14

Blind and semiblind MIMO channel estimation

Dirk Slock
Institut Eurécom

Abdelkader Medles
Lucent Technologies Bell Labs

The goal of this chapter is to expose a number of key ideas in blind and, especially, semiblind (SB) channel estimation (CE), and attract attention to various considerations that should be kept in mind in this context. As will become clear, the topic considered is vast. Due to space limitations, the inclusion and discussion of references is far from exhaustive. See also de Carvalho and Slock (2001) for an overview of semiblind single-input multiple-output (SIMO) channel estimation approaches. The use of blind information in digital communications is motivated by a desire to limit capacity loss due to training. Such capacity loss potentially increases with increasing time variation, occupied bandwidth, and number of transmitters. In other applications, blind techniques may be the only option (e.g., acoustic dereverberation). A particularity of digital communications, however, is that the sources are discrete time, white, and finite alphabet.

14.1 Signal model

In a first instance, the nonblind information considered will be provided by training or pilot information. As for terminology, the term *training sequence* (TS) tends to be used for a limited consecutive sequence of known symbols, whereas pilot symbols are typically isolated known symbols. A pilot signal is a continuous stream of known symbols, superimposed on the data signal.

Consider an (in a first instance time-invariant) discrete-time multiple-input multiple-output (MIMO) system with N_t inputs and N_r outputs,

$$\underbrace{\mathbf{y}_k}_{N_r\times 1} - \underbrace{\mathbf{v}_k}_{N_r\times 1} = \sum_{i=1}^{N_t}\sum_{l=0}^{L_i} \underbrace{\mathbf{h}_{i,l}}_{N_r\times 1}\,\underbrace{x_{i,k-l}}_{1\times 1} = \sum_{l=0}^{L} \underbrace{\mathbf{H}_l}_{N_r\times N_t}\,\underbrace{\mathbf{x}_{k-l}}_{N_t\times 1} = \underbrace{\mathbf{H}(q)}_{N_r\times N_t}\,\underbrace{\mathbf{x}_k}_{N_t\times 1}, \quad (14.1)$$

where $\mathbf{H}(q) = \sum_{l=0}^{L} \mathbf{H}_l\, q^{-l}$, $L = \max\{L_i, i = 1, \ldots, N_t\}$, and we introduced the one-sample delay operator: $q^{-1}\,\mathbf{x}_k = \mathbf{x}_{k-1}$; $\mathbf{v}_k$ is the additive noise. For the case of an (orthogonal frequency-division multiplexing (OFDM) or single-carrier) cyclic prefix (CP) block transmission system with N samples per block, the introduction of a cyclic prefix of $K \geq L$ samples means that the last K samples of the current block (of N samples) are repeated before the actual block. If we assume without loss of generality that the current block starts at time 0, then samples $\mathbf{x}_{N-K} \cdots \mathbf{x}_{N-1}$ are repeated at time instants $-K, \ldots, -1$. This means that the output at sample periods $0, \ldots, N-1$ can be written in matrix form as

$$\begin{bmatrix} \mathbf{y}_0 \\ \vdots \\ \mathbf{y}_{N-1} \end{bmatrix} = \mathbf{Y}_0 = \mathcal{C}(\mathbf{h})\ \mathbf{X}_0 + \mathbf{V}_0, \tag{14.2}$$

where the matrix $\mathcal{C}(\mathbf{h})$ is not only (block) Toeplitz, but even (block) circulant: each (block) row is obtained by a (block) cyclic shift to the right of the previous row. Consider now applying an N-point FFT to both sides of (14.2) at block m:

$$F_{N,N_r}\mathbf{Y}_m = F_{N,N_r}\mathcal{C}(\mathbf{h})F^{-1}_{N,N_t} \quad F_{N,N_t}\mathbf{X}_m \quad + \quad F_{N,N_r}\mathbf{V}_m, \tag{14.3}$$

or with new notation:

$$\boldsymbol{Y}_m = \mathcal{H} \qquad\qquad\qquad \boldsymbol{X}_m \qquad + \quad \boldsymbol{V}_m, \tag{14.4}$$

where $F_{N,p} = F_N \otimes I_p$ (Kronecker product: $A \otimes B = [a_{ij}B]$), F_N is the N-point $N \times N$ DFT matrix, $\mathcal{H} = \mathrm{diag}\{\boldsymbol{H}_0, \ldots, \boldsymbol{H}_{M-1}\}$ is a block-diagonal matrix with diagonal blocks $\boldsymbol{H}_n = \sum_{l=0}^{L} \mathbf{H}_l e^{-j2\pi ln/N}$, the $N_r \times N_t$ channel transfer function at tone (subcarrier) n (frequency = n/N times the sample frequency). In OFDM, the transmitted symbols are in $\boldsymbol{X}_m$ and, hence, are in the frequency domain. The corresponding time domain samples are in $\mathbf{X}_m$. The OFDM symbol period index is m. In single-carrier (SC) CP systems, the transmitted symbols are in $\mathbf{X}_m$ and, hence, are in the time domain. The corresponding frequency domain data are in $\boldsymbol{X}_m$. The components of $\mathbf{V}_m$ are assumed to be white noise; hence, the components of $\boldsymbol{V}_m$ are also white. At tone $n \in \{0, \ldots, N-1\}$, we get the following input-output relation

$$\underbrace{\boldsymbol{y}_n[m]}_{N_r \times 1} = \underbrace{\boldsymbol{H}_n}_{N_r \times N_t} \underbrace{\boldsymbol{x}_n[m]}_{N_t \times 1} + \underbrace{\boldsymbol{v}_n[m]}_{N_r \times 1}, \tag{14.5}$$

where the elements (symbols) of $\boldsymbol{x}_n[m]$ belong to some finite alphabet (constellation) in the case of OFDM.

14.2 Structured deterministic and stochastic channel models

The MIMO channel for the spatial multiplexing case is a special form of the multiuser channel, one in which the multiple users are colocated. As a result, the channel lengths L_i are usually equal. This remains the case even if sensors with different polarizations are used, but not necessarily if pattern (beam) diversity is used. Also refer to Chapter 1 for channel models.

14.2.1 Pathwise channel models

To go beyond the finite impulse response (FIR) channel model, more structured channel models, such as infinite impulse response (IIR) models, could be useful for applications in which the output may return to the input in some sense, creating natural modes, as in acoustic applications in an enclosed medium, or for transmission lines. For wireless communications, however, structured channel models can be based on the physics of the propagation mechanism. Since attenuation increases rapidly after a few reflections or diffractions, direct input-to-output transfer models are most appropriate. So, apart from the nonparametric FIR model (in which knowledge of the pulse shape may also be expressed, see Section 14.6.1), parametric pathwise channel models may be considered. The specular time-varying MIMO channel impulse response $\mathbf{H}(t, \tau)$ is of the form

$$\mathbf{H}(t, kT) = \sum_{i=1}^{N_p} A_i(t)\, e^{j2\pi f_i t}\, \mathbf{a}_R(\phi_i)\mathbf{a}_T^T(\theta_i)\, p(kT - \tau_i),$$

to which each path contributes a rank-one component in three dimensions: delay, direction of arrival (DOA), and direction of departure (DOD). The N_p pathwise contributions involve: complex attenuation A_i; Doppler shift $f_i \in (-f_d, f_d)$ (where f_d is the Doppler frequency); angle of departure θ_i; angle of arrival ϕ_i; path delay τ_i; $\mathbf{a}_R(\cdot)$, $\mathbf{a}_T(\cdot)$ are the Rx/Tx[†] antenna array responses, $p(\cdot)$ is the pulse shape (Tx filter), and T is the symbol period. Note that the DOA and DOD may involve more than one angle parameter. Consider stacking the columns of the consecutive impulse response matrix coefficients to obtain

$$\mathbf{h}(t) = \mathrm{vec}\{\mathbf{H}(t, kT)\} = \sum_{i=1}^{N_p} \mathbf{h}^{(i)} e^{j2\pi f_i t} A_i(t) = \mathbf{P}\mathbf{D}(t)\mathbf{A}(t), \tag{14.6}$$

[†]For the sake of brevity, Rx may alternatively refer to receiver(s)/receive/receiving etc., Tx to transmitter(s)/transmit/transmitting and similarly for other abbreviations.

where $\mathbf{h}$ is $N \times 1$ with $N = N_t N_r L$, $\mathbf{P} = [\mathbf{h}^{(1)} \cdots \mathbf{h}^{(N_p)}]$, $\mathbf{A}(t)$ is $N_p \times 1$, containing the $e^{j2\pi f_i t} A_i(t)$. The "fast" parameters $A_i(t)$ are band-limited (often to much less than f_d), so that $e^{j2\pi f_i t} A_i(t)$ is a modulated lowpass signal. Fast fading is essentially due to the Doppler shifts f_i, and leads to much faster variation of the channel coefficients with a more complex Doppler spectrum (compared with the $A_i(t)$), due to a superposition of path contributions. The "slow" parameters f_i, τ_i, θ_i, ϕ_i vary much more slowly, according to the slow fading (varying multipath structure).

14.2.2 Deterministic and stochastic time-varying channel models

Two approaches can be introduced for time variation: modeling $\mathbf{h}_k = \mathbf{h}(kT)$ as a stationary vector process, called the stationary model for short, or using a basis expansion model (BEM), in which the time-varying channel coefficients are expanded into known time-varying basis functions, and the unknown channel parameters are now no longer the channel coefficients but the combination coefficients in the BEM. The BEM model was introduced by Y. Grenier around 1980 for time-varying filtering, by E. Karlsson in the early 1990s for time-varying channel modeling, and by M. Tsatsanis and G. Giannakis in 1996 for blind time-varying channel estimation. The BEM has also been revived and generalized in the canonical coordinates concept of A. Sayeed, in which the basis functions are signal independent (nonparametric).

The choice of the channel model interacts with the design of the modulation format. For instance, a stationary model may be more appealing for the case of long transmission packets (e.g., corresponding to a packet of data within one convolutional coding operation), whereas the BEM model might be more appealing in the case of shorter packets (e.g., OFDM symbols), the length of which would correspond to a potential subsampling period of the channel variation (related to maximum Doppler spread) appearing in the BEM. In the case of the stationary model, the stationarity suggests Wiener filtering of brute channel estimates, but the transients at both edges of the packet may be more properly treated with a Kalman filter or smoother. For the BEM model, the question arises whether to model potential correlations between basis expansion coefficients, or just their variances (stationary and BEM models are equivalent if all correlations are accounted for in the BEM model). A (nonparametric) BEM is well designed if the expansion coefficients are fairly uncorrelated. A parametric BEM approach is obtained when working with the Karhunen–Loève expansion of the channel temporal variation correlation. Nonparametric BEMs typically correspond to subsampling and interpolation of lowpass signals. The *block-fading* model is a BEM with rectangular basis functions, leading to a subsampling of the channel variation. In Tong *et al.* (2004), it is mentioned that a stationary channel estimation error is obtained when the downsampled version of the channel satisfies the Nyquist criterion (i.e., allows reconstruction of the continuous-time lowpass channel variation). The block-fading and stationary models can be combined (subsampling first). BEMs can be applied to $\mathbf{h}(t)$ directly or to $\mathbf{A}(t)$. Modeling the $A_i(kT)$ (and hence the $e^{j2\pi f_i kT} A_i(kT)$) as independent autoregressive (AR) processes leads to a subspace AR model (Slock, 2004a) for $\mathbf{h}_k$ with spectrum $S_{\mathbf{hh}}(z) = \mathbf{P} S_{\mathbf{AA}}(z) \mathbf{P}^H$. So the channel may be quite predictable (especially in the wideband MIMO case), since $S_{\mathbf{hh}}(z)$ may be doubly singular, owing to limited bandwidth of $S_{\mathbf{AA}}(z)$ and limited rank, $N_p < N$. The block-fading model applied to the $A_i(t)$ leads to a four-dimensional rank-one contribution per path to $S_{\mathbf{hh}}(z)$. Although it would rarely make sense in practice, a four-dimensional separable correlation model of the form $S_{\mathbf{hh}}(f) = R_\tau \otimes R_T \otimes R_R \, S_d(f)$ may be considered for the purpose of performance analysis, where R_τ is the correlation matrix (R) between delays (typically diagonal with power delay profile), R_T is the Tx side R, R_R is the Rx side R, and $S_d(f)$ is the scalar common Doppler spectrum (shape) of all channel coefficients. For the estimation of structured channels, an unstructured version may be estimated first, with the structure being imposed in a second stage.

14.3 Performance indicators

14.3.1 Channel capacity

See also Chapters 4 and 6 for channel capacity and its dependence on channel state information at the receiver (CSIR), or a relevant discussion and references in Tong *et al.* (2004). However, in those references, mostly the gap to channel capacity with CSIR of schemes based on training only is studied.

Consider a possibly time-selective frequency-flat channel $\mathbf{y}_k = \mathbf{H}_k \, \mathbf{x}_k + \mathbf{v}_k$. Within the block duration $T = N_{TS} + N_B$, the pilot symbols are contained in $\mathbf{X}^{TS}$, whereas the data (blind part) is contained in $\mathbf{X}^B$, where we define $\mathbf{X}_i^k = [\mathbf{x}_i, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_k]$.

Mutual information (MI) decomposition

The MI is a crucial quantity, since the channel capacity is the MI when an optimal input distribution is used. Assuming the Rx has no side information about the channel, apart from the Rx signal, the MI between the Tx and

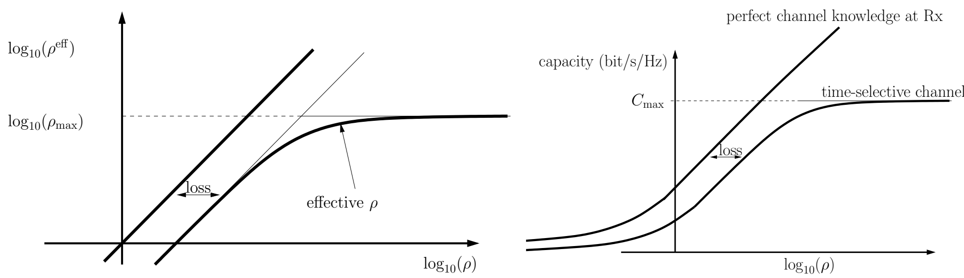

Fig. 14.1. Effective SNR and capacity vs. SNR behavior for a time selective channel.

Rx signals is then

$$I(\mathbf{Y}^{TS},\mathbf{Y}^{B};\mathbf{X}^{B}|\mathbf{X}^{TS}) = I(\mathbf{Y}^{B};\mathbf{X}^{B}|\mathbf{X}^{TS},\mathbf{Y}^{TS})\,.$$

The sequential expansion of this expression (Medles and Slock, 2003b) leads to

$$I(\mathbf{Y}^{B};\mathbf{X}^{B}|\mathbf{X}^{TS},\mathbf{Y}^{TS}) = \sum_{i=1}^{N_B} I(\mathbf{y}_i;\mathbf{x}_i|\mathbf{X}^{TS},\mathbf{x}_1^{i-1},\overline{\mathbf{Y}}_i)\,,$$

where $\overline{\mathbf{Y}}_i = [\mathbf{Y}^{TS},\mathbf{y}_1^{i-1},\mathbf{y}_{i+1}^{N_B}]$ contains all the Rx signal apart from $\mathbf{y}_i$. From this expression, we conclude that an optimal way of processing is to use the past detected symbols as training, and the future (not detected) symbols as blind information for the CE, hence SB CE. This is consistent with the DFE canonical Rx concept: not only Rx but also CE is based on all Rx signal plus past symbols. Furthermore, in Medles and Slock (2003b), bounds on the MI are obtained, assuming white Gaussian inputs, corresponding to modeling CE error induced noise as an additive independent Gaussian noise. These bounds lead to a decrease in effective SNR ρ and ensuing MI loss as depicted in Fig. 14.1. Upper and lower bounds are obtained by considering CE based on different types of information:

	TS-type channel estimate based on past symbols (as TS)
$\le$	channel capacity without CSIR
$=$	semiblind channel estimate based on known past symbols and unknown Gaussian future symbols
$\le$	TS-type channel estimate based on past and future symbols

Asymptotic behavior of the capacity for block-fading channels

In this case, $\mathbf{H}_k = \mathbf{H}$ for $k = 1, \ldots, T$. The average MI is defined as

$$I_{\text{avg}}(T) = \frac{1}{T} I(\mathbf{Y}^B; \mathbf{X}^B | \mathbf{X}^{TS}, \mathbf{Y}^{TS}) = \frac{1}{T} \sum_{k=1}^{N_B} I(\mathbf{y}_k; \mathbf{x}_k | \mathbf{X}^{TS}, \mathbf{x}_1^{k-1}, \overline{\mathbf{Y}}_k). \tag{14.7}$$

For infinite blocks, we obtain the following limit (Medles and Slock, 2003b)

$$\lim_{T \to \infty} I_{\text{avg}}(T) = I(\mathbf{y}; \mathbf{x} | \mathbf{H}),$$

where $I(\mathbf{y}; \mathbf{x}|\mathbf{H})$ is the average MI with perfect CSIR. For block-fading channels, there is no loss in MI for large blocks when optimal SB CE algorithms are used.

Remark 1 As T grows, the use of detected data only to estimate the channel allows us to achieve asymptotically the average MI I_{avg} of the system. But for finite T, it is also necessary to use future undetected symbols to reach it.

Remark 2 The MI expression of (14.7) does not differentiate between training and past detected symbols. Then, for a fixed T, and when all the entries of $\mathbf{X}$ (training and data) are i.i.d., it is easy to see that the average MI I_{avg} of the system is maximized when the number of TS symbols N_{TS} is as small as possible (i.e., allows SB identifiability of the channel).

14.3.2 Other performance criteria

Once the Tx signal structure has been fixed (including the pilot structure), performance may be characterized in terms of the channel estimation quality, via, e.g., the Cramér–Rao bound (CRB), or in terms of the bit error rate (BER). A BER upper bound is obtained by considering the channel estimation error induced noise as additional additive independent Gaussian. This BER upper bound is, hence, described by this noise variance increase and ensuing shift in SNR. More discussion on performance criteria can be found in Tong *et al.* (2004).

14.4 Basic semiblind techniques

Basic refers here to the only nonblind information coming from time-multiplexed (TM) pilots, the channel being deterministic (block fading), the noise being spatio-temporally (ST) white Gaussian, the model for the unknown symbols being deterministic symbols (DSB) or i.i.d. Gaussian symbols (GSB). Consider the blind case first (DB/GB).

In the DB case, the compressed likelihood, after eliminating the unknown symbols via $\mathbf{x}_k = (\mathbf{H}^\dagger(q)\mathbf{H}(q))^{-1}\mathbf{H}^\dagger(q)\,\mathbf{y}_k$ (MMSE-ZF equalizer), is a function of $P_{\mathbf{H}(z)} = \mathbf{H}(z)(\mathbf{H}^\dagger(z)\mathbf{H}(z))^{-1}\mathbf{H}^\dagger(z)$, or the column space of $\mathbf{H}(z)$. Obviously, $N_r > N_t$ is required, and if $\mathbf{H}(z)$ is irreducible (no zeros) and column reduced, with columns ordered in nonincreasing length, then the relation between $\mathbf{H}(z)$ and a $\widehat{\mathbf{H}}(z)$ that can be deterministically identified is $\widehat{\mathbf{H}}(z) = \mathbf{H}(z)\mathbf{L}(z)$, where $\mathbf{L}(z)$ is block lower triangular, the diagonal blocks have dimensions commensurate with the groups of columns of $\mathbf{H}(z)$ that have identical length L_i (order $L_i{-}1$); the diagonal blocks are instantaneous mixtures, and the lower triangular entries (i,j) are FIR of order $L_j - L_i$. So, if all L_i are identical, then $\mathbf{L}$ is a square instantaneous mixture.

In the GB case, we get for the Rx spectrum $S_{\mathbf{yy}}(z) = \sigma_x^2\,\mathbf{H}(z)\,\mathbf{H}^\dagger(z) + S_{\mathbf{vv}}(z)$, with the usual assumption $S_{\mathbf{vv}}(z) = \sigma_v^2\,I_{N_r}$ (ST white), although $S_{\mathbf{vv}}(z) = S_{vv}(z)\,I_{N_r}$ (spatially white) with scalar $S_{vv}(z) = \lambda_{\min}(S_{\mathbf{yy}}(z))$ is sufficient with $N_r > N_t$. With columns of $\mathbf{H}(z)$ arranged in nonincreasing length, we get

$$\mathbf{H}(z) = (S_{\mathbf{yy}}(z) - S_{vv}(z)\,I_{N_r})^{\frac{1}{2}}\;\mathbf{L},$$

where $(\cdot)^{1/2}$ denotes a minimum-phase spectral factor (with columns in nonincreasing length), and $\mathbf{L}$ is block diagonal (same structure as the $\mathbf{L}(z)$ of the DB case) with unitary diagonal blocks. So, when all L_i are identical, the unidentifiable part $\mathbf{L}$ is a unitary instantaneous mixture. Although GB reduces the number of unidentifiable parameters, DB leads to consistency in SNR (zero error in absence of noise) for the DB-identifiable part (while GB does not for the GB extra identifiable part). Note that GB provides blind information even when $N_t \geq N_r$. Going from blind to semiblind, a proper design should introduce enough nonblind information to allow complete channel identifiability, regardless of the channel realization. The nonblind information thus required is much less in the GB case than in the DB case.

14.4.1 Frequency-flat MIMO channels

In this case, the channel impulse response in (14.1) is limited to $\mathbf{H} = \mathbf{H}_0$, and there is no intersymbol interference (ISI). The TS has a length of N_{TS} pilot symbol vectors $\mathbf{x}_k$, and the blind data part is composed of N_B data symbol vectors. The second-order statistics (SOS) of the Rx signal are given by

$$\widehat{\mathbf{R}} = \frac{1}{N_B}\sum_{k=1}^{N_B}\mathbf{y}_k\mathbf{y}_k^H = \mathbf{U}_e\mathbf{S}_e\mathbf{U}_e^H\,.$$

Using the singular value decomposition (SVD) of the channel

$$\mathbf{H} = \mathbf{UDQ} = \mathbf{WQ}\,,$$

it is easy to see that only $\mathbf{W}$ can be identified blindly (GB), whereas $\mathbf{Q}$ has to be identified using TS. However, this does not mean that estimation of $\mathbf{W}$ and $\mathbf{Q}$ is decoupled. The optimal technique is the Maximum Likelihood (ML) approach, involving the complete Rx signal. However, in Medles and Slock (2003a) it was shown that for $N_B \gg (\sigma_x^2/\sigma_v^2)\,N_{TS}$, ML can be simplified by first estimating $\mathbf{W}$ from $\widehat{\mathbf{R}}$ and then estimating $\mathbf{Q}$ from TS as follows:

(i) Estimation of $\mathbf{W}$:
- $\widehat{\mathbf{U}}$ corresponds to the p dominant eigenvectors in $\mathbf{U}_e$ (where $p = \min\{N_t, N_r\}$)
- $\widehat{\mathbf{D}}$ matches the p dominant eigenvalues of $(1/\sigma_x)\left(\lfloor \mathbf{S}_e - \sigma_v^2\mathbf{I}_{N_r}\rfloor_+\right)^{1/2}$
- $\widehat{\mathbf{W}} = \widehat{\mathbf{U}}\widehat{\mathbf{D}}$

(ii) Estimation of $\mathbf{Q}$:
$\widehat{\mathbf{Q}} = \mathbf{V}\,\mathbf{S}^H$, where $\mathbf{S}$ and $\mathbf{V}$ denote the unitary factors in the SVD of $\sum_{k=1}^{N_{TS}} \mathbf{x}_k^{TS}\mathbf{y}_k^{TS\,H}\mathbf{W} = \mathbf{S}\Sigma\mathbf{V}^H$, assuming $\sum_{k=1}^{N_{TS}} \mathbf{x}_k^{TS}\mathbf{x}_k^{TS\,H} \sim I_{N_t}$.

The evaluation of the performance shows that this technique achieves the GSB CRB for $N_B \gg (\sigma_x^2/\sigma_v^2)\,N_{TS}$; see Fig. 14.2 for an illustration, and Medles and Slock (2005) for more details.

14.4.2 Frequency-selective MIMO channels

Time domain approaches

In this case, there is ISI and nonorthogonality between pilot and data symbols, as depicted in Fig. 14.3 for a single burst of pilots. If we regroup all pilots in $\mathbf{X}_{TS}$, and hence $\mathbf{X}_U$ collects all unknown symbols, we can write $\mathcal{T}(\mathbf{h})\,\mathbf{X} = \mathcal{T}_K(\mathbf{h})\,\mathbf{X}_{TS} + \mathcal{T}_U(\mathbf{h})\,\mathbf{X}_U$, where $\mathcal{T}_K$ produces the channel output due to known TS. The Gaussian SB (GSB) ML criterion becomes

$$\ln\det(C_U) + (\mathbf{Y} - \mathcal{T}_K(\mathbf{h})\,\mathbf{X}_{TS})^H\,C_U^{-1}\,(\mathbf{Y} - \mathcal{T}_K(\mathbf{h})\,\mathbf{X}_{TS}),$$

where $C_U = \sigma_x^2\,\mathcal{T}_U(\mathbf{h})\,\mathcal{T}_U^H(\mathbf{h}) + C_{\mathbf{VV}}$. This GSB ML criterion often gets simplified to the following augmented TS (ATS) criterion (information from the mean only) (Medles *et al.*, 2001): $(\mathbf{Y}_{\rm TS} - \mathcal{T}_{\rm TS}(\mathbf{h})\mathbf{X}_{\rm TS})^H C_{\rm TS}^{-1}(\mathbf{Y}_{\rm TS} -$

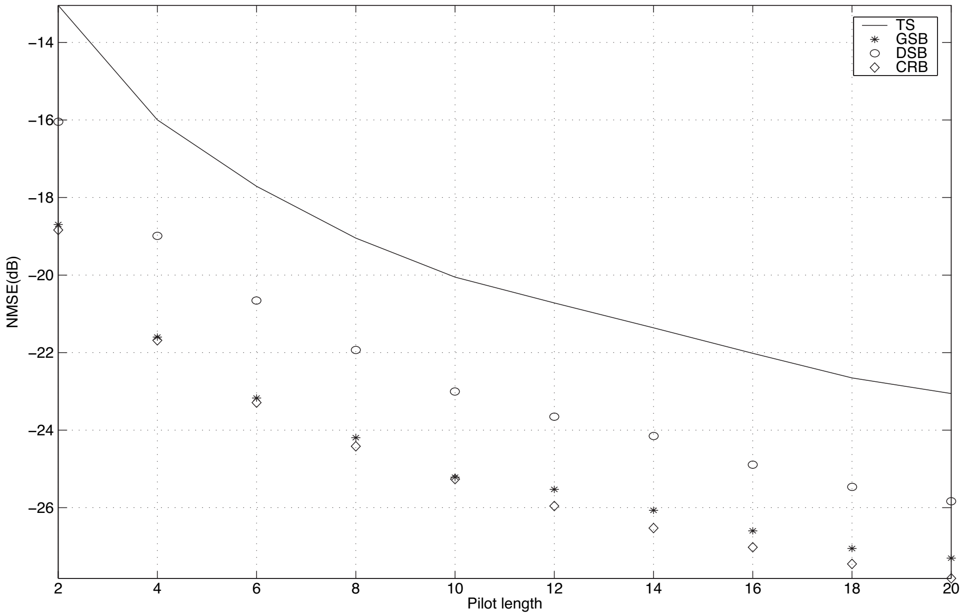

Fig. 14.2. Normalized channel estimation MSE vs. pilot length N_{TS}: frequency-flat channel, $N_t = 2$, $N_r = 4$, $N_B = 400$, SNR = 10 dB.

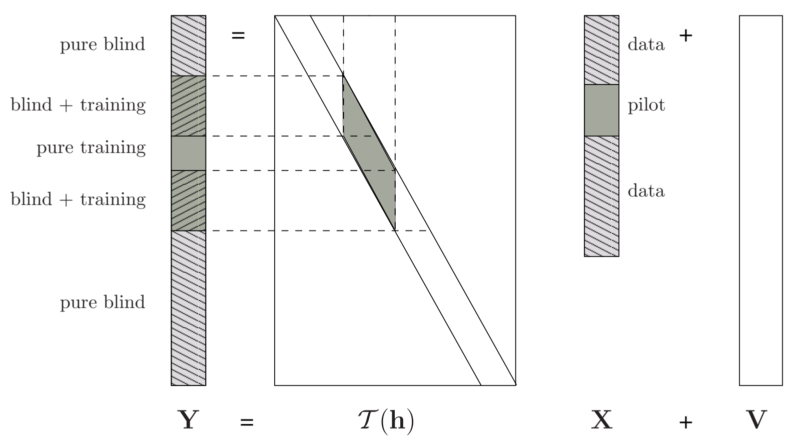

Fig. 14.3. Received signal structure for a frequency-selective channel.

$\mathcal{T}_{\mathrm{TS}}(\mathbf{h})\mathbf{X}_{\mathrm{TS}})$, where $\mathbf{Y}_{\mathrm{TS}}$ is the part of $\mathbf{Y}$ containing at least one pilot symbol. ATS can be solved iteratively as a weighted LS problem. See Pladdy *et al.* (2004) for a simplified ATS algorithmic approach, whereas in Rousseaux and Leus (2004) the interference from unknown data on pilots is canceled blindly in an iterative approach. In Rousseaux *et al.* (2003b), ATS is treated as a simplified ML criterion for the case in which $C_{\mathbf{VV}}$ is considered totally unknown (and unstructured), in which case C_{TS} is built from a sample covariance matrix. In Rousseaux *et al.* (2003a), distributed training bursts are treated. Though a significant improvement over TS, ATS only makes limited use of the blind information. Although the GSB ML problem is meaningful, and becomes straightforward to solve when the pilots are isolated, in some unpublished work we have found that grouped pilots lead to better performance here as well (when the channel does not vary fast).

To exploit the blind information, a general technique is to use a weighted least-squares (WLS) combination of TS and blind information:

$$\min_{\widehat{\mathbf{h}}}\left\{\left\|\mathbf{Y}_{\mathrm{TS}}-\mathcal{X}_{\mathrm{TS}}\,\widehat{\mathbf{h}}\right\|^2_{C^{-1}_{\mathrm{TS}}(\widehat{\mathbf{h}})}+\left\|\mathcal{B}\widehat{\mathbf{h}}\right\|^2_{C^{\#}(\widehat{\mathbf{h}})}\right\}, \tag{14.8}$$

where $\|\mathbf{Y}\|^2_C=\mathbf{Y}^HC\mathbf{Y}$, $\mathcal{X}_{\mathrm{TS}}\mathbf{h}=\mathcal{T}_{\mathrm{TS}}(\mathbf{h})\mathbf{X}_{\mathrm{TS}}$, $C(\mathbf{h})=\mathrm{E}\big[(\mathcal{B}\mathbf{h})(\mathcal{B}\mathbf{h})^H\big]$, and correlation between blind and TS parts is neglected. $\mathcal{B}$ is a matrix that (soft-) constrains the channel and expresses the blind information by parameterizing the noise subspace, or capturing the whiteness of the transmitted signal. $\mathcal{B}$ can be appropriately parameterized in terms of the prediction error filter $\mathbf{P}_K(q)=\mathbf{I}+\sum_{i=1}^{K-1}\mathbf{P}_iq^{-i}$ for the noise-free signal SOS: $\mathbf{P}_K(q)\mathbf{H}(q)=\mathbf{H}_0$. For $N_r\geq N_t$, the channel predictor generically exists and is FIR for $N_r>N_t$ with $K\geq\lceil(L-N_t)/(N_r-N_t)\rceil$; it can be evaluated from the Rx signal SOS $\widehat{\mathbf{R}}_{\mathbf{yy}}(k)$, $k=0,\ldots,K$, and leads to a parameterization of the channel: $\mathbf{H}(z)=\mathbf{P}^{-1}(z)\mathbf{W}\mathbf{Q}$, $(S_{\mathbf{yy}}(z)-\sigma_v^2\,I_{N_r})\mathbf{P}^{\dagger}(z)=\mathbf{H}(z)\sigma_x^2\mathbf{Q}^H\mathbf{W}^H$, and $\mathbf{W}$ is obtained from $\mathbf{P}(z)(S_{\mathbf{yy}}(z)-\sigma_v^2I_{N_r})\mathbf{P}^{\dagger}(z)=\sigma_x^2\mathbf{W}\mathbf{W}^H$. However, unlike the flat channel case, there is no simple ML variation to estimate $\mathbf{Q}$. If we reduce the exploitation of $\mathbf{P}(q)\mathbf{H}_k=\mathbf{H}_0\delta_{k0}$ to $\mathbf{W}^{\perp H}\mathbf{H}_0=0$ and $\mathbf{P}(q)\mathbf{H}_k=0$, $k>0$, and combine it with the TS part in a WLS approach, then the result is the quadratic criterion of (14.8). Further results on this approach can be found in Medles *et al.* (2001) and Medles and Slock (2005).

It is interesting to mention that the more conventional subspace-based techniques only exploit the set of equations $\mathbf{W}^{\perp H}\mathbf{P}(q)\mathbf{H}_k=0$, which corresponds to the (DB) information in the noise subspace (here parameterized by $\mathbf{W}^{\perp H}\mathbf{P}(q)$). The full criterion enhances the performance by further exploiting the temporal whiteness of the Tx symbols (GB), which is characterized by $\mathbf{P}(q)\mathbf{H}_k=0$, $k>0$. In Fig. 14.4, an illustration is given of the performance of basic TS and ATS, and DSB and GSB with basic TS or ATS (DSBA, GSBA). The number of pilots leads to unidentifiability for TS, ATS or DSB, and identifiability for GSB, DSBA and GSBA. This example illustrates the usefulness of blind information and the extra information present in GSB (be it not consistent in SNR).

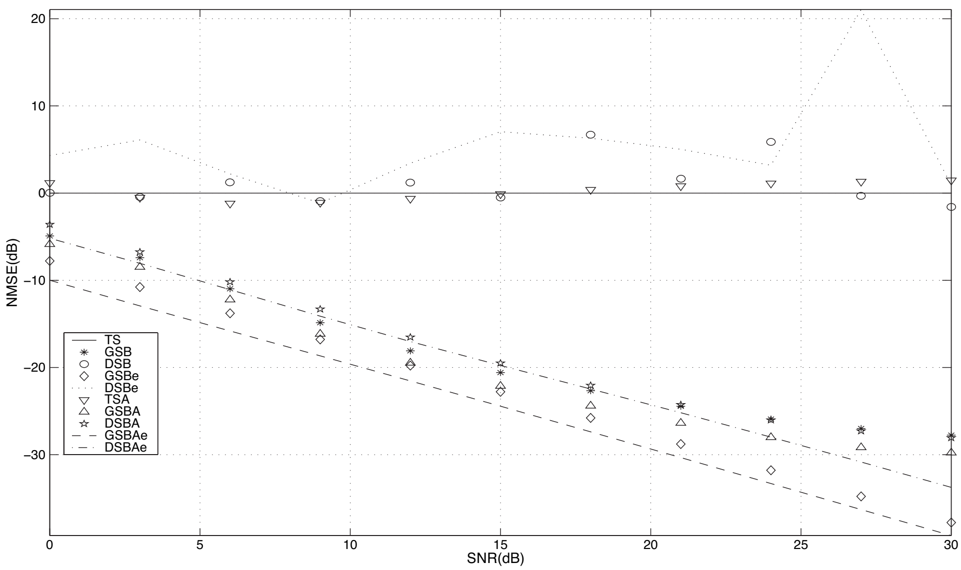

Fig. 14.4. Normalized channel estimation MSE vs. SNR: frequency-selective channel, $N_t = 2$, $N_r = 4$, channel lengths $[3, 1]$, $N_B = 300$, $N_{\mathrm{TS}} = 5$.

Frequency domain approaches

Note that time-domain-multiplexed pilots become embedded pilots (see further) in the frequency domain, and vice versa. In the CP case, we get for the SOS from (14.4)

$$\mathbf{R}_{\boldsymbol{YY}} = \sigma_x^2 \mathcal{H}\mathcal{H}^H + \sigma_v^2 I_{N_r N},$$

which shows the decoupling between tones, or from (14.5) per tone

$$\mathbf{R}_n = \mathbf{R}_{\boldsymbol{y}_n \boldsymbol{y}_n} = \sigma_x^2 \boldsymbol{H}_n \boldsymbol{H}_n^H + \sigma_{v_n}^2 I_{N_r} = V_{\mathcal{S},n} \Lambda_{\mathcal{S},n} V_{\mathcal{S},n}^H + \sigma_{v_n}^2 \, V_{\mathcal{N},n} V_{\mathcal{N},n}^H,$$

where $\sigma_{v_n}^2$ can vary between tones for a temporally colored but spatially white noise (which hence becomes straightforward to handle, approximating or assuming $\mathbf{R}_{\boldsymbol{VV}}$ to be block circulant). The coupling between tones comes from the FIR channel response. Let $\boldsymbol{h}_n = \mathrm{vec}(\boldsymbol{H}_n)$; then, the FFT relation leads to $\boldsymbol{h}_n = \boldsymbol{G}_n \mathbf{h}$ for some matrix $\boldsymbol{G}_n$. Now, if at tone n we have a cost function of the form $\boldsymbol{h}_n^H \boldsymbol{Q}_k \boldsymbol{h}_n$, then this induces a cost function for $\mathbf{h}$ of the form $\mathbf{h}^H \big[\sum_{n=0}^{N-1} \boldsymbol{G}_n^H \boldsymbol{Q}_n \boldsymbol{G}_n\big] \mathbf{h}$, and similarly for Fisher information matrices. So, one can just concentrate on the cost function for a given tone. For instance, if $\widehat{\mathbf{R}}_n = \widehat{\mathrm{E}}[\boldsymbol{u}_n \boldsymbol{u}_n^H] = \widehat{V}_{\mathcal{S},n} \widehat{\Lambda}_{\mathcal{S},n} \widehat{V}_{\mathcal{S},n}^H + \widehat{V}_{\mathcal{N},n} \widehat{\Lambda}_{\mathcal{N},n} \widehat{V}_{\mathcal{N},n}^H$, then we get the signal subspace fitting cost function: $\min_{\mathbf{h}} \sum_{n=0}^{N-1} \|\boldsymbol{H}_n^H \widehat{V}_{\mathcal{N},n}\|_F^2$, requiring only SVDs of small matrices.

Due to the decoupling between tones, one may also envisage the introduction of optimal weighting: $\max_{\mathbf{h}} \sum_{n=0}^{N-1} \mathrm{tr}\{P_{\boldsymbol{H}_n} \widehat{V}_{\mathcal{S},n} \widetilde{\widehat{\Lambda}}^2_{\mathcal{S},n} \widehat{\Lambda}^{-1}_{\mathcal{S},n} \widehat{V}^H_{\mathcal{S},n}\}$. More discussion can be found in Slock (2004b). See also Liu *et al.* (2001) for a per-tone application of the Constant Modulus Algorithm (CMA). In Zeng and Ng (2004), a subspace method (and semiblind version) is proposed based on zero padding, which leads to more robust channel estimates (compared with CP), but is computationally complex (SVDs are not per tone but of OFDM symbol size). A simpler version is discussed in Slock (2004b).

14.5 Bayesian semiblind (BSB) channel estimation

BSB CE has been introduced in Slock (2004a), where more details can be found. Separation property: BSB can be organized as basic SB followed by Bayesian filtering, especially if enough training for basic SB is available to lead to full identifiability. The basic semiblind CE leads to a measurement equation $\widehat{\mathbf{h}}_k = \mathbf{h}_k + \widetilde{\mathbf{h}}_k$, with $\widetilde{\mathbf{h}}_k \sim \mathcal{CN}(b(\mathbf{h}_k), C(\mathbf{h}_k))$, temporally decorrelated. The measurement equation then gets combined with the prior channel model (e.g., Rayleigh fading) in a Bayesian filtering operation. In the TS case, $b \equiv 0$, $C = \sigma^2_{\tilde{h}} I$, leading to Wiener filtering/smoothing. In the SB case, $b(\mathbf{h}_k), C(\mathbf{h}_k) \longrightarrow b(\widehat{\mathbf{h}}_k), C(\widehat{\mathbf{h}}_k)$, the measurement equation being possibly time-varying, Kalman filtering/smoothing may be required. The full Bayesian approach requires joint estimation of Bayesian channel and deterministic prior hyper-parameters, which may be done with a variety of approaches, including EM. The appropriate CRB here is for joint deterministic slow and Bayesian fast parameters. See also Hassan *et al.* (2004) (and references therein) for a combination of Kalman filtering and CMA, and Haykin *et al.* (2004) (and references therein) for an overview on the use of Kalman and particle filtering. Whereas most applications of Wiener/Kalman filtering to channel tracking proposed in the literature assume the state space model to be known, the estimation of the channel variation statistics is incorporated in Lenardi and Slock (2002), Montalbano and Slock (2003).

14.6 Other forms of side information

With perfect CSIR (and no CSIT, and i.i.d. channel elements), the optimal input signal is zero-mean ST white Gaussian noise. Any deviation from this (side information) will lower the perfect CSIR channel capacity. But of course, there usually is no CSIR, so any such deviation may allow channel estimation, hence leading to an increase in actual channel capacity (see

Zheng and Tse (2002) for optimal input distributions in the absence of CSIR). Possible forms of side information are:

- higher-order statistics of data symbols (Cardoso, 1998; Cichocki and Amari, 2002);
- finite alphabet (FA) of unknown symbols, exploited through iterative channel estimation and data detection, see, e.g., Talwar and Paulraj (1997), Li and Yang (2003), Zhu *et al.* (2003), Scaglione and Vosoughi (2004), Souza *et al.* (2004), or Yue *et al.* (2004) with two-level Kalman filtering. In Sadler *et al.* (2001), it is shown that when constraints such as FA constraints on the symbols only leave a discrete ambiguity, then the CRB (which is a local bound) for channel estimation is the same as for the case when the unknown symbols were known;
- channel coding in unknown symbols, exploited through turbo detection and estimation; in Scherb *et al.* (2004) a channel estimation CRB is provided when data symbol channel coding is exploited, involving the minimum distance amplification introduced by the channel code. As SNR increases from low to high values, this CRB moves from the case of the data symbols being unknown Gaussian to being known as a TS;
- partial FA knowledge: constant modulus (8-PSK in EDGE) (Safavi and Abed-Meraim, 2003; Hassan *et al.*, 2004; Liu *et al.*, 2001);
- some training/pilot symbols, only enough to allow iterative joint data detection/channel estimation to converge;
- symbol modulus variation pattern (a particular form of Tx-induced cyclostationarity); some of the techniques proposed here lead to wide-sense cyclostationarity (Tsatsanis and Giannakis, 1997), without consistency in SNR. The technique proposed in Leus *et al.* (2001), though, is deterministic;
- space-time coding redundancies through reduced rate linear precoding, introducing subspaces in the transmitted signal covariance, e.g., Alamouti or other orthogonal ST coding schemes, or Choi (2004) and Liu *et al.* (2001);
- guard intervals in time or frequency, see Scaglione *et al.* (1999), Zeng and Ng (2004), cyclic prefix structure;
- symbol stream color (see below);
- known pulse shape (see below);
- CDMA spreading code(s) (see below);
- Tx induced nonzero mean (see superimposed pilots below).

Spatial multiplexing schemes that achieve the optimal rate-diversity trade-off (see Chapter 8) typically do not introduce any blind information (other than GB) for the channel estimation. In Medles and Slock (2004), for instance, a previously introduced linear prefiltering scheme was shown to attain this optimal trade-off. Since the prefilter is a MIMO allpass filter, it leaves the white vector input white. However, perturbations of optimal trade-off achieving schemes can be derived that introduce side information (see Section 14.6.1). So, questions that so far are only partially answered are: what is the optimal amount of side information to maximize capacity, as more side information reduces capacity with CSIR, but also reduces channel estimation error and hence increases capacity? More importantly, what is the optimal distribution of side information over the various forms? Note that only DB and GB are, strictly speaking, blind approaches. The exploitation of any form of side information mentioned above should be called a semiblind approach.

14.6.1 Coloring linear precoding

In Hua and Tugnait (2000), it is shown that colored inputs can be separated if their spectra are linearly independent. Correlation can be introduced by linear convolutive precoding, which corresponds to MIMO prefiltering of $\mathbf{x}_k$ with a MIMO prefilter $\mathbf{T}(z)$, such that the Tx vector signal becomes $\mathbf{a}_k = \mathbf{T}(q)\,\mathbf{x}_k$. We consider full rate linear precoding, so that $\mathbf{T}(z)$ is square $(N_t \times N_t)$ (in Leus *et al.* (2001), an example of low rate precoding appears since the same symbol sequence gets distributed over all Tx antennas). We get for the Rx signal spectrum $S_{\mathbf{yy}}(z) = \sigma_x^2\,\mathbf{H}(z)\,\mathbf{T}(z)\,\mathbf{T}^\dagger(z)\,\mathbf{H}^\dagger(z) + \sigma_v^2\,I_m$. An appropriate $\mathbf{T}(z)$ may reduce the nonidentifiability to a phase factor per source, or even to a global phase factor. Consider a generic reducible channel that can be factored as $\mathbf{H}(z) = \mathbf{G}(z)\mathbf{C}(z)$, where $\mathbf{G}(z)$ is irreducible and column reduced with columns in order of., e.g., nonincreasing degree. If r is the (generic) rank of $\mathbf{H}(z)$, then $\mathbf{G}(z)$ is $N_r \times r$, whereas $\mathbf{C}(z)$ is $r \times N_t$. For $r \leq N_r - 1$, we can DB identify $\mathbf{G}(z)$. $\mathbf{G}(z)$ is unique up to a factor $\mathbf{L}(z)$ mentioned earlier. For whichever $\mathbf{G}(z)$ in this equivalence class, it remains to identify $\mathbf{C}(z)$ from

$$\mathbf{S}(z) = \mathbf{G}^{\#}(z)\,(S_{\mathbf{yy}}(z) - \sigma_v^2 I_m)\mathbf{G}^{\#\dagger}(z) = \mathbf{C}(z)\mathbf{S}_{\mathbf{aa}}(z)\mathbf{C}^\dagger(z), \qquad (14.9)$$

where $\mathbf{G}^{\#}(z)$ is a MMSE-ZF equalizer for $\mathbf{G}(z)$: $\mathbf{G}^{\#}(z)\mathbf{G}(z) = I_r$. The degree of $\mathbf{C}_j(z)$ is unpredictable and can be up to $L_j - 1$, the degree of the corresponding column $\mathbf{h}_j(z)$ of $\mathbf{H}(z)$. Two scenarios may be distinguished:

Noncooperative scenario

This scenario typically corresponds to the multiuser case (on the Tx side) without cooperation between users. Consider the simple case with users with a single Tx antenna. In this scenario, $\mathbf{H}(z)$ has no structure other than possibly being FIR, and $\mathbf{T}(z)$ and $S_{\mathbf{aa}}(z)$ are diagonal. This scenario has been considered in Abed-Meraim *et al.* (2001), Hua and Xiang (2001), Xavier *et al.* (2001). In Medles and Slock (2002), two approaches have been proposed for the identification of $\mathbf{C}(z)$ from (14.9).

Frequency-domain approach The idea here is to introduce zeros into the diagonal elements of $\mathbf{T}(z)$, or hence $\mathbf{S}_{\mathbf{aa}}(z)$, such that all other elements other than diagonal element j share N_j zeros

$$\mathbf{T}_{jj}(z) = \prod_{i=1,\neq j}^{p} \prod_{k=1}^{L_i} (1 - z_{i,k} z^{-1}).$$

This allows identifiability of $\mathbf{C}_j(z)$ from $\mathbf{S}(z)$ up to a phase, since

$$\mathbf{S}(z_{j,k}) = \mathbf{C}_j(z_{j,k}) \mathbf{S}_{a_j a_j}(z_{j,k}) \mathbf{C}_j^{\dagger}(z_{j,k}), \qquad k = 1, \ldots, L_j,$$

where $\mathbf{S}_{a_j a_j}(z) = \sigma_x^2 T_{jj}(z) T_{jj}^{\dagger}(z)$.

Time-domain approach The idea here is to introduce delay in the pre-filter, so that the correlations of each $\mathbf{C}_j(z)$ appear separately in certain delay portions of the correlation sequence of $\mathbf{S}(z)$. This can be obtained, for instance, with

$$\mathbf{T}_{jj}(z) = 1 - \alpha_j z^{-d_j}, \qquad d_j = \sum_{i=1}^{j-1} L_i.$$

Identification can be done with a correlation sequence peeling approach that starts with the last column $\mathbf{C}_{N_t}(z)$, of which the (single-sided) correlation sequence appears in an isolated fashion in the last L_{N_t} correlations of $\mathbf{S}(z)$. Identification of $\mathbf{C}_{N_t}(z)$ from its correlation sequence can be done up to a phase factor $e^{j\theta_{N_t}}$ (and up to the phase of zeros if $\mathbf{C}_{N_t}(z)$ has zeros). We can then subtract $\mathbf{S}_{a_j a_j}(z) \mathbf{C}_{N_t}(z) \mathbf{C}_{N_t}^{\dagger}(z)$ (which does not require $\mathbf{C}_{N_t}(z)$, but only its correlation sequence) from $\mathbf{S}(z)$, which will then reveal the correlation sequence of $\mathbf{C}_{N_t-1}(z)$ in its last L_{N_t-1} correlations, etc. The degree of $\mathbf{S}_{\mathbf{aa}}(z)$ is in this case the degree d_{N_t} of $\mathbf{S}_{a_{N_t} a_{N_t}}(z)$, which, in the case of all equal L_j, is again $(N_t{-}1)L_1$, which leads to a degree of $N_t L_1{-}1$ for $\mathbf{S}(z)$, or hence $N_t L_1$ correlations. Such a degree for $\mathbf{S}_{\mathbf{aa}}(z)$ is not only sufficient but also necessary, since when $r = 1$, there are $N_t L_1$ parameters to be identified, for which, indeed, at least $N_t L_1$ correlations are needed.

Cooperative scenario

This is the single-user spatial multiplexing scenario. $\mathbf{S}_{\mathbf{aa}}(z)$ is allowed to be nondiagonal. Noncooperative approaches can, of course, also be applied here. However, that would lead to at least an unknown phase per Tx antenna, and hence requires either differential encoding or training symbols per Tx antenna. By applying full prefiltering, such that $\mathbf{S}_{\mathbf{aa}}(z)$ is not block diagonal (in which case it is said to be fully diverse), $\mathbf{H}(z)$ can be identified up to a global phase factor only under certain conditions on $\mathbf{S}_{\mathbf{aa}}(z)$ (see Medles and Slock (2002)). Since this case results in better identifiability, better estimation quality may be another consequence.

Precoder optimization

We consider here the optimization of the ergodic capacity w.r.t. the precoder. As discussed earlier, the optimized prefilter is the result of a compromise, and is expected to be a perturbation of a paraunitary filter, transforming the white $\mathbf{x}_k$ into slightly colored $\mathbf{a}_k$, allowing channel identification. An example of this optimization for a frequency-flat channel is provided in Medles and Slock (2002).

Oversampling, known pulse shapes, and CDMA

So far, the multitude of outputs was assumed to stem from a multitude of sensors. Another output dimension may be added by oversampling the output w.r.t. the discrete-time input. If now also the Tx (and Rx) pulse shape is known, then it can be represented as an $N_{\mathrm{os}} \times 1$ vector prefilter $\mathbf{T}(z)$ per input, with N_{os} the oversampling factor. The channel $\mathbf{h}_i(z)$ for input i now also becomes a so-called pseudocirculant $N_r N_{\mathrm{os}} \times N_{\mathrm{os}}$ matrix filter. A known pulse shape is treated in, e.g., Ghauri and Slock (2000). The knowledge of the pulse shape helps to improve the channel estimation accuracy by reducing the remaining delay spread and capturing the ill-conditioning in time (tapering) and frequency domain (limited bandwidth). But it will not help in resolving inputs if they all use the same pulse shape.

Direct sequence spectrum spreading (or DS-CDMA) is a special case, in which the oversampling factor corresponds to the spreading factor. A sample is called a chip, and the column prefilter $\mathbf{T}(z)$ is static, corresponding to an instantaneous multiplication with the spreading code (which can be time-varying in the case of long/aperiodic/pseudorandom codes, or time-invariant as in the case of short/periodic/deterministic codes). Of course, CDMA can be combined with oversampling w.r.t. the chip rate, and exploitation of a chip pulse shape. The use of different spreading codes for different inputs allows for fairly robust blind source separation and channel estimation, see Ghauri and Slock (1998, 1999), and also Hochwald *et al.* (2001) and Liu *et al.* (2001). In Sung and Tong (2004), long-code CDMA and fast fading channels are considered.

14.7 Pilot structure optimization

Most existing work on pilot structure optimization considers channel estimation based on training only, see Chapter 17. Basic work on TS based MIMO CE appears in Hassibi and Hochwald (2003). In Barhumi *et al.* (2003), Ma *et al.* (2003), and Yang *et al.* (2004), TS-based CE in doubly selective MIMO OFDM systems is considered. See Dong *et al.* (2004) for Bayesian pilot based estimation of frequency-flat AR single-input single-output (SISO) channels, and Tong *et al.* (2004) for a tutorial. A (not so) recent twist on the training paradigm is, besides the usual time-multiplexed (TM) pilots, the appearance of superimposed pilots (SI, also called embedded). SI pilots are actually classical in CDMA standards, which use a pilot signal, sometimes combined with TM pilots. In Zhu *et al.* (2003), SI-pilot-based channel estimates are used to initialize an iterative receiver. In Berriche *et al.* (2004), optimization of a mixture of TM and SI pilots is considered. The continuous SI pilots actually form a pilot signal, and their large duration leads to quasi-orthogonality with the data. It is found that for large enough and equivalent pilot power, both pilot forms lead to similar performance. Only the channel estimation (CRB) is considered, though, as a performance indicator. In Vosoughi and Scaglione (2004b), the effect of both types of pilots on the throughput is considered, and TM pilots appear to be favored. Indeed, pilots not only allow channel estimation, but also influence the data detection. The presence of TM pilots leads to reduced ISI in frequency-selective channels with time-domain Tx. Semiblind channel estimation and detection with SI pilots is considered in Meng and Tugnait (2004). So, an important question here is: is orthogonality of pilots and data desirable? The answer may depend on how mixed information (pilot/data) is used.

14.8 Other research avenues ahead

We already mentioned the optimization of the side information mix.

Multiuser case In this case, the number of unknowns per received sample increases further. Whereas spatial multiplexing is the cooperative case of multiple-input, the multiuser case corresponds to the noncooperative version. Differentiation of users at the level of SOS can be obtained through coloring (e.g., CDMA) as mentioned earlier. In Zeng and Ng (2004), a semiblind multiuser scenario is considered.

Noncoherent approaches See also Chapter 10. In the comparison between noncoherent approaches (no CSIR) and coherent approaches based on channel estimates, the current trend in improving noncoherent approaches involves exploiting the Doppler structure (predictability) of the channel. This may be one indication that coherent approaches based on channel estimation should work better.

Semiblind direct receiver estimation Training optimization may depend on the receiver architecture (Vosoughi and Scaglione, 2004a). See Scaglione *et al.* (1999) for the blind determination of linear equalizers and Bugallo *et al.* (2002) for the semiblind determination of a MIMO DFE.

Channel estimation for the transmitter The availability of channel state information at the Tx (CSIT) allows us to improve transmission through adaptive modulation, see, e.g., Xia *et al.* (2004) and Chapter 6. Questions that arise here involve not only channel estimation, but also its possible quantization and (digital or analog) retransmission. A key issue here is also the degree of reciprocity of the channel, or, e.g., its pathwise parameters (direction, delay, Doppler shift, power). Another issue is the effect of sensor array design on channel estimation and reciprocity, e.g., beamspace (beam selection should be reciprocal).

Description of channel variation in terms of user mobility Such an approach would possibly allow a more compact description of temporal variation, lead to better channel predictability, and allow the separation of users with less side information and mobile localization applications (Amar and Weiss, 2004). See Bug and Jakoby (2004) for an approach in this direction. Variability of the environment also needs to be taken into account, however.

References

Abed-Meraim, K., Xiang, Y., Manton, J. H., and Hua, Y. (2001). Blind source-separation using second-order cyclostationary statistics. *IEEE Trans. Signal Process.*, **49** (4), 694–701.

Amar, A. and Weiss, A. J. (2004). Direct position determination of narrowband radio transmitters. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, 81–84.

Barhumi, I., Leus, G., and Moonen, M. (2003). Optimal training design for MIMO OFDM systems in mobile wireless channels. *IEEE Trans. Signal Process.*, **51** (6), 1615–1624.

Berriche, L., Abed-Meraim, K., and Belfiore, J.-C. (2004). Cramer-Rao bounds for MIMO channel estimation. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, 397–400.

Bug, S. and Jakoby, R. (2004). Modeling of the mobile radio channel using theory of dynamics—first derivations and results. In *Proc. Eur. Wireless Conf.*

Bugallo, M. F., Miguez, J., and Castedo, L. (2002). Decision-feedback semiblind channel equalization in space-time coded systems. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, 2425–2428.

Cardoso, J.-F. (1998). Blind signal separation: Statistical principles. *Proc. IEEE*, **86** (10), 2009–2025. Special Issue on Blind Identification and Estimation.

de Carvalho, E. and Slock, D. T. M. (2001). Semi-blind methods for FIR multichannel estimation. In Giannakis, G. B., Hua, Y., Stoica, P., and Tong, L. (eds.), *Signal Processing Advances in Wireless & Mobile Communications* (Prentice Hall).

Choi, J. (2004). Equalization and semi-blind channel estimation for space-time block coded signals over a frequency-selective fading channel. *IEEE Trans. Signal Process.*, **52** (3), 774–785.

Cichocki, A. and Amari, S. (2002). *Adaptive Blind Signal and Image Processing, Learning Algorithms, and Applications* (Wiley).

Dong, M., Tong, L., and Sadler, B. (2004). Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels. *IEEE Trans. Signal Process.*, **52** (5), 1403–1418.

Ghauri, I. and Slock, D. T. M. (1998). Blind and semi-blind single user receiver techniques for asynchronous CDMA in multipath channels. In *Proc. IEEE Global Telecommun. Conf.*, vol. 6, 3572–3577.

—— (1999). MMSE-ZF receiver and blind adaptation for multirate CDMA. In *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, vol. 1, 628–632.

—— (2000). Structured estimation of sparse channels in quasi-synchronous DS-CDMA. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, 2873–2876.

Hassan, M. A. S., Sharif, B. S., Woo, W. L., and Jimaa, S. (2004). Semiblind estimation of time varying STFBC-OFDM channels using Kalman filter and CMA. In *Proc. IEEE Int. Symp. Comp. Commun.*, vol. 2, 594–599.

Hassibi, B. and Hochwald, B. M. (2003). How much training is needed in multiple-antenna wireless links? *IEEE Trans. Inf. Theory*, **49** (10), 951–964.

Haykin, S., Huber, K., and Chen, Z. (2004). Bayesian sequential state estimation for MIMO wireless communications. *Proc. IEEE*, **92** (3), 439–454.

Hochwald, B., Marzetta, T. L., and Papadias, C. B. (2001). A transmitter diversity scheme for wideband CDMA systems based on space-time spreading. *IEEE J. Sel. Areas Commun.*, **19** (1), 48–60.

Hua, Y. and Tugnait, J. K. (2000). Blind identifiability of FIR-MIMO systems with colored input using second-order statistics. *IEEE Signal Process. Lett.*, **7** (12), 348–350.

Hua, Y. and Xiang, Y. (2001). The BID for blind equalization of FIR MIMO channels. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 1, 415–424.

Lenardi, M. and Slock, D. T. M. (2002). Estimation of time-varying wireless channels and application to the UMTS W-CDMA FDD downlink. In *Proc. Eur. Wireless Conf.*, 204–207.

Leus, G., Vandaele, P., and Moonen, M. (2001). Deterministic blind modulation-induced source separation for digital wireless communications. *IEEE Trans. Signal Process.*, **49** (1), 219–227.

Li, Y. and Yang, L. (2003). Semi-blind MIMO channel identification based on error adjustment. In *Proc. IEEE Conf. Neural Netw. Signal Process.*, vol. 2, 1429–1432.

Liu, Z., Giannakis, G. B., Barbarossa, S., and Scaglione, A. (2001). Transmit-antennae space-time block coding for generalized OFDM in the presence of unknown multipath. *IEEE J. Sel. Areas Commun.*, **19** (7), 1352–1364.

Ma, X., Giannakis, G. B., and Ohno, S. (2003). Optimal training for block transmission over doubly selective wireless fading channels. *IEEE Trans. Signal Process.*, **51** (5), 1351–1366.

Medles, A., de Carvalho, E., and Slock, D. T. M. (2001). Linear prediction based semi-blind estimation of MIMO FIR channels. In *Proc. IEEE Workshop Signal Process. Advances Wireless Commun.*, 58–61.

Medles, A. and Slock, D. (2004). Achieving the optimal diversity-vs-multiplexing tradeoff for MIMO flat channels with QAM space-time spreading and DFE equalization. *IEEE Trans. Inf. Theory.* Submitted.

—— (2005). Semiblind estimation of FIR MIMO channels in single-carrier systems. *IEEE Trans. Signal Process.* Submitted.

Medles, A. and Slock, D. T. M. (2002). Linear precoding for spatial multiplexing MIMO systems: Blind channel estimation aspects. In *Proc. IEEE Int. Conf. Commun.*, vol. 1, 401–405.

—— (2003a). Augmenting the training sequence part in semiblind estimation for MIMO channels. In *Proc. 37th Asilomar Conf. Signals, Syst., Comput.*, vol. 2, 1825–1829.

—— (2003b). Mutual information without channel knowledge at the receiver. In *Proc. IEEE Workshop Signal Process. Advances Wireless Commun.*, 398–402.

Meng, X. and Tugnait, J. K. (2004). Semi-blind channel estimation and detection using superimposed training. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, 417–420.

Montalbano, G. and Slock, D. T. M. (2003). Joint common-dedicated pilots based estimation of time-varying channels for W-CDMA receivers. In *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, vol. 2, 1253–1257.

Pladdy, C., Özen, S., Nerayanuru, S. M., Fimoff, M. J., and Zoltowski, M. (2004). Taylor series approximation of semi-blind best linear unbiased channel estimates for the general linear model. In *Proc. Asilomar Conf. Signals, Syst., Comput.*, vol. 2, 2208–2212.

Rousseaux, O. and Leus, G. (2004). An iterative method for improved training-based estimation of doubly selective channels. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, 889–892.

Rousseaux, O., Leus, G., Stoica, P., and Moonen, M. (2003a). Generalized training based channel identification. In *Proc. IEEE Global Telecommun. Conf.*, vol. 5, 2432–2436.

—— (2003b). Training based maximum likelihood channel identification. In *Proc.*

IEEE Workshop Signal Process. Advances Wireless Commun., 334–338.

Sadler, B. M., Kozik, M. J., and Moore, T. (2001). Bounds on bearing and symbol estimation with side information. *IEEE Trans. Signal Process.*, **49** (4), 822–834.

Safavi, A. and Abed-Meraim, K. (2003). Blind channel identification robust to order overestimation: a constant modulus approach. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, 1248–1251.

Scaglione, A., Giannakis, G. B., and Barbarossa, S. (1999). Redundant filterbank precoders and equalizers part II: Blind channel estimation, synchronization, and direct equalization. *IEEE Trans. Signal Process.*, **47** (7), 2007–2022.

Scaglione, A. and Vosoughi, A. (2004). Turbo estimation of channel and symbols in precoded MIMO systems. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, 413–416.

Scherb, A., Kühn, V., and Kammeyer, K.-D. (2004). Cramer-Rao lower bound for semiblind channel estimation with respect to coded and uncoded finite-alphabet signals. In *Proc. Asilomar Conf. Signals, Syst., Comput.*, vol. 2, 2193–2197.

Slock, D. T. M. (2004a). Bayesian blind and semiblind channel estimation. In *Proc. IEEE Workshop Sensor, Array, Multich. Signal Process.*

—— (2004b). Blind FIR channel estimation in multichannel cyclic prefix systems. In *Proc. IEEE Workshop Sensor, Array, Multich. Signal Process.*

Souza, R. D., Garcia-Frias, J., and Haimovich, A. M. (2004). A semi-blind receiver for iterative data detection and decoding of space-time coded data. In *Proc. IEEE Wireless Commun. Netw. Conf.*, vol. 3, 1626–1630.

Sung, Y. and Tong, L. (2004). Tracking of fast-fading channels in long-code CDMA. *IEEE Trans. Signal Process.*, **52** (3), 786–795.

Talwar, S. and Paulraj, A. (1997). Blind separation of synchronous co-channel digital signals using an antenna array—part II: Performance analysis. *IEEE Trans. Signal Process.*, **45** (3), 706–718.

Tong, L., Sadler, B. M., and Dong, M. (2004). Pilot-assisted wireless transmission. *IEEE Signal Process. Mag.*, **21** (6), 12–25.

Tsatsanis, M. K. and Giannakis, G. B. (1997). Transmitter induced cyclostationarity for blind channel estimation. *IEEE Trans. Signal Process.*, **45** (7), 1785–1794.

Vosoughi, A. and Scaglione, A. (2004a). The best training depends on the receiver architecture. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, 409–412.

—— (2004b). On the effect of channel estimation error with superimposed training upon information rates. In *Proc. IEEE Int. Symp. Inf. Theory*, 316.

Xavier, J., Barroso, V. A. N., and Moura, J. M. F. (2001). Closed-form correlative coding (cfc_2) blind identification of MIMO channels: Isometry fitting to second order statistics. *IEEE Trans. Signal Process.*, **49** (5), 1073–1086.

Xia, P., Zhou, S., and Giannakis, G. B. (2004). Adaptive MIMO-OFDM based on partial channel state information. *IEEE Trans. Signal Process.*, **52** (1), 202–213.

Yang, L., Ma, X., and Giannakis, G. B. (2004). Optimal training for MIMO fading channels with time- and frequency-selectivity. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, 821–824.

Yue, J., Kim, K. J., Reid, T., and Gibson, J. D. (2004). Joint semi-blind channel estimation and data detection for MIMO-OFDM systems. In *Proc. IEEE Symp.*

Emerging Tech.: Mobile and Wireless Commun., vol. 2, 709–712.

Zeng, Y. and Ng, T.-S. (2004). A semi-blind channel estimation method for multiuser multiantenna OFDM systems. *IEEE Trans. Signal Process.*, **52** (5), 1419–1429.

Zheng, L. and Tse, D. (2002). Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple antenna channel. *IEEE Trans. Inf. Theory*, **48** (2), 359–383.

Zhu, H., Farhang-Boroujeny, B., and Schlegel, C. (2003). Pilot embedding for joint channel estimation and data detection in MIMO communication systems. *IEEE Trans. Signal Process.*, **7** (1), 30–32.