filter	—	—	1×
Correlator	—	—	—
Reduced Rank Correlator	—	1×	—

Note that any rank reduction for the Wiener or matched filter would increase their MSE.

Rank reduction based on the EVD (36) for the correlator improves the MSE over the whole range of SNR ($R = 10$), as it decreases errors due to intersymbol interference, which is rather large for $N_\mathrm{p} = 10$, as well as the variance due to the noise (Fig. 8). Rank $R = 1$ would not be sufficient as a major fraction of the channel is neglected.

E. Complexity

Evaluating and comparing the computational complexity for the design of these channel estimators, we assume that a *constant rank* R is assigned to the RR versions. For the RR versions, we also require computation of all eigenvectors as their knowledge is necessary for estimating the rank. If only few eigenvectors are needed, less-complex numerical methods for the EVD, e.g., power iterations, exist [46].

Table III lists the order of the most complex operations needed to compute the channel estimators: The powerful ML and WF are of comparable complexity. The *cheapest estimators* are the correlator and matched filter. Incorporating second-order statistics into the MF results in a quadratic order of complexity, which is also the complexity order of the whole MF in case of white noise.

As the RML requires additional computations on the same order as the full rank ML and does not achieve a better performance than the WF (see Fig. 6), this effort is wasted. In case of the RR correlator, which performs better than the MF for an appropriate choice of R, the additional computations for the EVD result in a complexity similar to the WF. Here, we have to keep in mind that the *RR correlator* is used together with the generalized rake in the receiver, which relies on the same rank reduction (see Section V) and reduces the complexity of a linear MMSE equalizer. Thus, the overall computational requirements depend on the rate of change of $\boldsymbol{R}_{\boldsymbol{h}}$ and are decreased for rare updates of $\boldsymbol{R}_{\boldsymbol{h}}$.

The computations can be reduced further approximating $\boldsymbol{R}_{\boldsymbol{h}}$ by a block-diagonal matrix (Space-time eigenrake; Section III-B2), which reduces the cost to L_c EVDs of order $O(M^3)$ but also decreases performance due to a larger amount of neglected signal power for fixed R. The inversion of $\boldsymbol{R}_{\boldsymbol{\eta}}$ can be implemented with N_p inversions of order of $O(M^3)$ for temporally white noise.

The *final choice* of the *channel estimation method* depends on the sensitivity of the the equalizer and the point of operation: For low SNR, i.e., bad channel estimates, the matched filter is sufficient, or an RR correlator might be used together with the generalized rake; for higher SNRs, a Wiener filter is superior, and it should be worth spending the additional computational power.

F. Estimation of Second-Order Channel Statistics $\boldsymbol{R}_{\boldsymbol{h}}$

Up to now, the channel correlation matrix $\boldsymbol{R}_{\boldsymbol{h}}$ was assumed to be known perfectly. In practice, it has to be estimated, e.g., by temporal averaging of F channel estimates $\hat{\boldsymbol{h}}_\bullet$:

$$\hat{\boldsymbol{R}}_{\hat{\boldsymbol{h}}_\bullet}^{F} = \frac{1}{F}\sum_{f=1}^{F} \hat{\boldsymbol{h}}_\bullet^{f} \hat{\boldsymbol{h}}_\bullet^{f,\mathrm{H}}. \tag{44}$$

For nonstationary channels, the estimate $\hat{\boldsymbol{R}}_{\hat{\boldsymbol{h}}_\bullet}^{f}$ based on f channel estimates can be performed recursively by

$$\hat{\boldsymbol{R}}_{\hat{\boldsymbol{h}}_\bullet}^{f} = \rho\, \hat{\boldsymbol{R}}_{\hat{\boldsymbol{h}}_\bullet}^{f-1} + (1-\rho)\, \hat{\boldsymbol{h}}_\bullet^{f} \hat{\boldsymbol{h}}_\bullet^{f,\mathrm{H}} \tag{45}$$

with an exponential forgetting factor ρ (see [41p. 277]), which is set according to the degree of nonstationarity of the channel.

For the *RML estimator* (26), the optimum estimate of $\boldsymbol{R}_{\boldsymbol{h}}$ is obtained substituting the full-rank ML estimate $\hat{\boldsymbol{h}}_\mathrm{ML}$ from (20) for $\hat{\boldsymbol{h}}_\bullet^{f}$ [cf. (67)] [5]. For the *Wiener filter* (18), one may also use the ML estimate $\hat{\boldsymbol{h}}_\mathrm{ML}$ to estimate $\boldsymbol{R}_{\boldsymbol{h}}$. It can be obtained before applying $\boldsymbol{P}_\mathrm{WF}$ in (30), resulting in an increased complexity due to the factorization of $\boldsymbol{W}_\mathrm{WF}$. Thus, it loses its complexity advantage compared with the RML (see Table III) for fixed rank.

For the *matched filter* (22) or the *RR correlator* (33), only the channel estimate $\hat{\boldsymbol{h}}_\mathrm{c}$ from the correlator (23), which is the first stage in both estimators, is available without increasing complexity. Using this estimate in (44) leads to a significant bias due to the remaining interference of the correlator estimate. In [47], a technique to compensate the interference correlation matrix $\boldsymbol{R}_\mathrm{I}$ is proposed based on the system of linear equations

$$\begin{aligned}\boldsymbol{R}_{\hat{\boldsymbol{h}}_\mathrm{c}[\ell]} &= \mathrm{E}[\hat{\boldsymbol{h}}_\mathrm{c}[\ell]\hat{\boldsymbol{h}}_\mathrm{c}[\ell]^\mathrm{H}] = \boldsymbol{R}_{\boldsymbol{h}[\ell]} + \frac{1}{N_\mathrm{p}P_\mathrm{p}}\boldsymbol{R}_\mathrm{I}\\ \boldsymbol{R}_{\boldsymbol{y}_\mathrm{p}[n]} &= \mathrm{E}[\boldsymbol{y}_\mathrm{p}[n]\boldsymbol{y}_\mathrm{p}[n]^\mathrm{H}] = P_\mathrm{p}\boldsymbol{R}_{\boldsymbol{h}[\ell]} + \boldsymbol{R}_\mathrm{I}\end{aligned} \tag{46}$$

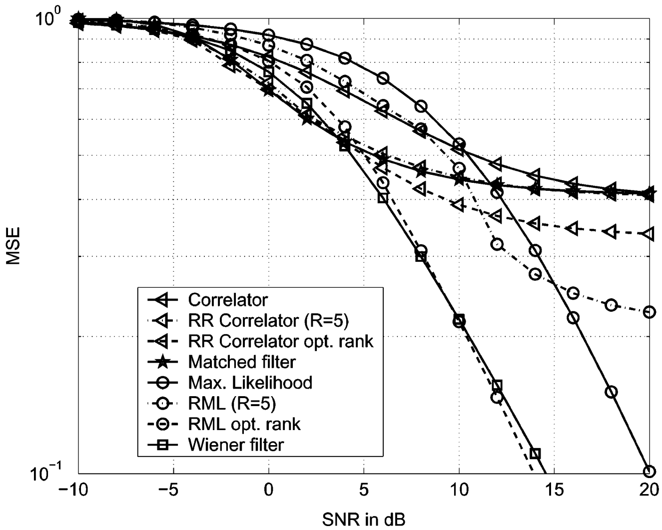


Fig. 9. MSE of channel estimators for a frequency-selective channel of (full) rank equal to 40 for $N_\mathrm{p} = 10$, where second-order channel statistics $\boldsymbol{R_h}$ were estimated from $F = 50$ independent channel estimates, as described in Section III-F and are assumed to have block diagonal structure.

which can be used to compute $\boldsymbol{R}_{\boldsymbol{h}[\ell]}, \forall \ell \in \{1, \ldots, L_\mathrm{c}\}$

$$\boldsymbol{R}_{\boldsymbol{h}[\ell]} = \frac{1}{P_\mathrm{p}(N_\mathrm{p}-1)} \left(N_\mathrm{p} P_\mathrm{p} \boldsymbol{R}_{\hat{\boldsymbol{h}}_\mathrm{c}[\ell]} - \boldsymbol{R}_{\boldsymbol{y}_\mathrm{p}[n]} \right) \tag{47}$$

yielding a block diagonal approximation of $\boldsymbol{R_h}$.

Channel measurements in [32] showed that approximately 120 independent channel realizations are available for estimation in a suburban environment. Thus, the MSE based on $\hat{\boldsymbol{R}}^F_{\hat{\boldsymbol{h}}_\bullet}$ with $F = 50$ and the scenario as in Section III-D ($\boldsymbol{R_h}$ block diagonal) is shown in Fig. 9 for given SNR. It can be seen that the conclusions from Section III-D still hold with the restriction that the WF is more sensitive to errors than the RML. The matched filter and RR correlator outperform the latter approaches due to their low sensitivity to errors in $\hat{\boldsymbol{R}}^F_{\hat{\boldsymbol{h}}_\bullet}$. This will no longer be true if the impact of channel estimation on LMMSE equalization and the BER are considered, as in Section VI.

IV. Analysis for Equivalent Flat-Fading Channels

An analysis of the presented channel estimation methods in Section III for *equivalent flat fading channels* provides sufficient insights to understand the differences and performance tradeoffs involved in parameterization of the estimators and selection of the optimal method.

We assume that all ML spatial-temporal diversity branches can be separated perfectly, e.g., due to perfect autocorrelation properties of the spreading sequence. With these assumptions, the resulting space-time channel can be described by the equivalent flat (nondispersive) Rayleigh fading channel model below (e.g., [48]). The receiver performs *maximum ratio combining* (MRC) of the branches using the estimated channel coefficients (see Fig. 10) $\hat{\boldsymbol{h}}$, which maximizes the SNR. BPSK symbols $s_\mathrm{d}[i]$ are transmitted, i.e.,

$$\boldsymbol{y}_\mathrm{d}[i] = \boldsymbol{h} s_\mathrm{d}[i] + \boldsymbol{\eta}_\mathrm{d}[i] \in \mathbb{C}^{ML} \tag{48}$$

$$\hat{s}_\mathrm{d}[i] = \mathrm{Real}(\hat{\boldsymbol{h}}^\mathrm{H} \boldsymbol{y}_\mathrm{d}[i]). \tag{49}$$

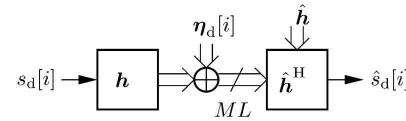


Fig. 10. Equivalent flat-fading channel with ML diversity paths and maximum ratio combining (MRC) receiver.

For channel estimation N_p pilot symbols are arranged in $\boldsymbol{S}_\mathrm{p} = \boldsymbol{s}_\mathrm{p} \otimes \boldsymbol{I}_{ML}$ with $\boldsymbol{s}_\mathrm{p} = [s_\mathrm{p}[1], \ldots, s_\mathrm{p}[N_\mathrm{p}]]^\mathrm{T}$, and we assume white noise $\boldsymbol{R_\eta} = \sigma_\mathrm{n}^2 \boldsymbol{I}_{MN_\mathrm{p}}$.

For flat-fading channels, the correlator and ML estimator are identical as no ISI is present. Consequently, RML (28) and the RR correlator (33) are equivalent.

From Table IV, we see that for flat channels, all estimators can be written as an ML estimator followed by an appropriate post-processing stage weighting the channel subspaces: The Wiener filter performs an optimum weighting using the eigenvectors $\boldsymbol{U}$ of $\boldsymbol{R_h}$ (30). Clearly, a truncation of signal subspaces as for the RML estimator is suboptimum w.r.t. MSE.

The MSE of all channel estimates can be written as

$$\begin{aligned} \mathrm{MSE}_\bullet &= \mathrm{E}\left[\| \hat{\boldsymbol{h}}_\bullet - \boldsymbol{h} \|_2^2 \right] \\ &= \mathrm{E}\left[\| \boldsymbol{W}_\bullet \boldsymbol{S}_\mathrm{p} \boldsymbol{h} - \boldsymbol{h} + \boldsymbol{W}_\bullet \boldsymbol{\eta}_\mathrm{p} \|_2^2 \right] \\ &= \mathrm{trace}(\boldsymbol{W}_\bullet \boldsymbol{R_\eta} \boldsymbol{W}_\bullet^\mathrm{H}) \\ &\quad + \mathrm{trace}\left(\boldsymbol{R_h} + \boldsymbol{W}_\bullet \boldsymbol{S}_\mathrm{p} \boldsymbol{R_h} \boldsymbol{S}_\mathrm{p}^\mathrm{H} \boldsymbol{W}_\bullet^\mathrm{H} \right. \\ &\quad \left. -\ 2\,\mathrm{Real}(\boldsymbol{W}_\bullet \boldsymbol{S}_\mathrm{p} \boldsymbol{R_h}) \right) \end{aligned} \tag{50}$$

since $\boldsymbol{h}$ and $\boldsymbol{\eta}_\mathrm{p}$ are uncorrelated. The *bias* [second term in (50)] represents a systematic estimation error independent of the noise $\boldsymbol{\eta}_\mathrm{p}$, whereas the *variance* (first term) describes the error due to the noise. These two contributions to the MSE allow a more detailed analysis of the estimators' behavior. The MSE of the considered channel estimators split into bias and variance contribution is given in Table IV.

In case of few pilot symbols and high noise, the bias dominates the performance of the WF as the estimator does not "trust" the received pilot signal. It relies more on the *a priori* information, i.e., the second-order moment of the channel $\boldsymbol{h}$. For large N_p and high SNR, the WF "trusts" the estimates it obtains from the ML estimator, i.e., no weighting of the subspace is needed. Remember, that for low rank channels, i.e., $\mathrm{Rank}(\boldsymbol{R_h}) < ML$, the Wiener method inherently drops the channel dimensions containing noise only.

The MSE of the (unbiased) ML estimate is equivalent to the estimation variance (see Table IV) linearly increasing with the number of coefficients ML to be estimated as known from classical estimation theory [34]. It is unbounded in low SNR situations: The MSE exceeds the power of the channel $\mathrm{trace}(\boldsymbol{R_h})$, which is avoided by the Wiener approach due to the *a priori* knowledge of the channel statistics.

To compute its MSE, the matched filter is scaled[8] by $\alpha = \sigma_\mathrm{n}^2/\lambda_1$: Generally, $\sum_{m=1}^{ML} (\lambda_m/\lambda_1)^2 \leq ML$ holds, i.e., the variance of the matched filter is smaller than for the ML estimator. The more correlated the channel, the larger the variance gain, but a bias independent of the noise variance remains for high

[8]Scaling with a scalar Wiener filter would be less heuristic and show similar results but would be less obvious to interpret.

TABLE IV
FULL AND RR CHANNEL ESTIMATORS AND THEIR BIAS/VARIANCE FOR EQUIVALENT FLAT CHANNELS

	Channel Estimator	Bias	Variance
ML/Correlator	$\boldsymbol{W}_{\mathrm{ML}} = \boldsymbol{W}_{\mathrm{C}} = \frac{1}{P_{\mathrm{p}} N_{\mathrm{p}}} \boldsymbol{s}_{\mathrm{p}}^{\mathrm{H}} \otimes \boldsymbol{I}_{ML}$	0	$ML \, \frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{p}} N_{\mathrm{p}}}$
MF	$\boldsymbol{W}_{\mathrm{MF}} = \frac{1}{\lambda_1} \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^{\mathrm{H}} \boldsymbol{W}_{\mathrm{ML}}$	$\sum_{m=1}^{ML} \lambda_m \left(\frac{\lambda_m}{\lambda_1} - 1 \right)^2$	$\sum_{m=1}^{ML} \left(\frac{\lambda_m}{\lambda_1} \right)^2 \frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{p}} N_{\mathrm{p}}}$
WF	$\boldsymbol{W}_{\mathrm{WF}} = \boldsymbol{U} \underbrace{\left(\boldsymbol{I}_{ML} + \frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{p}} N_{\mathrm{p}}} \boldsymbol{\Lambda}^{-1} \right)^{-1}}_{\text{Optimal weighting of subspaces}} \boldsymbol{U}^{\mathrm{H}} \boldsymbol{W}_{\mathrm{ML}}$	$\sum_{m=1}^{ML} \lambda_m \left(\frac{1}{\frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{p}} \lambda_m N_{\mathrm{p}}} + 1} - 1 \right)^2$	$\sum_{m=1}^{ML} \frac{1}{\left(\frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{p}} \lambda_m N_{\mathrm{p}}} + 1 \right)^2} \frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{p}} N_{\mathrm{p}}}$
RR Correlator	$\boldsymbol{W}_{\mathrm{RRC}} = \boldsymbol{U}_R^{\mathrm{H}} \boldsymbol{W}_{\mathrm{ML}}$	$\sum_{m=R+1}^{ML} \lambda_m$	$R \, \frac{\sigma_{\mathrm{n}}^2}{P_{\mathrm{p}} N_{\mathrm{p}}}$
RML	$\boldsymbol{W}_{\mathrm{RML}} = \boldsymbol{U}_R \boldsymbol{U}_R^{\mathrm{H}} \boldsymbol{W}_{\mathrm{ML}}$		

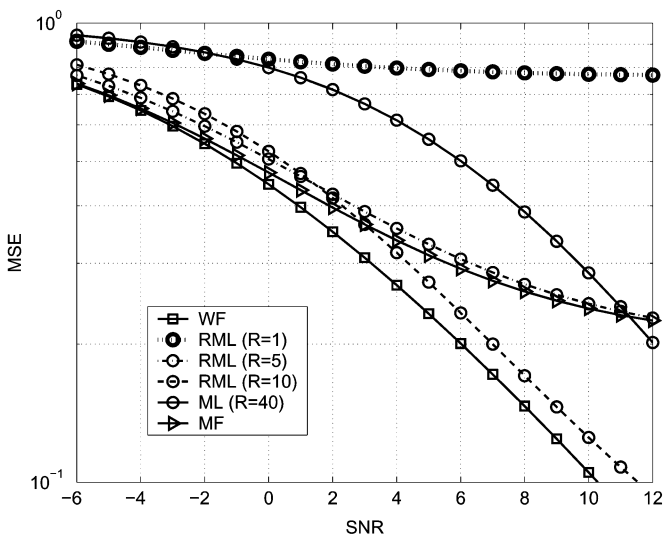


Fig. 11. MSE of channel estimators for equivalent flat-fading channel ($\mathrm{ML} = 40, N_{\mathrm{p}} = 10$). All estimates are scaled with a scalar Wiener filter.

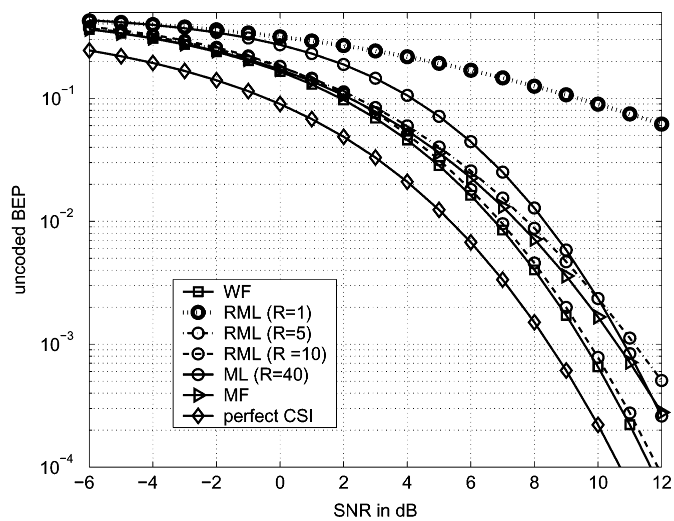


Fig. 12. Bit error probability of MRC with different channel estimators for equivalent flat-fading channel ($\mathrm{ML} = 40, N_{\mathrm{p}} = 10$) and BPSK modulation.

SNR (see Table IV). Thus, a crossover point in performance exists between the MF and ML estimator.

In addition, the RML and correlator introduce a bias to decrease estimation variance finding the best *trade-off between bias and variance*. Their MSE is defined as

$$\mathrm{MSE}_{\mathrm{RML}} = \mathrm{MSE}_{\mathrm{RRC}} = \mathrm{E}\left[\|\boldsymbol{U}_R \hat{\boldsymbol{h}}_{\mathrm{R}} - \boldsymbol{h}\|_2^2 \right]. \tag{51}$$

The bias is proportional to the received power in the neglected signal/channel dimensions. To find the best performance, it is necessary to optimize R, which is very cheap for flat-fading channels, but more involved in the general case (cf. Section III-C)

In the sequel, the same scenario is considered as in Section III-D, but the spatial-temporal paths are considered as noninterfering diversity branches (48).

As in Section III-D the channel estimates are scaled by a scalar Wiener filter preserving their structure but allowing for a comparison w.r.t. the MSE.[9] The MSE for the scaled estimates is shown in Fig. 11 for $N_{\mathrm{p}} = 10$. As in the frequency-selective case, the MF and RML/ML have a crossover point since the MF saturates at its bias (see Table IV). This point approximately occurs at the same SNR w.r.t. the BEP (see Fig. 12). All results in Fig. 12 are obtained by evaluating the analytical BEP expression, which is given in Appendix D with a brief derivation. For a BEP of 10^{-1}, the ML estimator loses about 4.5 dB due to errors in the channel estimates. For the Wiener filter, the loss is about 2 dB. The MF is as close as 0.3 dB to the WF. Above 2.5 dB, the MF is outperformed by the RML with $R = 10$, which is considerably more complex due to the EVD. At low SNR and for small N_{p}, the MF should be the preferred channel estimator.

[9]For flat fading channels and BPSK, the real-valued scaling factor does not affect the BEP.

V. APPLICATION TO EQUALIZATION

Considering the *classical separate design* of channel estimation and equalization, we briefly introduce two equalization concepts to illustrate the consequences of different channel estimators on their design and performance (see Section VI). The generalized rake in particular can be well understood from the concept of the RR correlator (see Section III-B2).

1) Joint Equalization and Despreading: For short spreading codes, the data symbols $s_{\mathrm{d}}[i]$ can be directly estimated from the received chip sequence (joint equalization and despreading; see Fig. 13)

$$\hat{s}_{\mathrm{d}}[i] = \boldsymbol{g}^{\mathrm{H}} \boldsymbol{y}_{\mathrm{d}}^{(N)}[n], \; n = Qi \tag{52}$$

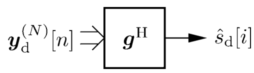


Fig. 13. Joint equalization and despreading.

where all contributions from symbol $s_{\rm d}[i]$ in $\boldsymbol{y}_{\rm d}[n]$ are captured for $N = Q + L_{\rm c} - 1$. An LMMSE FIR equalizer [18] $\boldsymbol{g} \in \mathbb{C}^{MN}$ as solution of the Wiener–Hopf equation

$$\boldsymbol{R}_{\boldsymbol{y}}\, \boldsymbol{g} = \boldsymbol{r}_{\boldsymbol{y} s_{\rm d}[i]^*} \tag{53}$$

with covariance matrix of $\boldsymbol{R}_{\boldsymbol{y}} = \mathrm{E}[\boldsymbol{y}_{\rm d}^{(N)}[n]\boldsymbol{y}_{\rm d}^{(N)}[n]^{\rm H}]$ and cross-correlation vector $\boldsymbol{r}_{\boldsymbol{y} s_{\rm d}[i]^*} = \mathrm{E}[\boldsymbol{y}_{\rm d}^{(N)}[n] s_{\rm d}[i]^*]$ is used. It has MN degrees of freedom, yielding a better performance at the expense of a large complexity, which is $O(M^3 N^3)$, assuming that the matrix structure is not exploited, compared with chip-level equalization in the next subsection. It also requires knowledge about the spreading sequence $\boldsymbol{c}$.

2) Generalized Rake Receiver: For long spreading sequences, the approach in (52) is far too complex, as the spreading sequence changes from symbol to symbol. LMMSE *equalization at the chip level* [49] estimates the transmitted chip sequence $x_{\rm d}[n]$, i.e., is independent of the spreading code. Its design is based on (15) with $N = L_{\rm c}$:

$$\boldsymbol{y}_{\rm d}^{(L_{\rm c})}[n] = \boldsymbol{\mathcal{H}}^{(L_{\rm c})} \boldsymbol{x}_{\rm d}^{(L_{\rm c})}[n] + \boldsymbol{\eta}_{\rm d}^{(L_{\rm c})}[n] \in \mathbb{C}^{ML_{\rm c}}. \tag{54}$$

The *generalized rake receiver* architecture was developed to further reduce the computational complexity of the chip-level equalizer at the price of a small loss in signal power employing the notion of low-rank approximation from Section III-B2 [11], [21]–[23]:

$$\boldsymbol{z}_{\rm d}[n] = \boldsymbol{M}_R^{\rm H} \boldsymbol{y}_{\rm d}^{(L_{\rm c})}[n] \in \mathbb{C}^R. \tag{55}$$

The LMMSE filter $\boldsymbol{g} = \boldsymbol{R}_{\boldsymbol{z}}^{-1} \boldsymbol{r}_{\boldsymbol{z} x_{\rm d}[n+L_{\rm c}]^*} \in \mathbb{C}^R$ estimates $\hat{x}_{\rm d}[n] = \boldsymbol{g}^{\rm H} \boldsymbol{z}_{\rm d}[n]$ with parameters

$$\boldsymbol{R}_{\boldsymbol{z}} = \boldsymbol{M}_R^{\rm H} \boldsymbol{\mathcal{H}}^{(L_{\rm c})} \boldsymbol{\mathcal{H}}^{(L_{\rm c}),{\rm H}} \boldsymbol{M}_R + \boldsymbol{M}_R^{\rm H} \boldsymbol{R}_{\boldsymbol{\eta}} \boldsymbol{M}_R \tag{56}$$

$$\boldsymbol{r}_{\boldsymbol{z} x_{\rm d}[n+L_{\rm c}]^*} = \boldsymbol{M}_R^{\rm H} \boldsymbol{\mathcal{H}}^{(L_{\rm c})} \boldsymbol{e}_{L_{\rm c}} = \boldsymbol{M}_R^{\rm H} \boldsymbol{h} = \boldsymbol{h}_R \in \mathbb{C}^R \tag{57}$$

where the correlation matrix $\boldsymbol{R}_{\boldsymbol{z}} = \mathrm{E}[\boldsymbol{z}_{\rm d}[n]\boldsymbol{z}_{\rm d}[n]^{\rm H}]$ can be directly estimated as a time average from the received data or indirectly estimating the coefficients in $\boldsymbol{\mathcal{H}}^{(L_{\rm c})}$ with a correlator before despreading. The cross-correlation vector $\boldsymbol{r}_{\boldsymbol{z} x_{\rm d}[n+L_{\rm c}]^*} = \mathrm{E}[\boldsymbol{z}_{\rm d}[n] x_{\rm d}[n+L_{\rm c}]^*] = \boldsymbol{h}_R$ can be obtained as an RR (32) estimate from the RR correlator (see Section III-B2). Alternatively, $\boldsymbol{g}$ is often chosen as an MRC with $\boldsymbol{g} = \boldsymbol{h}_R$.

Considering the received signal (15) for $N = Q + L_{\rm c} - 1$, as before, its RR version is

$$\boldsymbol{z}_{{\rm d},R}[n] = \boldsymbol{T}_R \boldsymbol{y}_{\rm d}^{(N)}[n] \in \mathbb{C}^{RQ}, \quad 1 \le R \le ML_{\rm c} \tag{58}$$

using the block FIR filter (block-Toeplitz)

$$\boldsymbol{T}_R = \sum_{\ell=0}^{L_{\rm c}-1} \boldsymbol{J}_{(\ell M, Q, L_{\rm c}-1)} \otimes \boldsymbol{M}_R[\ell]^{\rm H} \in \mathbb{C}^{RQ \times MN}. \tag{59}$$

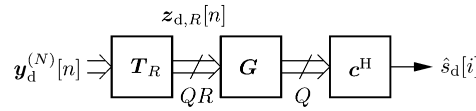


Fig. 14. Generalized rake receiver: Chip-level equalization.

The block coefficient $\boldsymbol{M}_R[\ell]$ in $\boldsymbol{M}_R$ (35) maximizing signal energy is the ℓth $M \times R$ submatrix of $\boldsymbol{U}_R$ for $\ell \in \{0, \ldots, L_{\rm c} - 1\}$. Possible implementations are (see Fig. 14)

$$\hat{s}_{\rm d}[i] = \boldsymbol{c}^{\rm H} \boldsymbol{G} \boldsymbol{T}_R \boldsymbol{y}_{\rm d}^{(N)}[n] = \boldsymbol{g}^{\rm H} \boldsymbol{D} \boldsymbol{T}_R \boldsymbol{y}_{\rm d}^{(N)}[n] \tag{60}$$

with $\boldsymbol{G} = \boldsymbol{I}_Q \otimes \boldsymbol{g}^{\rm H}$ and despreading matrix $\boldsymbol{D} = \boldsymbol{c}^{\rm H} \otimes \boldsymbol{I}_R \in \mathbb{C}^{R \times RQ}$. For example, with $\boldsymbol{M}_R[\ell] = \boldsymbol{U}_R^{({\rm T})} \otimes \boldsymbol{I}_M$ (42), it is the conventional temporal rake receiver with LMMSE combining of the rake fingers.

The LMMSE equalizer together with a generalized rake has a complexity of order $O(R^3)$ compared with the full-rank chip equalizer $O(M^3 L_{\rm c}^3)$, which results in a considerable gain for typical suburban or rural environments.

Besides using the RR correlator for estimating $\boldsymbol{h}_R$ in (57), the conclusions drawn for it w.r.t. channel estimation performance also apply to the generalized rake, which uses the same transformation $\boldsymbol{M}_R$. Thus, application of the generalized rake as CDMA receiver has the following *advantages*.

1) It presents an add-on to existing CDMA receiver architectures, which use a simple correlator for channel estimation, as an extension to space-time processing.
2) It improves the quality of the channel estimates by rank reduction.
3) It reduces complexity of the equalizer design by rank reduction. (On the complexity of the EVD, see Section III-E.)
4) It exploits the slowly varying spatial and temporal channel properties.

On the other hand, it reduces the degrees of freedom available for equalization to R. Thus, the performance gain is largest at low SNR, where the channel estimate is improved, and the BER is dominated by the noise and not the ISI.

VI. Simulation

Simulation Scenario: The same environment is chosen as in Section III-D with $M = 8$, but now, the $L = 5$ channel delays are uniformly distributed in $[0, 4T_{\rm c}]$ and not quantized by the sampling period. A root-raised cosine impulse shape with roll-off factor 0.2 is used and truncated for comparison such that $L_{\rm c} = 11$ [$\beta = 3$, (9)]. Furthermore, we assume a Rayleigh fading channel and that 300 independent random realizations of the environment are considered for Monte Carlo simulation of the uncoded BER. In each realization, a block of 500 QPSK symbols is transmitted. The number of pilot symbols is $N_{\rm p} = 20$, which are chosen to illustrate the difference among the channel estimation approaches. We consider a single user with spreading factor $Q = 4$ and (short) random spreading sequence of length 4. All MAI is modeled as noise $\boldsymbol{\eta}_{\rm d}[n]$ for simplicity as we restrict our evaluation in this context

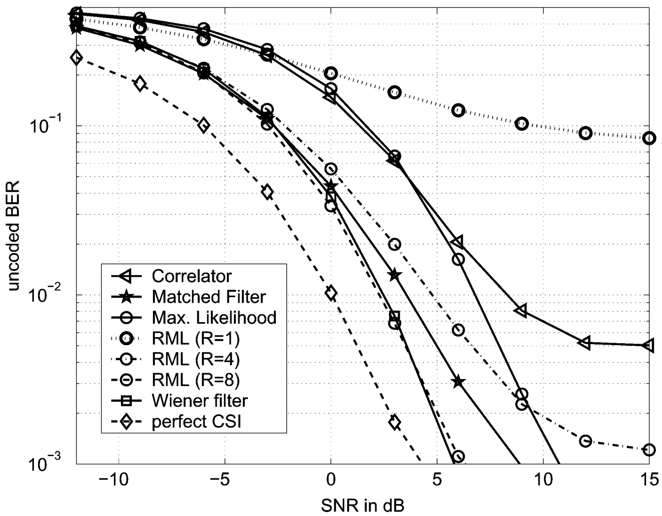


Fig. 15. BER of LMMSE equalizer (*joint equalization and despreading*) with different channel estimators (Section III) for frequency-selective channel $R \in \{1, 4, 8\} (N_{\rm P} = 20)$.

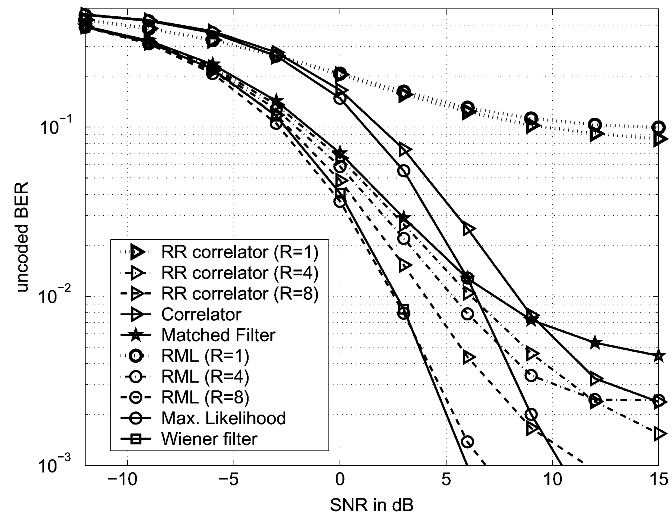


Fig. 16. BER of LMMSE equalizer (*chip-level equalization*) with generalized rake (EVD) and RR correlator in comparison with other estimators for frequency-selective channel $R \in \{1, 4, 8\} (N_{\rm P} = 20)$.

to the sensitivity of linear single-user equalizers (LMMSE and generalized rake) to channel estimation errors.

Results: At first, we evaluate the sensitivity of joint equalization and despreading as given in (52) and (53) to estimation errors for full-rank channel estimators and the RML (see Fig. 15). At a BER of 10^{-1}, the loss of the ML estimator is about 7.5 dB due to the short pilot sequence. We gained 4.4 dB with the MF and WF. We conclude that second-order statistics for channel estimation results in a significant improvement: The cheapest version is the MF performing as well as the WF at the considered uncoded BER of 10^{-1}, which is a typical point of operation for speech services in UMTS. The RML estimator $(R = 8)$ achieves a similar performance as the WF (Fig. 15), but the rank R needs to be optimized or chosen *a priori* according to the environment of the base station (e.g., the loss for $R = 4$ is considerable).

For chip-level equalization with the generalized rake (60) and a RR correlator (33), we have a gain of 4.4 dB for rank $R = 8$ (3.3 dB for $R = 4$) compared with full rank (see Fig. 16). The BER of the MF is comparable with the RR correlator for $R = 4$. For low SNR, the RR correlator provides estimates with the quality of the WF, whereas the equalizer is significantly less complex using the same rank reduction as the generalized rake. Thus, it improves performance and, at the same time, reduces complexity, exploiting the two time-scales of channel variations. The sensitivity of chip-level equalization to estimation errors of the different channel estimators (see Fig. 16) differs considerably from joint equalization and despreading (see Fig. 15).

Channel measurements show that the RR subspace is stable over a period much longer than the channel coherence time [32], e.g., $F = 120$ independent stationary channel realization are available for estimating $\boldsymbol{R_h}$ in an urban environment. Evaluating the impact of $\hat{\boldsymbol{R}}_{\boldsymbol{h}}$, we choose a considerably smaller number of $F = 20$ (see Section III-F). The methods described in Section III-F are employed for the estimation of $\boldsymbol{R_h}$. The MF and RR correlator (generalized rake) lose most in performance (see Figs. 17 and 18). This is mainly due to the fact that only the spatial correlations are exploited in the estimate of $\boldsymbol{R_h}$ (see Section III-F). All other channel estimators still achieve comparable BER performance gains, as in case of perfect knowledge of $\boldsymbol{R_h}$.

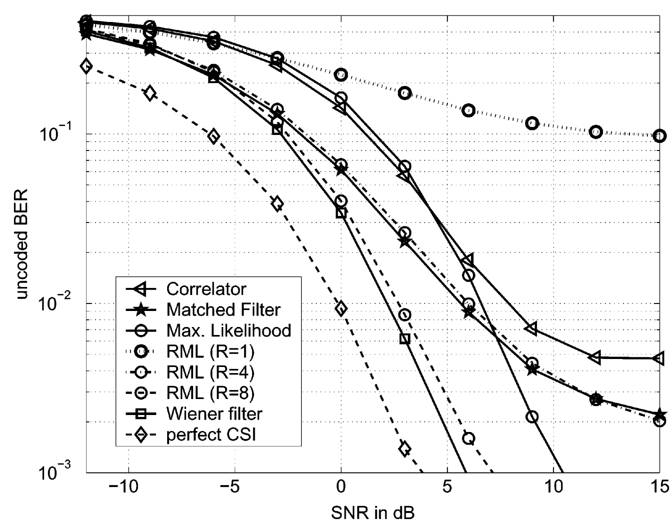


Fig. 17. BER of LMMSE equalizer (*joint equalization and despreading*) with different channel estimators (Section III) for frequency-selective channel $R \in \{1, 4, 8\} (N_{\rm P} = 20)$. Correlation matrix $\boldsymbol{R_h}$ estimated from $F = 20$ channel realizations as described in Section III-F. Only spatial channel correlations were estimated for matched filter, i.e., channel assumed uncorrelated for distinct delays.

VII. Conclusion

The channel estimation quality for space-time wireless communication channels improves significantly when the *different time-scales* of channel parameter variation are *exploited*. It was shown in the examples that second-order statistics, which describe the slowly changing spatial and temporal channel characteristics, can be estimated with sufficient accuracy to realize the performance gains in practice.

Based on the *comparison of linear channel estimators* w.r.t. the underlying channel model (stochastic/deterministic/RR), performance (analysis for correlated flat fading and simulation),

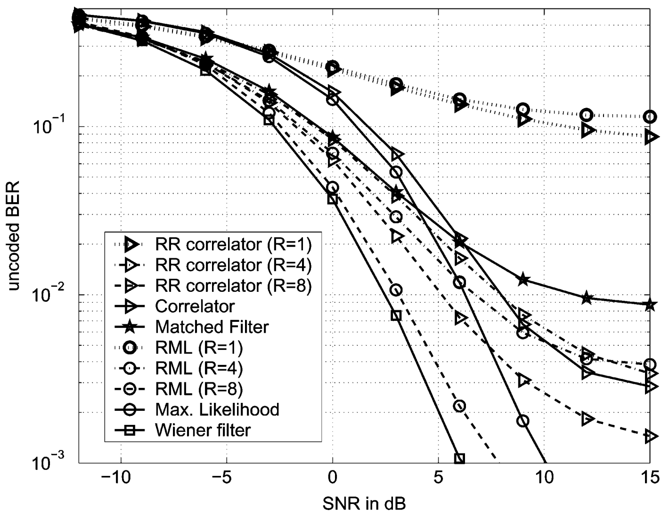


Fig. 18. BER of LMMSE equalizer (*chip-level equalization*) with generalized rake (EVD) and RR correlator in comparison with other estimators for frequency-selective channel $R \in \{1,4,8\} (N_{\mathrm{p}} = 20)$. Correlation matrix $\boldsymbol{R_h}$ estimated from $F = 20$ channel realizations, as described in Section III-F. Only spatial channel correlations were estimated for matched filter and RR correlator, i.e., channel assumed uncorrelated for distinct delays.

and complexity, we conclude the following: The ML approach and correlator do not take into account second-order statistics at all, resulting in a considerable performance loss compared with the Wiener approach. Considering channel estimation only, the Wiener estimator optimally exploits the second-order statistics with similar complexity as the RR correlator and ML. The presented *matched filter* is a *novel* low-complexity channel estimator with good performance for low SNR and short training sequences. Recent approaches like the RML are shown to be more complex *and* perform worse than the Wiener filter in case of perfectly known second-order statistics. Hence, *RR* techniques for channel estimation based on second-order statistics can only be recommended if the equalizer is designed in the subspace of the RR channel estimate, as for the *generalized rake receiver*, and if this subspace changes on a low rate, since the sum complexity will be reduced.

APPENDIX A
DERIVATION OF MATCHED FILTER CHANNEL ESTIMATION

From criterion (22) with (16), it follows that

$$\boldsymbol{W}_{\mathrm{MF}} = \underset{\boldsymbol{W}}{\operatorname{argmax}} \frac{|\mathrm{trace}(\boldsymbol{W}\boldsymbol{S}_{\mathrm{p}}\boldsymbol{R}_h)|^2}{\mathrm{trace}(\boldsymbol{W}\boldsymbol{R}_{\boldsymbol{\eta}}\boldsymbol{W}^{\mathrm{H}})} \tag{61}$$

which can be solved setting the complex derivative [2] to zero (note that $(\partial \mathrm{tr}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{B}))/(\partial \boldsymbol{A}^*) = \boldsymbol{B}, (\partial \mathrm{tr}(\boldsymbol{A}\boldsymbol{B}))/(\partial \boldsymbol{A}^*) = \boldsymbol{0}$):

$$\frac{\partial}{\partial \boldsymbol{W}^*} \mathrm{tr}(\boldsymbol{W}\boldsymbol{S}_{\mathrm{p}}\boldsymbol{R_h}) \frac{\mathrm{tr}\left(\boldsymbol{R_h}\boldsymbol{S}_{\mathrm{p}}^{\mathrm{H}}\boldsymbol{W}^{\mathrm{H}}\right)}{\mathrm{tr}(\boldsymbol{W}\boldsymbol{R}_{\boldsymbol{\eta}}\boldsymbol{W}^{\mathrm{H}})} = \frac{\mathrm{tr}(\boldsymbol{W}\boldsymbol{S}_{\mathrm{p}}\boldsymbol{R_h})}{(\mathrm{tr}(\boldsymbol{W}\boldsymbol{R}_{\boldsymbol{\eta}}\boldsymbol{W}^{\mathrm{H}}))^2} \times \left(\mathrm{tr}(\boldsymbol{W}\boldsymbol{R}_{\boldsymbol{\eta}}\boldsymbol{W}^{\mathrm{H}})\boldsymbol{R_h}\boldsymbol{S}_{\mathrm{p}}^{\mathrm{H}} - \mathrm{tr}\left(\boldsymbol{R_h}\boldsymbol{S}_{\mathrm{p}}^{\mathrm{H}}\boldsymbol{W}^{\mathrm{H}}\right)\boldsymbol{W}\boldsymbol{R}_{\boldsymbol{\eta}}\right) = \boldsymbol{0}.$$

As expected, from (22), the solution is unique up to a complex scalar α:

$$\boldsymbol{W}_{\mathrm{MF}} = \alpha \boldsymbol{R_h}\boldsymbol{S}_{\mathrm{p}}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\eta}}^{-1}. \tag{62}$$

APPENDIX B
DERIVATION OF RR ML CHANNEL ESTIMATION

To derive the RML channel estimator, we first assume that a finite number of blocks F are available. The proof is based on the work of Nicoli *et al.* [3], [5].

As in (25), maximizing the likelihood function is equivalent to minimizing (63).

$$\begin{aligned}\Theta(\boldsymbol{U}_{\mathrm{ST}}, \{\boldsymbol{\zeta}^f\}; F) &= \frac{1}{F}\sum_{f=1}^{F} \left\| \hat{\boldsymbol{h}}_{\mathrm{ML}}^f - \boldsymbol{U}_{\mathrm{ST}}\boldsymbol{\zeta}^f \right\|^2_{\boldsymbol{R}_{\boldsymbol{\varepsilon}}^{-1}} \\ &= \frac{1}{F}\sum_{f=1}^{F} \left(\hat{\boldsymbol{h}}_{\mathrm{ML}}^f - \boldsymbol{U}_{\mathrm{ST}}\boldsymbol{\zeta}^f\right)^{\mathrm{H}} \boldsymbol{R}_{\boldsymbol{\varepsilon}}^{-1} \left(\hat{\boldsymbol{h}}_{\mathrm{ML}}^f - \boldsymbol{U}_{\mathrm{ST}}\boldsymbol{\zeta}^f\right). \end{aligned} \tag{63}$$

First, we search for the optimum slot-dependent parameters $\boldsymbol{\zeta}^f$, evaluating the complex derivative [2]

$$\frac{\partial}{\partial \boldsymbol{\zeta}^{f,*}} \Theta(\boldsymbol{U}_{\mathrm{ST}}, \{\boldsymbol{\zeta}^f\}; F) = \boldsymbol{0}_{WL\times 1} \tag{64}$$

which yields

$$\hat{\boldsymbol{\zeta}}^f = \left(\boldsymbol{V}_R^{\mathrm{H}}\boldsymbol{V}_R\right)^{-1} \boldsymbol{V}_R^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{\varepsilon}}^{-1/2}\hat{\boldsymbol{h}}_{\mathrm{ML}}^f \tag{65}$$

with $\boldsymbol{V}_R = \boldsymbol{R}_{\boldsymbol{\varepsilon}}^{-1/2}\boldsymbol{U}_{\mathrm{ST}}$. These parameters are substituted in (63), resulting in a modified cost function. Minimizing this function gives the same solution as maximizing

$$\tilde{\Theta}(\boldsymbol{V}_R; F) = \mathrm{trace}\left(\hat{\boldsymbol{R}}_F\boldsymbol{V}_R\left(\boldsymbol{V}_R^{\mathrm{H}}\boldsymbol{V}_R\right)^{-1}\boldsymbol{V}_R^{\mathrm{H}}\right) \tag{66}$$

with $\hat{\boldsymbol{R}}_F = \boldsymbol{R}_{\boldsymbol{\varepsilon}}^{-1/2}((1/F)\sum_{f=1}^F \hat{\boldsymbol{h}}_{\mathrm{ML}}^f\hat{\boldsymbol{h}}_{\mathrm{ML}}^{f,\mathrm{H}})\boldsymbol{R}_{\boldsymbol{\varepsilon}}^{-1/2}$. Now we take the limit $\tilde{\Theta}(\boldsymbol{V}_R) = \lim_{F\to\infty}\tilde{\Theta}(\boldsymbol{V}_R; F)$. For (mean-square) ergodic processes, the time average converges to the ensemble average, and the estimate of the correlation matrix in (66) is equivalent to

$$\lim_{F\to\infty} \hat{\boldsymbol{R}}_F = \boldsymbol{R} + \boldsymbol{I}_{ML_{\mathrm{c}}}. \tag{67}$$

With $\boldsymbol{R} = \boldsymbol{R}_{\boldsymbol{\varepsilon}}^{-1/2}\boldsymbol{R_h}\boldsymbol{R}_{\boldsymbol{\varepsilon}}^{-1/2}$, (66) can be written in the limit as

$$\tilde{\Theta}(\boldsymbol{V}_R) = \mathrm{trace}\left(\boldsymbol{R}\boldsymbol{V}_R\left(\boldsymbol{V}_R^{\mathrm{H}}\boldsymbol{V}_R\right)^{-1}\boldsymbol{V}_R^{\mathrm{H}}\right) \tag{68}$$

which is the generalized Rayleigh quotient [40]. The projector on the subspace spanned by the R leading eigenvectors of $\boldsymbol{R}$ maximizes $\tilde{\Theta}(\boldsymbol{V}_R)$. The projector on the subspace is unique but not its basis $\boldsymbol{V}_R$. For convenience, we choose the unitary basis $\boldsymbol{V}_R$ given by the R eigenvectors associated with the largest eigenvalues of $\boldsymbol{R}$. This leads to the RML estimator in (26). Note the similarity to the independently motivated optimization of the RR correlator (36), which does not include information about the noise statistics.

APPENDIX C
EQUIVALENCE OF PRE- AND POSTPROCESSING FOR THE RR CORRELATOR

First, we observe that the following formulations for the model of the received pilot (12) are equivalent [cf. (1)]:

$$\begin{aligned}\boldsymbol{y}_{\mathrm{p}} &= (\boldsymbol{S}'_{\mathrm{p}} \otimes \boldsymbol{I}_M)\mathrm{vec}(\boldsymbol{H}) + \boldsymbol{\eta}_{\mathrm{p}} = \mathrm{vec}\left(\boldsymbol{H}\boldsymbol{S}'^{,\mathrm{T}}_{\mathrm{p}}\right) + \boldsymbol{\eta}_{\mathrm{p}} \\ \boldsymbol{Y}_{\mathrm{p}} &= \boldsymbol{H}\boldsymbol{S}'^{,\mathrm{T}}_{\mathrm{p}} + \boldsymbol{N}_{\mathrm{p}}\end{aligned} \tag{69}$$

$$\mathrm{P_b}(\{\lambda_m\},\gamma_\mathrm{p},\gamma_\mathrm{d}) = \sum_{m=1}^{R}\left[\prod_{i=1}^{R}\left(1-\frac{\beta_i\lambda_i\left(1+\sqrt{\left(1+\frac{1}{\lambda_i\gamma_\mathrm{p}}\right)\left(1+\frac{1}{\lambda_i\gamma_\mathrm{d}}\right)}\right)}{\beta_m\lambda_m\left(1-\sqrt{\left(1+\frac{1}{\lambda_m\gamma_\mathrm{p}}\right)\left(1+\frac{1}{\lambda_m\gamma_\mathrm{d}}\right)}\right)}\right)\right.$$
$$\left.\times\prod_{\substack{i=1\\ i\neq m}}^{R}\left(1-\frac{\beta_i\lambda_i\left(1-\sqrt{\left(1+\frac{1}{\lambda_i\gamma_\mathrm{p}}\right)\left(1+\frac{1}{\lambda_i\gamma_\mathrm{d}}\right)}\right)}{\beta_m\lambda_m\left(1-\sqrt{\left(1+\frac{1}{\lambda_m\gamma_\mathrm{p}}\right)\left(1+\frac{1}{\lambda_m\gamma_\mathrm{d}}\right)}\right)}\right)\right]^{-1} \quad (75)$$

where $\boldsymbol{H}$ is defined as in (9), $\boldsymbol{y}_\mathrm{p} = \mathrm{vec}(\boldsymbol{Y}_\mathrm{p})$, and $\boldsymbol{Y}_\mathrm{p} = [\boldsymbol{y}_\mathrm{p}[1],\ldots,\boldsymbol{y}_\mathrm{p}[N_\mathrm{p}]] \in \mathbb{C}^{M\times N_\mathrm{p}}$ (equivalently for $\boldsymbol{N}_\mathrm{p}$).

In the sequel, we show the equivalence of (32) and (34). From (32), it follows, with (69), that

$$\hat{\boldsymbol{h}}_R = \frac{1}{N_\mathrm{p}P_\mathrm{p}}\boldsymbol{M}_R^\mathrm{H}\mathrm{vec}(\boldsymbol{Y}_\mathrm{p}\boldsymbol{S}_\mathrm{p}^{\prime,*}) \quad (70)$$

Rearranging, we obtain (34) with (1):

$$\hat{\boldsymbol{h}}_R = \boldsymbol{M}_R^\mathrm{H}(\boldsymbol{I}_{L_\mathrm{c}}\otimes\boldsymbol{Y}_\mathrm{p})\frac{1}{N_\mathrm{p}P_\mathrm{p}}\mathrm{vec}(\boldsymbol{S}_\mathrm{p}^{\prime,*}). \quad (71)$$

APPENDIX D
BEP FOR COMBINING CORRELATED SIGNALS BASED ON CHANNEL ESTIMATES

An analytical expression for the BEP in the case of ML correlated (space-time) diversity branches is of particular interest in channel estimation for space-time receivers (see Section IV). It is derived under the following assumptions: BPSK data symbols $s_\mathrm{d}[i] \in \{-1,+1\}$ are transmitted, channel coefficients are distributed as $\boldsymbol{h} \sim \mathcal{N}_\mathrm{c}(\boldsymbol{0},\boldsymbol{R_h})$, and maximum ratio combining based on noisy channel coefficients from linear channel estimators at the receiver. Moreover, $N_\mathrm{p}P_\mathrm{p} \neq P_\mathrm{d}$ is needed for the results below. We generalize the results in [50] for ML channel estimation to general linear estimators. The derivations are mainly based on the results in [51].

Equation (49) can be written as a quadratic form

$$\hat{s}_\mathrm{d}[i] = \frac{1}{2}\left(\hat{\boldsymbol{h}}^\mathrm{H}\boldsymbol{U}_R\boldsymbol{y}_\mathrm{d}[i] + \hat{\boldsymbol{h}}^\mathrm{T}\boldsymbol{U}_R^*\boldsymbol{y}_\mathrm{d}[i]^*\right) \quad (72)$$
$$= \sum_{m=1}^{R}\boldsymbol{v}_m^\mathrm{H}\boldsymbol{Q}_2\boldsymbol{v}_m = \boldsymbol{v}^\mathrm{H}\boldsymbol{Q}_{2R}\boldsymbol{v} \quad (73)$$

with $\boldsymbol{v}_m = [\boldsymbol{u}_m^\mathrm{H}\hat{\boldsymbol{h}}, y_{\mathrm{d},m}[i]]^\mathrm{T} = [\hat{h}_m, y_{\mathrm{d},m}[i]]^\mathrm{T}$, $\hat{h}_m \hat{=} \boldsymbol{u}_m^\mathrm{H}\hat{\boldsymbol{h}}$, $\boldsymbol{v} = [\boldsymbol{v}_1^\mathrm{T},\ldots,\boldsymbol{v}_R^\mathrm{T}]^\mathrm{T}$, and $\boldsymbol{Q}_{2R} = 1/2\,(\boldsymbol{I}_R\otimes[\boldsymbol{e}_2,\boldsymbol{e}_1])$. $\boldsymbol{u}_m$ is the mth column of $\boldsymbol{U}_R$. $\boldsymbol{v}$ is distributed as $\mathcal{N}_\mathrm{c}(\boldsymbol{0},\boldsymbol{L})$ with correlation matrix $\boldsymbol{L} = \mathrm{E}[\boldsymbol{v}\boldsymbol{v}^\mathrm{H}] = \mathrm{diag}(\{\boldsymbol{L}_1,\ldots,\boldsymbol{L}_R\})$, and

$$\boldsymbol{L}_m = \begin{bmatrix}\mathrm{E}[|\hat{h}_m|^2] & \mathrm{E}[\hat{h}_m y_{\mathrm{d},m}^*[i]]\\ \mathrm{E}[\hat{h}_m^* y_{\mathrm{d},m}[i]] & \mathrm{E}[|y_{\mathrm{d},m}[i]|^2]\end{bmatrix}$$
$$= \begin{bmatrix}\beta_m^2\left(\lambda_m+\frac{\sigma_\mathrm{n}^2}{NP_\mathrm{p}}\right) & \sqrt{P_\mathrm{d}}\,\lambda_m\,\beta_m\\ \sqrt{P_\mathrm{d}}\,\lambda_m\,\beta_m & P_\mathrm{d}\,\lambda_m+\sigma_\mathrm{n}^2\end{bmatrix} \quad (74)$$

if $s_\mathrm{d}[i] = +\sqrt{P_\mathrm{d}}$ is the transmitted symbol.

Based on the distribution of the quadratic form in complex circular symmetric Gaussian random variables $\boldsymbol{v}$, the average BEP in (75), shown at the top of the page, can be derived [50], [51]. The parameter β_m characterizing the channel estimation scheme is $\beta_m = 1$ for ML and correlator, $\beta_m = \lambda_m$ for the MF, and $\beta_m = \lambda_m/(\lambda_m + (1)/(\gamma_\mathrm{p}))$ for the WF. For $R = \mathrm{ML}$, we have the case of full-rank channel estimation, and $R < \mathrm{ML}$ describes RR schemes and the generalized rake receiver. The ratio of transmit power to noise at the receiver in the data channel is $\gamma_\mathrm{d} = (P_\mathrm{d})/(\sigma_\mathrm{n}^2)$, and the effective SNR in the pilot channel is $\gamma_\mathrm{p} = (N_\mathrm{p}P_\mathrm{p})/(\sigma_\mathrm{n}^2)$, where we have (75).

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their valuable comments that significantly improved the quality of the paper.

REFERENCES

[1] A. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Process. Mag.*, vol. 14, no. 6, pp. 49–83, Nov. 1997.

[2] S. M. Kay, *Fundamentals of Statistical Signal Processing—Estimation Theory*, First ed. Englewwod Cliffs, NJ: PTR Prentice-Hall, 1993.

[3] M. Nicoli, O. Simeone, and U. Spagnolini, "Multi-slot estimation of fast-varying space-time communication channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1184–1195, May 2003.

[4] ——, "Multislot estimation of fast-varying space-time channels in TD-CDMA systems," *IEEE Commun. Lett.*, vol. 6, no. 9, pp. 376–378, Sep. 2002.

[5] O. Simeone and U. Spagnolini, "Multi-slot estimation of space-time channels," in *Proc. IEEE Int. Conf. Commun.*, Apr./May 2002, pp. 802–806.

[6] P. Stoica and M. Viberg, "Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions," *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 3069–3078, Dec. 1996.

[7] P. Forster and T. Asté, "Maximum likelihood multichannel estimation under reduced-rank constraint," in *Proc. IEEE Conf. Acoust. Speech Signal Process.*, 1998, pp. 3317–3320.

[8] J. Jelitto and G. Fettweis, "Reduced dimension space-time processing for multi-antenna wireless systems," *IEEE Wireless Commun. Mag.*, vol. 9, no. 6, pp. 18–25, Dec. 2002.

[9] J. Jelitto, M. Bronzel, and G. Fettweis, "From smart antennas to reduced dimension space-time processing," *Frequenz, Zeitschrift für Telekommunikation*, vol. 55, no. 5–6, pp. 165–170, May/June 2001.

[10] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.

[11] F. Dietrich and W. Utschick, "On the effective spatio-temporal rank of wireless communication channels," in *Proc. 13th IEEE Symp. Pers. Indoor Mobile Radio Commun.*, Lisbon, Portugal, Sep. 2002.

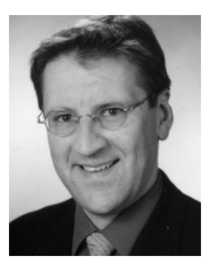
[12] M. Stege, P. Zillmann, and G. Fettweis, "MIMO channel estimation with dimension reduction," in *Proc. Wireless Pers. Multimedia Commun.*, Oct. 2002.

[13] 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.

[14] J. K. Cavers, "An analysis of pilot symbol assisted modulation for rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.
[15] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.
[16] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers*. New York: Wiley, 1998.
[17] M. Tüchler, A. Singer, and R. Koetter, "Minimum mean squared error equalization using *a priori* information," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 673–683, Mar. 2002.
[18] M. Honig and M. K. Tsatsanis, "Adaptive techniques for multiuser CDMA receivers," *IEEE Signal Process. Mag.*, vol. 17, no. 3, pp. 49–61, May 2000.
[19] R. Price and P. E. Green Jr., "A communication technique for multipath channels," in *Proc. IRE*, vol. 46, Mar. 1958, pp. 555–570.
[20] J. G. Proakis, *Digital Communications*, Third ed. New York: McGraw-Hill, 1995.
[21] C. Brunner, "Efficient Space-Time Processing Schemes for WCDMA," Ph.D. dissertation, Inst. Network Theory Signal Process., Tech. Univ. Munich, Munich, Germany, 2000.
[22] C. Brunner, J. S. Hammerschmidt, A. Seeger, and J. A. Nossek, "Space-time eigenrake and downlink eigenbeamformer: Exploiting long-term and short-term channel properties in WCDMA," in *Proc. IEEE GLOBECOM*, Nov. 2000.
[23] J. S. Hammerschmidt, "Adaptive Space and Space-Time Signal Processing for High-Rate Mobile Data Receivers," Ph.D. dissertation, Inst. Integrated Circuits, Technical Univ. Munich, Munich, Germany, 2000.
[24] C. Brunner, M. Haardt, and J. A. Nossek, "On space-time rake receiver structures for WCDMA," in *Proc. 33rd Asilomar Conf. Signals, Syst., Computers*, Pacific Grove, CA, Oct. 1999. invited paper.
[25] A. F. Naguib, "Adaptive Antennas for CDMA Wireless Networks," Ph.D. dissertation, Stanford Univ., Stanford, CA, Aug. 1996.
[26] F. Dietrich, M. T. Ivrlac, and J. A. Nossek, "On performance limits of optimum reduced-rank channel estimation," in *Proceedings of IEEE Globecom*, Taipei, Taiwan, R.O.C., Nov. 2002.
[27] I. Viering, T. Frey, and G. Schnabl, "Hybrid beamforming: Eigen beamforming on beam signals," in *Proc. Int. Zurich Seminar Broadband Commun.*, Feb. 2002.
[28] F. Dietrich and W. Utschick, "Nonlinear receiver concepts in reduced dimensions for WCDMA," in *Proc. Fourth Int. ITG Conf. Source Channel Coding*, Berlin, Germany, Jan. 2002.
[29] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. CAS-25, pp. 772–781, Sep. 1978.
[30] I. Viering, T. Frey, and G. Schnabl, "Potential of coefficient reduction in delay, space and time based on measurements," in *Proc. CISS*, Mar. 2003.
[31] H. Hofstetter, I. Viering, and P. H. Lehne, "Spatial and temporal long term properties of typical urban base stations at different heights," in *Proc. COST273 Sixth Meet.*, Barcelona, Spain, Jan. 2003.
[32] I. Viering, H. Hofstetter, and W. Utschick, "Validity of spatial covariance matrices over time and frequency," in *Proc. IEEE Globecom*, Taipei, Taiwan, R.O.C., Nov. 2002.
[33] H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, Second ed. Englewood Cliffs, NJ: Prentice-Hall, 1994.
[34] H. V. Poor, *An Intoduction to Signal Detection and Estimation*, Second ed. New York: Springer-Verlag, 1994.
[35] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, to be published.
[36] T. A. Hua and R. F. Gunst, "Generalized ridge regression: A note on negative ridge parameters," *Commun. Statist. Theory Methods*, vol. 12, no. 1, pp. 37–45, 1983.
[37] Y. Hua, M. Nikpour, and P. Stoica, "Optimal reduced-rank estimation and filtering," *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 457–469, Mar. 2001.
[38] J. H. Manton, R. Mahony, and Y. Hua, "The geometry of weighted low-rank approximations," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 500–514, Feb. 2003.
[39] W. Utschick, "Tracking of signal subspace projectors," *IEEE Trans. Signal Process.*, vol. 50, no. 4, pp. 769–778, Apr. 2002.
[40] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*, First ed. New York: Springer-Verlag, 1994.
[41] E. F. Deprettere, *SVD and Signal Processing*, First ed. Amsterdam, The Netherlands: North-Holland, 1988.
[42] J. S. Lim, *Two-Dimensional Signal and Image Processing*. Englewood Cliffs, NJ: Prentice-Hall PTR, 1990.
[43] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, pp. 387–392, Apr. 1985.
[44] M. Nicoli and U. Spagnolini, "Reduced-rank channel estimation and rank order selection for CDMA systems," in *Proc. IEEE Int. Conf. Commun.*, Helsinki, Finland, Jun. 2001, pp. 2737–2741.
[45] P. E. Mogensen, K. I. Pedersen, B. Fleury, F. Frederiksen, P. Leth-Espensen, K. Olesen, and S. L. Larsen, "TSUNAMI II: 2D-Channel Characteristics," Aalborg Univ., Aalborg, Denmark, Tech. Rep., Dec. 1997.
[46] G. H. Golub and C. F. V. Loan, *Matrix Computations*, Second ed. Baltimore, MD: John Hopkins Univ. Press, 1989.
[47] I. Viering, T. Grundler, and A. Seeger, "Improving uplink adaptive antenna algorithms for WCDMA by covariance matrix compensation," in *Proc. IEEE Veh. Technol. Conf.*, Vancouver, BC, Canada, Sep. 2002.
[48] F. Ling, "Optimal reception, performance bound, and cutoff rate analysis of reference-assisted coherent cdma communications with applications," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 1583–1592, Oct. 1999.
[49] M. D. Zoltowski, T. P. Krauss, and S. Chowdhury, "Chip-level MMSE equalization for high-speed synchronous CDMA in frequency selective multipath," in *Proc. SPIE*, Apr. 2000.
[50] F. A. Dietrich and W. Utschick, "Maximum ratio combining of correlated rayleigh fading channels with imperfect channel knowledge," *IEEE Commun. Lett.*, vol. 7, no. 9, pp. 419–421, Sep. 2003.
[51] G. L. Turin, "The characteristic function of hermitian quadratic forms in complex normal variables," *Biometrika*, vol. 47, pp. 199–201, Jun. 1960.

Frank A. Dietrich (S'99) was born in Fürstenfeldbruck, Germany, in 1976 and received the Diploma degree (summa cum laude) in electrical engineering from Munich University of Technology, Munich, Germany, in 2001. Funded by a Fulbright scholarship and a grant from the American European Engineering Exchange program, he studied computer and systems engineering at the Rensselaer Polytechnic Institute, Troy, NY, from 1998 to 1999. Currently, he is working as a research assistant, pursuing the Ph.D. degree at the Institute for Circuit Theory and Signal Processing, Munich University of Technology.

During summer 1999, he was involved in array processing research at Siemens Corporate Research, Princeton, NJ. His main research interests are in adaptive antennas, channel estimation, and robust space-time signal processing at the transmitter and receiver with focus on wireless communication systems.



Wolfgang Utschick was born in Ingolstadt, Germany, in 1964. He completed several industrial education programs before receiving the Dipl.-Ing. and Dr.-Ing. degrees (both with honors) in electrical engineering from the Munich University of Technology (TUM), in 1993 and 1998, respectively. During this period, he held a scholarship from the Bavarian Ministry of Education for exceptional students and a scholarship from Siemens AG.

In 1993, he became a part-time lecturer at a technical school for industrial education. From 1998 to 2002, he was head of the Signal Processing Group at the Institute of Circuit Theory and Signal Processing, TUM. He has been a guest researcher at the ETH Zurich, Zurich, Switzerland, in the summer of 2000, and from 2000 to 2001, he was instrumental in the 3rd Generation Partnership Project as an academic consultant in the field of multielement antenna wireless communication systems. Since October 2002, he has been Professor of signal processing with the Department of Electrical Engineering and Information Technology, TUM. He completed the biennial PROLEHRE program for professional academic teaching. He gives national and international undergraduate and graduate courses in the field of signal processing, communication systems, and smart antennas and conducts research on array signal processing, multielement antenna communication systems, x-layer optimization, neural networks, etc. He has authored and co-authored many papers on the field of neural networks and signal processing with regard to communication applications and holds several patents in the field of wireless communications.