Space-Time Wireless Systems:
From Array Processing to MIMO Communications

Edited by
H. Bölcskei, D. Gesbert, C. Papadias, and A. J. van der Veen
Contents

Part I  Multiantenna basics  page 1

Part II  Space-time modulation and coding  3

Part III  Receiver algorithms and parameter estimation  5

17 Training for MIMO communications  Y. Sung, T. Sung, B. M. Sadler, and L. Tong  7

Part IV  System-level issues of multiantenna systems  29

Part V  Implementations, measurements, prototypes, and standards  31

Author index  33

Index  34
Part I
Multiantenna basics
Part II
Space-time modulation and coding
Part III
Receiver algorithms and parameter estimation
In this chapter, an overview of training signal design for multi-input multi-output (MIMO) systems is provided with basic theoretical frameworks related to parameter estimation and information theory, as well as generalization and practical issues.

17.1 Introduction

Many MIMO communication systems and space-time techniques, e.g., BLAST (Foschini, 1996), are designed for coherent detection that requires channel state information for successful decoding. To facilitate channel estimation and synchronization in such systems, training or pilot signals are usually embedded in transmitted data streams. The design of these training signals can affect significantly the overall performance of a wireless system.

Since the use of training signals reduces effective data throughput, it is natural to seek optimal design of these embedded signals; one may ask “how many training symbols are necessary?” or “what is the optimal pilot sequence and its placement within data streams?”

Optimal training design for MIMO systems is a challenging task since the number of channel parameters to estimate increases rapidly as the number of transmitting and receiving antennas increases. Optimality of design depends on various factors such as receiver implementation, channel model, and design criteria. Although receiver architecture must be taken into account, training design is primarily a transmitter technique. Once a training scheme is chosen, it may be standardized for a specific application. It is therefore important that a training scheme is optimal or near optimal for a wide range of channel conditions and receiver implementation. Furthermore, since designers may have different design constraints and objectives, it is preferable that the training scheme is optimal for different design criteria.
The problem of optimal training design has been investigated extensively for various systems in recent years, and a substantial literature is available; see Tong et al. (2004) and references therein. Typical design criteria include the information-theoretic metrics such as the cut-off rate, information (ergodic) capacity, and outage probability (Foschini, 1996; Telatar, 1999). For example, training signals for MIMO systems are optimized by maximizing data throughput under outage probability constraint (Marzetta, 1999), and by maximizing a lower-bound on training-based information capacity as a function of training design parameters (Hassibi and Hochwald, 2003; Ma et al., 2002; Yang et al., 2004; Baltersee et al., 2001). Typically, these information-theoretic approaches to training design have been applied to a class of systems constrained to using training in a specific way, where channel is estimated using classical estimators based on training signal only and the channel estimate is subsequently used for data decoding.

The information-theoretic metrics are global measures of a communication system, yielding good insights into the trade-off between the quality of channel estimate and data throughput. Often, we are interested in the optimality of specific system components. For example, one may be interested in the training scheme that minimizes channel estimation error for a given amount of training. In this case, a sensible measure is the mean square error (MSE) of the channel estimator. Classical training-based estimators form a channel estimate based only on the observations of training signals. For these channel estimators the problem of training placement is straightforward; training symbols should be clustered in a single block since corrupted observations by unknown data are discarded. (In the case of a flat-fading channel the symbol-time location of the training symbols does not matter.) Hence, only the design of optimal pilot sequence remains and has been investigated by minimizing the MSE for flat-fading MIMO channels (Marzetta, 1999; Silverstein, 1997; Guey et al., 1999; Wong and Park, 2004) and for frequency-selective MIMO channels (Yang and Wu, 2002), by minimizing auto- and cross-correlations (Fan and Mow, 2004), and by maximizing a lower bound on the training-based capacity (Hassibi and Hochwald, 2003). These works confirm the merit of orthogonal training sequences, e.g., Hadamard codes, between multiple transmitting antennas.

On the other hand, a more sophisticated class of estimators, referred to as semiblind, use all available observations (i.e., training and data) to estimate channel. The semiblind methods treat unknown data as nuisance parameters that can either be marginalized or estimated jointly with a channel. These semiblind approaches improve the estimation performance for a given amount of training. Because it is desirable that the design of an optimal
training scheme does not depend on the specific algorithm used at the receiver, the Cramér-Rao bound (CRB) (Van Trees, 1968) is a natural choice as a figure of merit in this case. Semiblind CRB’s have been developed under deterministic and stochastic data assumption (de Carvalho and Slock, 1997; Dong and Tong, 2002), and for MIMO channels (Sadler et al., 2001b; Kozick et al., 2003). For semiblind channel estimation, the position of training symbols makes a difference since both training and data are modelled in estimation. Using the CRB as metric, training design has been investigated for linear finite impulse response (FIR) MIMO channels (Dong and Tong, 2002), for orthogonal space-time coded systems (Budianu and Tong, 2002), and for FIR MIMO systems using affine precoding (Vosoughi and Scaglione, 2004, 2003).

Multiple transmitting antennas are also considered for multicarrier systems ( Bölcskei et al., 2002). Using training-based estimators and their MSE’s, optimal design of a training signal for MIMO orthogonal frequency division multiplexing (OFDM) systems is considered (Li et al., 1999; Tung et al., 2001; Barhumi et al., 2003).

Although frameworks and methods for training design have been vigorously investigated in recent years, there are still many open problems and practical issues such as superimposed versus multiplexed training, optimal design for tracking of MIMO channels, the trade-off between optimality and complexity, etc.

The remainder of this chapter is organized as follows. Section 17.2 describes MIMOs system models. Section 17.3 provides fundamental frameworks emphasizing parameter estimation and information-theoretic approaches. Some generalization and practical issues are given in Section 17.4, followed by the conclusion in Section 17.5.

17.2 System model

Section 17.2 describes MIMO system models.

17.2.1 FIR MIMO channel

Channel models are crucial for the further development of theoretical frameworks. We first consider the block fading (quasi-static) channel model (Biglieri et al., 1998) which provides a good approximation to slowly varying channels. In this model, channels are assumed to be time-invariant within a block of length $B$ symbol intervals and then to change to a different fading state. The value of $B$ is designed to be approximately less than or equal to the
Fig. 17.1. A MIMO system with $M$ transmitting and $N$ receiving antennas.

coherence time of the fading channel. Fig. 17.1 illustrates a typical model of a MIMO system with $M$ transmitting and $N$ receiving antennas, where the channel between each pair of transmitting and receiving antennas can be viewed as a single-input single-output (SISO) channel. When the bandwidth of a transmitted signal is wide and intersymbol interference occurs due to multipaths, it is usually assumed that the channel has a linear finite impulse response of order $L$ which captures the delay spread of the channel. Thus, in complex baseband representation the output of a typical discrete-time FIR MIMO channel is described by

$$
y(i) = \sum_{l=0}^{L} H(l)s(i - l) + w(i), \quad (17.1)
$$

where $s(i) \triangleq [s_1(i), \ldots, s_M(i)]^T$ is the input $M$-vector at symbol interval $i$, $y(i) = [y_1(i), \ldots, y_N(i)]^T$ is the output $N$-vector, and $w(i) \triangleq [w_1(i), \ldots, w_N(i)]^T$ is the receiver noise vector. ($()^T$ denotes the matrix transpose.) The noise vector $w(i)$ is usually modeled as complex Gaussian $\mathcal{N}(\mathbf{0}, \sigma_w^2 \mathbf{I})$, independent over $i$. The channel is represented by a sequence of $M \times N$ matrices, $\{H(l), l = 0, 1, \ldots, L\}$. (Here, the block index for the channel matrices is omitted.) For antenna pair $(m, n)$, the channel impulse response is given by

$$
h^{(mn)} = [H_{mn}(0), H_{mn}(1), \ldots, H_{mn}(L)]^T, \quad (17.2)
$$

where $H_{mn}(l)$ is the $(m, n)$ element of $H(l)$. Depending on the approach, there are several choices for the assumption on the elements of $\{H(l)\}$; one may view the elements as deterministic but unknown quantities or as realizations of random variables with a known joint probability distribution.

Stacking the observation vectors for an entire block except the observa-
Training for MIMO communications

When the bandwidth of the transmitted signal is narrow so that all spectral components of the signal experience almost the same fading magnitude and phase, the flat-fading MIMO channel model can be applied as a special case of (17.1), and the channel output is given by

$$ y(i) = Hs(i) + w(i), \quad i = 1, \ldots, B, $$

where the channel between a pair of transmitting and receiving antennas is described by a single scalar coefficient. In the flat-fading case we can write the observation block of length $B$ in a matrix form:

$$ \mathbf{Y} = \mathbf{HS} + \mathbf{W}, $$

where $\mathbf{Y} \triangleq [y(1), \ldots, y(B)]$, $\mathbf{S} \triangleq [s(1), \ldots, s(B)]$, and $\mathbf{W} \triangleq [w(1), \ldots, w(B)]$.

17.2.2 Pilot-assisted transceiver structure

While the Shannon theory does not mandate channel estimation, the idea of acquiring channel state before decoding using training signals, is widely
employed in practice and has been proposed for many space-time systems (Foschini, 1996; Marzetta, 1999; Wolniansky et al., 1998; Alamouti, 1998). So-called pilot-assisted transmissions (PAT) embed pilot signals into data streams. The presence of training signals implies that they will be used at the receiver explicitly or implicitly. Explicit approaches, as illustrated in Figure 17.2, estimate channel parameters and subsequently use the channel estimate for demodulation and decoding. The channel estimator takes the training signal \( s_t \) (and possibly the entire observation \( y \)), produces a channel estimate \( \hat{H} \), and feeds the estimate to the decoder. Practical decoders may assume that the estimated channel parameters are perfect. Such an assumption is of course not valid in a strict sense, and the corresponding scheme is referred to as a mismatched decoder (Merhav et al., 1994; Lapidoth and Narayan, 1998). An alternative is to treat the estimated channel parameters as part of the observation. The decoder exploits the joint statistics of \((\hat{H}, y)\).

Implicit approaches, on the other hand, treat the training signal as side information. The channel estimator in Fig. 17.2 is bypassed, and the training signal is used to tune the receiver directly. In adaptive equalization techniques, for example, channel estimates might not be obtained explicitly, rather the training is used to adaptively update an equalizer.
17.3 Pilot signal design: Framework

In this section, we present the frameworks on which various optimal training designs are based. We consider two fundamental issues. The first is how much training is needed. The problem is not well-posed without constraints on data rates. The proper figures of merit are of information-theoretic nature, where there is a trade-off between having more training for better channel estimation and more channel uses for higher data rates. The second question focuses on how a training signal is embedded into a data stream when there is a fixed allocation of training—the problem of optimal placement. Here, we can fix the amount of training and optimize the placement of training symbols, and both information theoretic and parameter estimation measures can be used.

17.3.1 Information-theoretic approaches

The use of information-theoretic metrics is crucial to revealing trade-offs among training designs. In these settings a training scheme provides side information about unknown channels. Along this line there is an extensive literature on reliable communications under channel uncertainty (see Lapidoth and Narayan (1998) and references therein), which provides useful tools for obtaining achievable rate expressions that may be optimized with respect to training design parameters. Specifically, when channel estimates are available to the receiver, bounds on mutual information (Lapidoth and Narayan, 1998; Caire and Shamai, 1999; Medard, 2000; Lapidoth and Shamai, 2002) are the primary equations used in many of the optimal training designs. These expressions, however, do not incorporate the resources required to obtain channel estimates. For the analysis of training-based schemes one must take into account the resources allocated for pilot transmission.

For MIMO systems the problem of training design for the BLAST system was first investigated in Marzetta (1999). Based on the non-ergodic channel model (i.e., one codeword is contained within one block length of a quasi-static channel) and the outage probability due to the error of a pure training-based channel estimator, the required training period was obtained by maximizing data throughput for the block fading flat MIMO channel (17.7) assuming that $H_{mn} \sim \mathcal{N}(0,1)$. It is shown in Marzetta (1999) that for maximum data throughput the required training period is half of the block length and no less than the number of transmitting antennas.

For the class of ergodic block fading† MIMO channels, the problem of

† In this case, one codeword is spread over many blocks, which may be a reasonable scenario for a system employing an interleaver after the channel encoder.
training design was investigated using the information capacity as metric with the flat fading model (17.7) (Hassibi and Hochwald, 2003). Assuming time-division multiplexed (TDM) training within a block, Hassibi and Hochwald maximized a lower bound on training-based channel capacity with respect to the number of pilots used in a block and the power allocated to pilots. Their work provides a useful framework for analyzing the capacity achievable by training-based schemes in general, and leads to several interesting insights. They identified that, in the low SNR regime and when the coherence time (block length) was short, training-based schemes incurred a substantial penalty; training can lead to bad channel estimates, and no training may be preferable. On the other hand, training-based schemes were close to being optimal in high SNR and long coherence time regimes. This is consistent with the intuition that, with a negligible price paid to obtaining high quality estimates, we can assume that the channel is approximately known at the receiver. Later, similar approaches have been applied for the training design for frequency-selective and time- and frequency-selective MIMO channels (Ma et al., 2002; Yang et al., 2004).

Information-theoretic formulation

We illustrate an example of information-theoretic formulation for MIMO training design by Hassibi and Hochwald (2003), which is germane to many scenarios considered in the literature. A typical information-theoretic setup assumes the explicit transceiver structure described in Fig. 17.2, where the channel estimate is formed by a classical estimator based on training signals only. In this section, we also assume the transceiver structure in which the transmitter does not know the channel and the receiver learns the channel using the minimum mean-square error (MMSE) channel estimator. The presentation follows that in Hassibi and Hochwald (2003).

Consider the blockwise time-varying flat MIMO channel (17.8) with normalization such that the elements of $\mathbf{H}$ and $\mathbf{W}$ have unit mean square:

$$
\mathbf{Y} = \sqrt{\frac{\rho}{M}} \mathbf{HS} + \mathbf{W},
$$

(17.9)

where $\mathbf{Y}$ is the $N \times B$ channel output observed for the block, $\mathbf{S}$ is the $M \times B$ channel input matrix consisting of the training subblock $\mathbf{S}_t$ and the data subblock $\mathbf{S}_d$ (i.e., $\mathbf{S} = [\mathbf{S}_t, \mathbf{S}_d]$), $\mathbf{W}$ is additive white Gaussian noise with zero mean and unit variance, and $\rho$ is the expected SNR at each receiving antenna. For simplicity, we assume that the channel matrix $\mathbf{H}$ is Gaussian, with independent entries of zero mean and unit variance, independent over each block. We assume that $B_t$ symbol intervals are used for transmitting
training within a block, and $\rho_t$ and $\rho_d$ are the received SNR at each receiving antenna for the training period and the data period, respectively. Thus, we have

$$B = B_t + B_d, \quad \text{and} \quad \rho B = \rho_t B_t + \rho_d B_d.$$  \hfill (17.10)

Notice that for given $B$ and $\rho$, $\{B_t, \rho_t\}$ is sufficient to specify the training parameters together with the values of $S_t$. Assuming the training-based MMSE channel estimate $\hat{H}$ available at the decoder, the model in (17.9) can be rewritten for the data period as

$$Y_d = \sqrt{\frac{\rho_d}{M}}  \hat{H} S_d + Z, \quad Z = \sqrt{\frac{\rho_d}{M}}  \hat{H} S_d + W_d,$$ \hfill (17.11)

where $S_d$ is the unknown data with power constraint $\text{tr} \mathbb{E} \{S_d S_d^H\} = M B_d$, and $\hat{H} = H - \hat{H}$ is the estimation error, which is uncorrelated (independent for the Gaussian assumption) to $\hat{H}$ by the orthogonal principle for the MMSE estimation. The difference between (17.9) and (17.11) is that $\hat{H}$ of (17.11) is known to the receiver, whereas $\hat{H}$ of (17.9) is unknown. In addition, $Z$ is not independent of $S_d$ in (17.11), and its normalized variance is given by

$$\sigma_Z^2 \triangleq \frac{1}{NB_d} \text{tr} \mathbb{E} \{ZZ^H\},$$

$$= \frac{\rho_d}{MNB_d} \text{tr} \left[ \mathbb{E} \{\hat{H}\hat{H}^H\} \mathbb{E} \{S_d^H S_d\} \right] + 1.$$ \hfill (17.12)

If the decoder takes $\hat{H}$ as part of the observation and uses it along with $Y_d$ to decode $S_d$, the capacity of such a scheme is lower bounded by

$$C \geq \inf_{p_x: \text{tr} \mathbb{E} \{ZZ^H\} = N B_d \sigma_Z^2} \sup_{p_{S_d}: \text{tr} \mathbb{E} \{S_d^H S_d\} = M B_d} \frac{B - B_t}{B} I(S_d; Y_d; \hat{H}),$$

$$= \inf_{p_x: \text{tr} \mathbb{E} \{ZZ^H\} = N B_d \sigma_Z^2} \sup_{p_{S_d}: \text{tr} \mathbb{E} \{S_d^H S_d\} = M B_d} \frac{B - B_t}{B} I(S_d; Y_d; \hat{H}),$$ \hfill (17.13)

where the equality is the result of applying the chain rule under the assumption that the data $S_d$ is independent of the channel estimate $\hat{H}(S_t, Y_t)$ based on the training signals only. The right-hand side of (17.13) is simply the channel capacity of the scheme added by the worst-case noise. If we choose $S_d$ to be zero mean Gaussian with $\mathbb{E} \{S_d^H S_d\} = B_d I_M$ and Gaussian noise with the power constraint in (17.13), the capacity is lower bounded by

$$C \geq C_{LB} \triangleq \mathbb{E}_H \left\{ \frac{B - B_t}{B} \log \det \left( I_N + \frac{\rho_d \sigma_H^2}{1 + \rho_d \sigma_H^2} \frac{\hat{H}^H \hat{H}}{M} \right) \right\}.$$ \hfill (17.14)
where $\sigma_H^2 = 1 - \sigma_H^2$, $\sigma_H^2 = (1/NM) \text{tr} E\{\hat{H}^H \hat{H}\}$, and the normalized estimate $\hat{H} = (1/\sigma_H^2) \hat{H}$. The proof can be found in Hassibi and Hochwald (2003). The fact that Gaussian noise is the worst uncorrelated additive noise for the Gaussian model can also be found in Medard (2000); Diggavi and Cover (2001); Lapidoth and Shamai (2002).

Notice that the lower bound is a function of training design parameters. The problem of training design can be formulated as maximizing the lower bound on capacity (17.14). Let $\mathcal{P}$ denote the set of training parameters that specify $B_t$, $\rho_t$, and the values and placement of pilot symbols. For the considered training scheme, the optimal training scheme $\mathcal{P}^*$ is given by

$$\mathcal{P}^* = \sup_{\mathcal{P}} C_{LB}.$$ 

The above optimization is performed with respect to the training percentage $B_t/B$, power allocation $\rho_t$, and training symbol values and placement. This framework can be extended to channel estimators with some constraint other than the MMSE estimator and to more general channel models with frequency-selectivity and time-variation (Baltersee et al., 2001; Ma et al., 2002; Yang et al., 2004).

### 17.3.2 Signal processing perspectives on MIMO training design

We now consider MIMO training design, when there is a fixed allocation of training, using performance measures related to parameter estimation. We focus here on the FIR MIMO channel in Section 17.2 and its training design, especially on the optimal placement of training symbols.

#### Multidimensional training placement models

Training symbols are traditionally time-division multiplexed. The use of an antenna array in MIMO systems extends multiplexing to the spatial dimension. In addition to interleaving training and data by time and spatial domain, we may consider superimposing pilots and data. The combination of these factors multiplies possible scenarios to be examined.

The key to unifying various schemes is to view the problem of training design as one of power allocation. Specifically, in each design dimension—time, space, frequency, etc.—a pair of power allocation parameters are used to describe the design. In the case of a single-carrier MIMO system with $M$ transmitting antenna illustrated in Fig. 17.3 (a), each transmitted symbol $s_m(i)$ is specified by the time index $i$ and the transmitting antenna index
Training for MIMO communications

Fig. 17.3. An illustration of the multidimensional training placement model (a partially shaded square indicates a pilot symbol superimposed onto a data symbol.)

If we assume that pilots may be superimposed in any position, we have

\[ s_m(i) = \sqrt{\phi_{mi}} s^t_m(i) + \sqrt{\gamma_{mi}} s_d^t_m(i), \quad m = 1, \ldots, M, \quad i = 1, \ldots, B, \]  

(17.15)

where \( s^t_m(i) \) satisfying \( |s^t_m(i)| = 1 \) are known with allocated power \( \phi_{mi} \geq 0 \), and \( s_d^t_m(i) \) are unknown data with zero mean and unit variance with \( \gamma_{mi} \geq 0 \). For the transmission of a block of size \( B \) the training scheme is parameterized by a \( M \times B \) training matrix \( S_t = [s^t_m(i)]_{(m,i)} \) and non-negative power allocation matrices \( \Phi = [\phi_{mi}] \) and \( \Gamma = [\gamma_{mi}] \). The formulation can be extended to higher dimensions illustrated in Fig. 17.3 (b).

Constraints can be imposed on training design in several ways. For example, a constraint on the total power of training can be given by

\[ \frac{1}{MB} \sum_{i=1}^{B} \sum_{m=1}^{M} \mathbb{E}\{|\sqrt{\phi_{mi}} s^t_m(i)|^2\} = \frac{1}{MB} \sum_{i=1}^{B} \sum_{m=1}^{M} \phi_{mi} = P. \]  

(17.16)

Training design for block fading channels

For block fading MIMO channel models typically training symbols are inserted in each block, and channel estimation and symbol detection are performed within each block with the help of training signals. These training signals can be exploited in different ways (see Fig. 17.2). A classical training-based estimator forms a channel estimate \( \hat{H} \) based only on the observations \( y_t \) corresponding to the training signal \( s_t \), whereas semiblind estimators use all available observations \( y \) (i.e., training and data) to obtain a channel estimate.

When training-based estimators are used for single carrier systems over MIMO channels with flat or frequency-selective fading, the design of a given
amount of training is straightforward, and reduces to the problem of optimal sequence design (Marzetta, 1999; Silverstein, 1997; Guey et al., 1999; Wong and Park, 2004; Yang and Wu, 2002; Fan and Mow, 2004; Hassibi and Hochwald, 2003). Hence, we here focus on semiblind estimation of FIR MIMO channels with the Cramér-Rao bound on mean square error (MSE) as design metric. Specifically, the MSE of any unbiased estimator $\hat{H}$, under regularity conditions, is lower bounded by

$$\mathbb{E}\{||\hat{H} - H||^2_F\} \geq \text{tr}(F^{-1}(\Gamma, \Phi, S_t)),$$

(17.17)

where $F(\Gamma, \Phi, S_t)$ is the Fisher information matrix (FIM) which is a function of the training design parameters to be optimized. Thus, optimal pilot placement is found by minimizing the FIM over all feasible allocations $\{\Gamma, \Phi, S_t\}$ satisfying given constraints. For the parameter estimation formulation the FIM serves as the counterpart of the lower bound $C_{LB}$ of (17.14) in the information-theoretic setup.

The CRB may be formulated with both random and deterministic parameter models; random models lead to useful insights into ensemble behavior, whereas deterministic models provide means for assessing specific realizations of channels and sources. The channel CRB may also be a function of the unknown data transmitted simultaneously with the training (de Carvalho and Slock, 1997). When the unknown data are treated as random parameters, they need to be marginalized to obtain the likelihood function and the CRB is a function of the (prior) data distribution. If, on the other hand, these unknown data are treated as deterministic unknown parameters, then the data may be viewed as nuisance parameters that affect the CRB of the channel estimator. Note that, although the CRB can be achieved with finite data samples in some estimation problems such as in the case of a linear system model, the achievability is not always guaranteed with finite data samples. However, the existence of asymptotically efficient algorithms (e.g., the maximum likelihood estimator (MLE) in many cases) justifies the use of the CRB as a design criterion.

Using the CRB as design metric and a random channel model, optimal placement of training symbols and power allocations is derived for FIR SISO and MIMO channels in Dong and Tong (2002). It is shown that the optimal training placement scheme that minimizes the CRB is independent of the prior channel distribution and the SNR. The placement depends on the power allocation, and the number and placement of training symbols in a block. Using the CRB as metric, the design of affine precoder is optimized as well in Vosoughi and Scaglione (2004, 2003), where it is shown that different assumptions about unknown data lead to different affine precoder designs.
The CRB has also been employed to study the impact of side information, including training, with MIMO channels in Kozick et al. (2003). Identifiability of the unknown channel(s) may also be studied by considering the minimum conditions under which the CRB exists (that is, the minimum conditions under which the Fisher information matrix [FIM] becomes full rank, which is referred to as the FIM identifiability). Necessary and sufficient conditions for this to occur are provided in Moore and Sadler (2004) covering cases from SISO to FIR MIMO channels. Minimum conditions include the length of the training.

Superimposed training schemes are also considered for MIMO systems. A superimposed pilot scheme for space-time coded transmission over flat block fading MIMO channels was considered in Budianu and Tong (2002), where the problem setting is general but the optimal placement was not obtained. However, the analysis revealed the weakness of superimposed training in the block stationary case, showing that TDM training scheme had smaller CRB than that of superimposed training. On the other hand, if training should be included in every block and the channel estimation is accurate, the superimposed scheme provides higher mutual information.

**Block fading channels: A design example**

In this section, we present a particular example from Dong and Tong (2002). We consider optimal training placement of the block-fading FIR MIMO channel in Section 17.2.1 with the semiblind CRB with random-parametric formulation as design metric. (The notations and definitions follow those in Section 17.2.1.)

For the placement problem, each transmitted symbol is either training or data, i.e.,

$$s_m(i) = \sqrt{\phi_{mi}}s^t_m(i) + \sqrt{\gamma_{mi}}s^d_m(i), \ m = 1, \ldots, M, \ i = 1, \ldots, B; \quad (17.18)$$

where $\phi_{mi}$ is one for the training locations in space-time and $\gamma_{mi} = 1 - \phi_{mi}$. Thus, the set $\mathcal{P}$ of all possible placements can be specified as follows. Let $B^t_m$ and $B^d_m$ be the number of training and data symbols, respectively, within a block of length $B$ for transmitting antenna $m$. For transmitting antenna $m$ we define $\mu^{(m)}$ and $\nu^{(m)}$; $\mu^{(m)} = [\mu_1^{(m)}, \ldots, \mu_{\beta_m+1}^{(m)}]$ a data subblock length vector, $\nu^{(m)} = [\nu_1^{(m)}, \ldots, \nu_{\beta_M}^{(m)}]$ a training cluster length vector, where $\beta_m$ is the number of training clusters in a transmitted block at antenna $m$, as illustrated in Fig. 17.3.2. Constraining the total number of data and pilot symbols, we have $\sum_{j=1}^{\beta_{m+1}} \mu_j^{(m)} = B^d_m$ and $\sum_{j=1}^{\beta_m} \nu_j^{(m)} = B^t_m$. Thus, the set
of all possible placements is given by
\[ P = \{ (\mu, \nu) : \mu = [\mu^{(1)}, \ldots, \mu^{(M)}], \nu = [\nu^{(1)}, \ldots, \nu^{(M)}] \} \] satisfying the power constraint.

We make the following assumptions:

A1) The data symbols \{s^d_m(i), i = 1, \ldots, B\} for transmitting antenna \( m \) form an i.i.d. sequence (independent over \( m \)) drawn from the probability density function (pdf) \( p_s(\cdot) \) with zero mean and variance \( \sigma_{d;m}^2 \). The power of training symbols for antenna \( m \) is defined as
\[ \sigma_{t;m}^2 \triangleq \frac{1}{B^t_m} \sum_{i=1}^{B^t_m} |s^t_m(i)|^2. \]

A2) The \((L + 1)MN\) coefficients of the channel matrices \{\( H(l), l = 0, \ldots, L \)\} are i.i.d. random variables with pdf \( p_h(\cdot) \).

A3) The data \( s \), channel \( h \) and noise \( w \) are jointly independent, where \{\( w_m(i) \)\} are i.i.d. circular complex Gaussian noise with zero mean and variance \( \sigma_w^2 \).

A4) \( B^t_m \sigma_{t;m}^2 > \sigma_{d;m}^2 \) for \( m = 1, \ldots, M \). That is, the total power of training symbols is at least larger than that of one data symbol for all transmitting antennas.

From (17.18) the data MB-vector \( s \) of (17.3) for the entire block is decomposed into two MB-vectors:
\[ s = s_t + s_d, \quad (17.19) \]
where \( s_t \) is made by putting the known symbols at the training locations in space-time and zero padding elsewhere, i.e., at the data locations. Similarly, the data matrix \( H(s) \) of (17.5) is decomposed by
\[ \mathcal{H}(s) = \mathcal{H}(s_t) + \mathcal{H}(s_d), \quad (17.20) \]
where \( \mathcal{H}(s_t) \) and \( \mathcal{H}(s_d) \) are formed by (17.6) using \( s_t \) and \( s_d \), respectively. We define the following matrices:
\[ \mathcal{R}_s \triangleq \mathcal{H}(s)^\dagger \mathcal{H}(s) \text{ and } \mathcal{R}_{s_t} \triangleq \mathcal{H}(s_t)^\dagger \mathcal{H}(s_t). \quad (17.21) \]
Training for MIMO communications

\[ B_i^{(m)} \geq 2L + 1 \]

\[ B_i^{(m)} \leq 2L \]

\( \mathcal{R}_{s_{1}} = \text{diag}(B_1^{(1)} \sigma_{L_1}^2, B_1^{(2)} \sigma_{L_2}^2) \otimes \mathbf{I}_N \)

Fig. 17.5. Optimal placements of training symbols for a MIMO channel with two transmitting antennas.

**Theorem 17.1** (Dong and Tong, 2002) Under the assumptions A1-A4 and appropriate regularity conditions (Van Trees, 1968; Weinstein and Weiss, 1988), the MSE matrix of any unbiased estimator \( \hat{\mathbf{h}}(\mathbf{y}, \mathbf{s}_t) \), defined as

\[ \mathcal{M}(\hat{\mathbf{h}}) \triangleq \mathbb{E}\{|\hat{\mathbf{h}}(\mathbf{y}, \mathbf{s}_t) - \mathbf{h}(\mathbf{y}, \mathbf{s}_t) - \mathbf{h}|^H\}, \quad (17.22) \]

satisfies the following inequality

\[ \mathcal{M}(\hat{\mathbf{h}}) \geq \Lambda(P, \mathbf{s}_t) \triangleq \left[ \frac{1}{\sigma_w^2} \mathbb{E}\{\mathcal{R}_s\} + \rho_h^2 \mathbf{I}_{(L+1)MN} \right]^{-1}, \quad (17.23) \]

where \( \Lambda(P, \mathbf{s}_t) \) is the complex CRB, the expectation in (17.23) is taken w.r.t. the unknown data symbols, and \( \rho_h^2 = \mathbb{E}\{\left| \frac{\partial \ln p_h(h)}{\partial h} \right|^2 \} \) with the expectation taken with respect to \( p_h(h) \).

Thus, the CRB is a function of the placement parameters \( P \), the training symbol values \( \mathbf{s}_t \) via \( \mathbb{E}\{\mathcal{R}_s\} \) and the channel distribution through \( \rho_h^2 \). The CRB is optimized over \( (P, \mathbf{s}_t) \) to obtain the best training scheme.

It is shown in Dong and Tong (2002) that if the number of training symbols is large enough to cover the block edges corrupted by interblock interference, it is optimal to place (known) zeros wasting no power at the two block edges and to place the remaining known symbols in the middle of the block (not corrupted by interblock interference) such that these symbols satisfy a certain orthogonal condition between transmitting antennas. On the other hand, if the known symbols cannot cover the block edges, it is optimal to place roughly half of the known symbols of value zero at each edge and one symbol with all available training power in the middle of the block such that this symbol does not interfere with the middle symbol of different transmitting antenna streams through intersymbol interference. Fig. 17.3.2 illustrates such a placement for the case of two transmitting antennas. The difference between training-based and semiblind estimations can be clearly...
For training-based estimators the locations of known symbols within a transmitted stream should be clustered as a single subblock, and the single cluster of one transmitting antenna should be aligned with the training cluster of the other transmitting antenna stream.

Generalizations to correlated channel taps, symbol-by-symbol power constraints, etc., are available in the reference. The framework illustrated here can be generalized by modifying the signal decomposition (17.19). For example, superimposed training schemes can be modelled by allowing $\phi_{mi}$ to take arbitrary values between 0 and 1. Also, the combination of linear precoding and training embedding (possibly including superimposing) can be considered (Vosoughi and Scaglione, 2004, 2003). In this case, (17.19) can be modified as

$$s = Fs_I + s_t,$$  \hspace{1cm} (17.24)

with a precoding matrix $F$ with full column rank, known symbols $s_t$ and information symbols $s_I$. The precoder $F$ can be optimized along with $s_t$ to obtain minimum CRB.

### 17.4 Generalization and other issues

#### 17.4.1 MIMO-OFDM systems

In the previous sections, we have considered training signal design for single carrier systems over MIMO channels. Multiple transmitting and receiving antennas can also be used in multi-carrier systems using orthogonal frequency division multiplexing (OFDM), which has been standardized for many applications such as digital television broadcasting, digital audio broadcasting, wireless local area networks (WLANs), etc. A typical configuration is to apply the OFDM processing at each transmitting and receiving antenna. For OFDM systems each OFDM symbol is preceded by a cyclic prefix (CP) or guard interval, and intersymbol interference in time is irrelevant if the guard interval is larger than the maximum channel dispersion, whereas the interference between subcarriers may occur due to the mobility of transmitter or receiver and the frequency offset of local oscillators.

For MIMO OFDM systems, assuming a TDM training period, optimal design has been considered for training-based schemes. In the TDM case, the problem reduces to a two-dimensional sequence design in space and subcarrier domain. Optimal sequence was obtained by minimizing the MSE of a training-based channel estimator (Li et al., 1999; Tung et al., 2001; Larsson and Li, 2001), showing the optimality of orthogonal sequences (in subcarrier domain) between different transmitting antennas. On the other
hand, the optimal placement of pilot tones in single or multiple OFDM symbols was considered when the training period contains both pilot tones and data tones (Barhumi et al., 2003). Barhumi et al. show that the optimal training requires some orthogonality between different transmitting antenna signals in addition to the well-known conditions of equal power and equal spacing for pilot tones for SISO OFDM systems.

17.4.2 Fast-fading MIMO channels

In the previous sections, we considered block fading channels where the channel is time-invariant for a block. However, for fast fading channels this assumption may not be valid, i.e., the channel varies from symbol to symbol.

In this case, we typically assume that channel variations are highly correlated, at least for a short time, which is consistent with mobile channel measurements (Jakes, 1974). Indeed, if the channel process behaves independently from sample to sample, then training is required for every sample and training placement is not an issue. A practical model is the first order autoregressive (AR) model of the channel process $h_i$ that leads to a state space representation

$$h_{i+1} = Ah_i + Bu_i,$$

where the channel vector $h_i$ at time $i$ is made by concatenating the MIMO channel matrix $H_i$, $A$ characterizes the fading rate, and $u_i$ is the driving noise. Higher order AR models have also been employed for mobile channel modelling. AR models provide a reasonable fit to the widely used Jake's model that characterizes the power spectral density of the channel process $h_i$.

Optimal training design for the fast-fading MIMO channels is a challenging task, and provides many interesting research topics such as the comparison of the superimposed scheme with TDM training, the investigation of optimal training amount as function of fading rate, etc. From a signal processing perspective one may consider the Kalman filter for MIMO channel tracking (Komninakis et al., 2002) and optimal training design. Since the CRB is not tractable for fast fading channels, one possible design metric is the steady-state MSE of the training-based optimal estimator, Kalman filter, as in Dong et al. (2004). While the CRB in the block fading model gives a lower bound which can be achieved by the MLE at least asymptotically, the steady-state MSE of the optimal filter gives a good figure of merit for the fast-fading model (17.25). Another useful technique for fast-fading channels is a basis expansion model (Giannakis and Tepedelenlioglu, 1998).
This method absorbs the time-variation of a channel into a set of known basis functions and converts a general time-varying channel model to a block frequency-selective fading model.

### 17.4.3 Semiblind approach with side information

In Section 17.3.2, we have noted how training can be treated as side information and incorporated into semiblind channel estimation and decoding. Often, the receiver has additional side information such as constant modulus signals (or a more generally known constellation), known power levels, known angles of arrival, space-time coding, precoding, etc. Exploiting this additional side information can result in enhanced channel estimates and cochannel interference rejection. Alternatively, the side information can be exploited to decrease the required amount of training for a given performance level. The constant modulus (CM) property is particularly useful in this regard, and leads to tractable algorithms. The impact of many forms of side information can be analyzed using the constrained CRB (Stoica and Ng, 1998; Sadler et al., 2001a). In Sadler and Kozick (2000) it is demonstrated that reduced training sizes are needed when the CM property is exploited.

An example from Kozick et al. (2003) is shown in Fig. 17.6, showing CRBs and symbol estimation against training size for a 2 × 2 system (block fading, block size = 30). Independent 8-PSK data streams come from each transmit antenna, and results are shown for recovering one of these streams. A 3-ray flat channel was used, with significant spatial angle overlap. The SNR is 20 dB. The top curve bounds performance of a semiblind system that
Training for MIMO communications

exploits both training and the unknown symbols, and indicates the value of training is most significant in the first few training samples, approaching the known channel bound as the training size grows. The bottom curve incorporates the side information that the signals are constant modulus; this bound is nearly coincident with the bound obtained assuming a known channel matrix $H$. Also shown are simulation results for a nonlinear least squares semiblind algorithm (NLS2) that exploits both training and CM. The algorithm comes close to the CRB for small training size. (Cases with better channels show that the CRB can be attained.) This example shows how the amount of training needed in a MIMO setting can be significantly reduced, and source estimation enhanced, when side information is exploited (in this case, constant modulus signaling).

17.5 Conclusion

In this chapter, we have presented an overview of training signal design for MIMO systems. General MIMO channel models have been given, and common design criteria have been reviewed. Also, information-theoretic and signal processing frameworks have been discussed with particular examples.

17.6 Acknowledgements

This work was supported in part by the Multidisciplinary University Research Initiative (MURI) under the Office of Naval Research Contract N00014-00-1-0564, Army Research Laboratory CTA on Communication and Networks under Grant DAAD19-01-2-0011, and National Science Foundation under Contract CCR-0311055.

References


Part IV

System-level issues of multiantenna systems
Part V

Implementations, measurements, prototypes, and standards
Author index
Index

Cramér-Rao bound, 18
autoregressive channel model, 23
block fading MIMO channel, 9
constant modulus property, 24
Fisher information matrix, 18
generalized Sylvester matrix, 11
Hankel matrix, 11
intersymbol interference, 10
optimal training design, 8
semiblind channel estimation, 17
training sequence design, 8