

Brief paper

# Subspace system identification for training-based MIMO channel estimation<sup>☆</sup>

Chengjin Zhang<sup>a</sup>, Robert R. Bitmead<sup>b,\*</sup>

<sup>a</sup>Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093-0407, USA

<sup>b</sup>Department of Mechanical and Aerospace Engineering, University of California, San Diego, La Jolla, CA 92093-0411, USA

Received 30 January 2004; received in revised form 16 March 2005; accepted 20 April 2005

Available online 16 June 2005

## Abstract

The application of state-space-based subspace system identification methods to training-based estimation for quasi-static multi-input–multi-output (MIMO) frequency-selective channels is explored with the motivation for better model approximation performance. A modification of the traditional subspace methods is derived to suit the non-contiguous nature of training data in mobile communication systems. To track the time variation of the channel, a new recursive subspace-based channel estimation is proposed and demonstrated in simulation with practical MIMO channel models. The comparison between the state-space-based channel estimation algorithm and the FIR-based Recursive Least Squares algorithm shows the former is a more robust modeling approach than the latter.

© 2005 Elsevier Ltd. All rights reserved.

**Keywords:** MIMO; Channel estimation; State-space models; Subspace system identification

## 1. Introduction

Digital communication using multiple transmit and receive antennas has been one of the most important technical developments in modern communications. In a rich scattering environment, multi-input–multi-output (MIMO) systems offer significant capacity gain at no cost of extra spectrum (Foschini & Gans, 1998). Furthermore, space–time channel codes (Tarokh, Jafarkhani, & Calderbank, 1999) can be applied to build spatial redundancy in the transmitted signal such that maximum spatial diversity is achieved. Coherent space–time processing schemes assume the availability of a channel model at the receiver. Therefore, this model needs to be estimated at the receiver end.

<sup>☆</sup> This paper was presented at the 13th IFAC Symposium on System Identification, Rotterdam, the Netherlands, 2003. This paper was recommended for publication in revised form by Associate Editor Brett Ninness under the direction of Editor T. Söderström.

\* Corresponding author. Tel.: +1 858 822 3477; fax: +1 858 822 3107.

E-mail addresses: [zhangc@ucsd.edu](mailto:zhangc@ucsd.edu) (C. Zhang), [rbitmead@ucsd.edu](mailto:rbitmead@ucsd.edu) (R.R. Bitmead).

It is the subject of this paper to explore the application of state-space models to represent MIMO frequency-selective channels in the hope for better model approximation performance and possible parsimonious parametrization. Since any channel model is essentially an approximation of the physical channel, the goal of channel modeling is to find a model that is as close to the real channel as possible and maintains manageable complexity as well. FIR models have been widely used for wireless channels for their simplicity and guaranteed stability. On the other hand, state-space models are able to provide better modeling performance since they are a more general class of models that includes FIR models as a subset. It is found that state-space models are able to provide low-order models of high-quality channel approximation.

Furthermore, when the channel is frequency selective and has multiple inputs and multiple outputs, an FIR model can be very non-parsimonious and contain excessive redundancy since it represents the sub-channels of a MIMO channel with *separately* parametrized finite-length impulse responses. In situations where there exists correlation between the

sub-channels, it may be beneficial to adopt a state-space model which models the whole channel as a single entity and hence captures the structure in the MIMO channel while allowing a more parsimonious description of it.

For quasi-static or slowly varying channels, training-based channel estimation is very common in practice. Therefore, an algorithm for estimating MIMO state-space models using training data is needed. We propose such a state-space channel estimation algorithm based on recursive subspace system identification (SSI). The simulations of this paper demonstrate comparable (but marginally slower) adaptation performance of state-space methods but improved approximation performance compared to FIR-based recursive least-squares (RLS) methods.

Subspace system identification (SSI) algorithms are a group of methods that identify a MIMO state-space system in a straightforward way using numerically robust computation tools such as singular value decomposition (SVD) and QR factorization (Verhaegen & Dewilde, 1992; Van Overschee & De Moor, 1996).

One issue with SSI-based MIMO channel estimation is that the training sequences need not be contiguous in the data stream in practical mobile communication systems. Instead, they appear as the overhead or mid-amble of a frame of data. More specifically, in the Global System for Mobile communications (GSM), a 26-bit-long segment in the middle of each 156-bit frame is allocated for the insertion of the training sequence which is the only data known to both the transmitter and the receiver (Steele, 1992). Since the traditional SSI methods assume the availability of a contiguous input–output data stream, they need to be modified to suit the situation of MIMO channel estimation. It will be shown that, when the length of the training sequence,  $N_t$ , is sufficiently large compared to the order or the McMillan degree of the model of the MIMO channel, the modified non-contiguous-data approach retains performance similar to the original contiguous-data approach.

Furthermore, a recursive version of the SSI algorithms is needed when the channel becomes time-varying as is fundamentally the case in mobile communications. In order to update the channel estimate adaptively upon receiving new training data, a recursive SSI algorithm is proposed based on the recursive ordinary multivariable output-error state space (MOESP) model identification algorithm in Lovera, Gustafsson, and Verhaegen (2000). The difference between the recursive MOESP algorithm in Lovera et al. (2000) and our modified recursive MOESP lies in the way they weight the input–output data with the exponential forgetting factor, which is normally included in recursive algorithms to adjust the effect of past input–output data on the current channel estimate. The newly proposed recursive MOESP algorithm for non-contiguous data is tested on trial channels by computer simulation and exhibits faster convergence than the one in Verhaegen and Deprettere (1991) and Lovera et al. (2000).

This paper is organized as follows. Section 2 reviews the problem of adaptive channel estimation. As the motivation for using state-space models for MIMO channels, numerical simulation results using practical channel models are presented to compare the state-space-based recursive channel estimation algorithm to the FIR-based recursive least-squares (RLS) channel estimator. Section 3 provides a detailed discussion of the state-space-based channel estimation algorithm used in the simulation study in Section 2. We present the state-space channel model and propose a modification of the traditional SSI algorithms to suit non-contiguous training data. A modified recursive MOESP is also developed for identifying time-varying channels. The paper is concluded in Section 4.

## 2. State-space models for MIMO wireless channels

### 2.1. Adaptive channel estimation

In wireless communications, the transmitted signal is reflected and attenuated by various objects before it reaches the receive antenna. The received signal is a sum of delayed and weighted versions of the original signal. When the maximum delay is significant enough compared to the symbol period of the transmission, the interaction between these waves causes the so-called inter-symbol interference (ISI). The single-input–single-output ISI channel can be well described by a symbol-sampled discrete-time FIR model as follows:

$$y_k = \sum_{i=0}^{L-1} h_i u_{k-i} + n_k, \quad (1)$$

where  $u_k$ ,  $y_k$  are the  $k$ th input and output symbol, respectively,  $h_i$  is length- $L$  channel impulse response, and  $n_k$  is the additive white Gaussian noise.

In order to reduce the ISI, an equalizer is employed at the receiver to compensate for the channel effect. In training-based equalization schemes, a pre-selected training sequence, known to both the transmitter and the receiver, is inserted in each data frame and transmitted through the channel, as shown in Fig. 1. The receiver uses the input–output data either to estimate the channel response and then compute the coefficients of the equalizer or to estimate directly the coefficients of the equalizer which could be regarded as some type of inverse of the channel.

In this paper we assume that the receiver works in the indirect manner, i.e., it uses the input–output data to estimate the channel response first and then computes the coefficients of the equalizer. Furthermore, the channel is assumed to be quasi-static, i.e., it is time-invariant within each frame and changes independently from frame to frame. Our main interest is the estimation algorithm the receiver uses to estimate the channel.

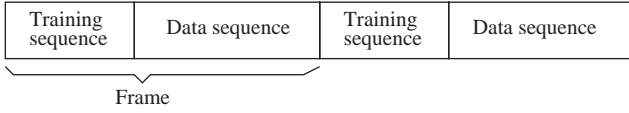


Fig. 1. Frame structure with training sequences.

With the advent of multiple antenna systems, the channel becomes an MIMO system so that MIMO system identification needs to be considered when estimating the channel. One approach is to build the MIMO model as a sequence of “stacked” or “vectorized” SISO channels using the fact that an FIR MIMO channel model has nearly the same form as its SISO counterpart except that each tap of the FIR MIMO channel is a gain matrix instead of a scalar as in (2) (Fragouli, Al-Dhahir, & Turin, 2003; Al-Dhahir & Sayed, 2000). This approach inherits the advantage of guaranteed stability of FIR models, and the channel estimation can usually be performed via RLS-type or LMS-type algorithms. However, state-space models may be able to provide better modeling accuracy.

## 2.2. State-space channel models

A computer experiment was carried out in order to compare the modeling performance of state-space models and that of FIR models. For a given MIMO reference channel, non-recursive batch versions of both the state-space-based ordinary MOESP algorithm (Verhaegen, 1994) and FIR-based least-squares algorithm are applied to produce estimated channel models with various orders. The estimation performance is measured in terms of the scalar  $H_2$  norm of the difference between the estimated channel models and the reference channel. Notice we only compare the batch estimation methods in this section. Discussion of the convergence behaviour comparison between the newly to-be-proposed recursive MOESP and MIMO recursive least-squares methods will be presented in Section 2.3.

### 2.2.1. Generation of reference MIMO channels

The reference MIMO channel is assumed to be composed of multiple finite impulse responses. Assuming all the sub-channels have the same length  $L$ , an FIR MIMO channel with  $m$  transmit antennas and  $p$  receive antennas can be represented by assembling all the taps of the sub-channel impulse responses with the same delay  $\tau$  into a matrix  $\mathbf{H}_\tau$  and representing the channel as a series of matrices

$$[\mathbf{H}_0, \mathbf{H}_1, \dots, \mathbf{H}_{L-1}],$$

$$\mathbf{H}_\tau = \begin{bmatrix} h_{11}(\tau) & h_{12}(\tau) & \dots & h_{1m}(\tau) \\ h_{21}(\tau) & h_{22}(\tau) & \dots & h_{2m}(\tau) \\ \vdots & \vdots & \ddots & \vdots \\ h_{p1}(\tau) & h_{p2}(\tau) & \dots & h_{pm}(\tau) \end{bmatrix}. \quad (2)$$

The simulated channels are taken to be square (i.e.,  $m = p$ ), and the length of the impulse responses is set to 6. The chan-

nel taps are generated as zero-mean circularly-symmetric complex Gaussian (ZMCSCG) random variables with equal covariance. The widely used exponential power delay profile (PDP) (Pedersen, Mogensen, & Fleury, 2000) is applied to all the sub-channels to control the distribution of the average power of each tap,

$$E[\|\mathbf{H}_l\|_F] = e^{-\alpha l}, \quad l = 0, 1, \dots, L-1, \quad (3)$$

where  $\|\cdot\|_F$  represents the Frobenius norm of a matrix and  $\alpha$  is a constant which decides how fast the tail of the impulse response decays. The simulations presented in this section were also performed using the COST 207 power delay profiles (Failli, 1989) and yielded conceptually similar results.

The MIMO channel is assumed to be uncorrelated in delay but correlated in spatial dimension, i.e.,

$$E(\text{vec}(\mathbf{H}_k)\text{vec}(\mathbf{H}_l)^H) = \begin{cases} \mathbf{0}, & k \neq l \\ \mathbf{R}, & k = l \end{cases},$$

where  $\mathbf{R}$  is a non-zero covariance matrix and  $\text{vec}(\cdot)$  is an operator that stacks the columns of a matrix on top of each other to form a vector. We adopt a common model for the spatial correlation structure of  $\mathbf{H}_\tau$  (Yu & Ottersten, 2002),

$$\mathbf{H}_\tau = (R_{RX})^{1/2} \mathbf{H}_w (R_{TX})^{T/2}, \quad (4)$$

where  $\mathbf{H}_w$  is a  $p \times m$  matrix with IID ZMCSCG elements.  $R_{TX}$  and  $R_{RX}$  are, respectively, the transmit covariance matrix and receive covariance matrix. For spatially white MIMO channels, both  $R_{TX}$  and  $R_{RX}$  are identity matrices. For spatially correlated MIMO channels with uniform linear arrays at both the transmitter and the receiver, under the assumption of uniformly distributed angle of arrival,  $R_{TX}$  and  $R_{RX}$  can be modeled with closed-form expressions as functions of angular spread  $\Delta$ , angle of arrival  $\phi$  and normalized antenna spacing  $D/\lambda_c$ , where  $D$  is the antenna spacing and  $\lambda_c$  the carrier wavelength. See Salz and Winters (1994) for detailed development of the closed-form expressions.

The antenna configuration used in the simulation for spatially correlated MIMO channels is summarized by Table 1. Since the matrix taps of the channel impulse response in (2) generally correspond to paths from various clusters, their angles of arrival are generated as a uniformly distributed random variable on  $[-90^\circ, 90^\circ]$ . However, for simplicity, all the taps are assumed to have the same transmitter and receiver angular spread.

Fig. 2 shows the channel estimation performance of both the state-space-based MOESP algorithm and the FIR-based least-squares algorithm. The  $H_2$  channel estimation error is presented as a function of model order with the full order being 15. The order of a model is defined as the McMillan degree of its transfer function matrix.

Several observations on the modeling performance can be made. Firstly, state-space models can provide better model approximation performance than FIR models for both

Table 1  
Configuration of the TX/RX antenna arrays

	TX/MU	RX/BS
Angular spread ( $\Delta$ )	$30^\circ$	$5^\circ$
Angle of arrival ( $\phi$ )	$-90^\circ$ – $90^\circ$	$-90^\circ$ – $90^\circ$
Baseline length ( $L_{bs}$ )	0.3 m	1.0 m
Antenna spacing ( $D/\lambda_c$ )	$L_{bs}/(m-1)\lambda_c$	$L_{bs}/(p-1)\lambda_c$

spatially white MIMO channels and spatially correlated ones except for the full-order models. The full-order FIR models have the advantage that they have exactly the same structure as the reference channel. However, it is more important to examine the performance of reduced-order models, because in practice, without knowing the true order of the physical channel, any channel model we adopt is only an approximation of the physical channel and generally has a lower order than the true one. The advantage of state-space models comes from the fact that the state-space model is a more general class of models which includes FIR models as a subset.

Secondly, the number of low-order FIR models is significantly smaller than that of the low-order state-space models. This is because the order of an FIR model can only be reduced by decreasing the length of the impulse responses, which corresponds to a step size of  $\min\{m, p\}$  in terms of model order. Therefore, state-space models provide us with many more choices for low-order approximation of the reference channel than FIR models.

It can also be seen that FIR models are not able to take advantage of the spatial correlation structure in MIMO channels since the two FIR curves for both spatially white channels and for correlated channels nearly overlap. In contrast, state-space models use the spatial correlation to improve the model approximation performance, although the improvement does not appear to be significant.

Furthermore, the channel power delay profile has a large impact on how the channel approximation performance degrades when the model order decreases away from the full order. The reduced-order models for channels with large tail taps (Fig. 2(a)) suffer bigger loss in channel approximation performance compared to those for channels with small tail taps (Fig. 2(b)). However, for either type of power delay profile, state-space models always exhibit more graceful performance degradation than FIR models.

### 2.3. Performance of the recursive state-space MIMO channel estimation algorithm

In this section, we present the performance of the state-space-based recursive MOESP channel estimation algorithm compared to the FIR-based recursive least-squares (RLS) channel estimator. It is found that the convergence rate of the state-space method and that of the FIR method are comparable.

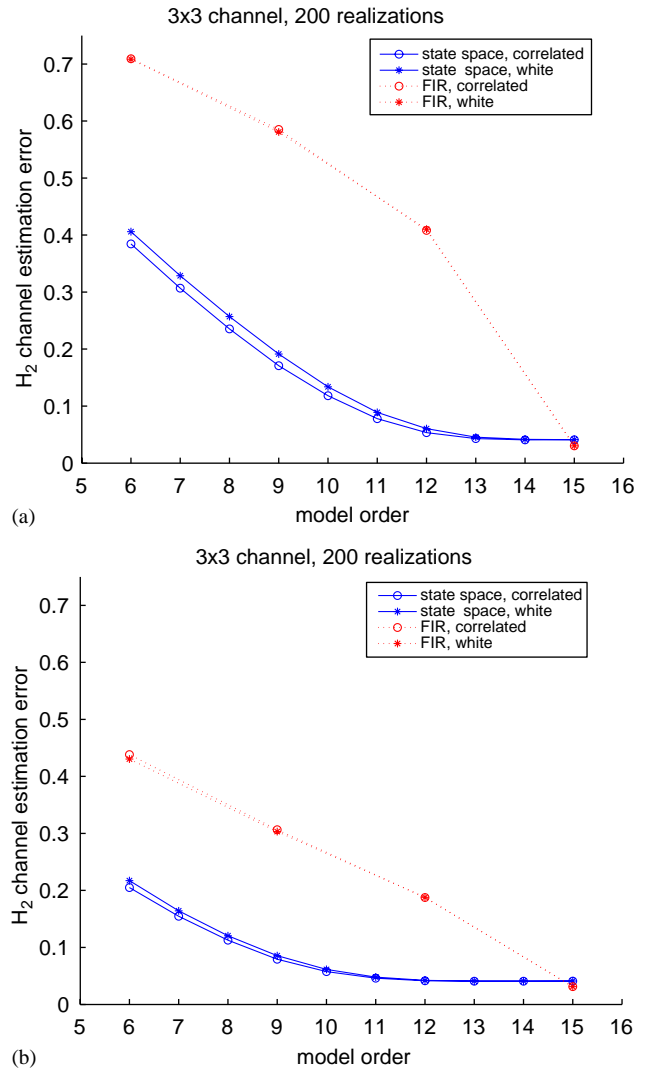


Fig. 2.  $H_2$  channel estimation error for  $3 \times 3$  channels. (a) Uniform power delay profile ( $\alpha = 0$ ). (b) Exponentially decaying power delay profile ( $\alpha = 0.5$ ).

Although the detailed development of the state-space algorithm used in the simulation is not discussed until Section 3, the purpose of presenting the results of the algorithm beforehand is to provide an immediate motivation.

The spatially correlated MIMO channels are assumed to be quasi-static and are generated in the same way as discussed in Section 2.2. The training sequences are contiguous random binary sequences that are uncorrelated in both the spatial dimension and the time dimension. The upper plots of Figs. 3(a) and (b) illustrate the transient performance of the estimators for  $2 \times 2$  channels with  $SNR = 20$  dB. There are separate curves for each different model order, with the dotted curves corresponding to the FIR models and the solid curves corresponding to state-space models. The lower figures are derived from the upper figures and plot one point for each model order and identification method. The horizontal

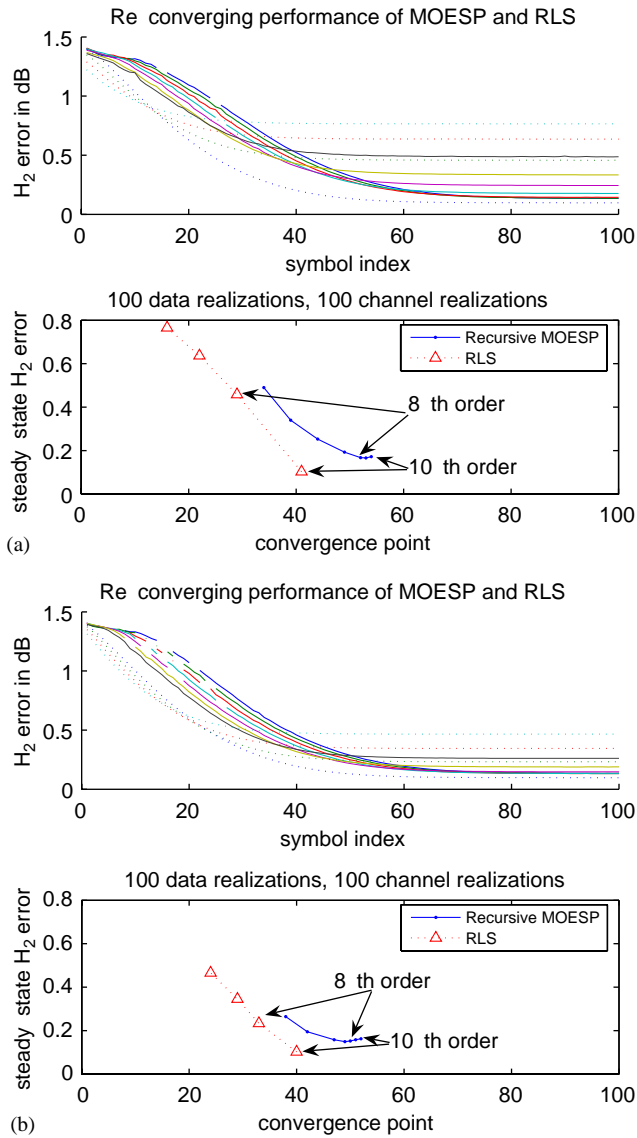


Fig. 3.  $H_2$  estimation error of recursive MOESP and RLS for  $2 \times 2$  spatially correlated channels. (a) Uniform power delay profile ( $\alpha=0$ ). (b) Exponential decaying power delay profile ( $\alpha=0.5$ ).

axis shows the input symbol number from which the channel model error remains below 110% of the steady-state value.

In terms of steady-state  $H_2$  channel estimation error, state-space models provide low-order models of high-quality approximation, whereas the reduced-length FIR models fail to provide comparable estimation error performance. This performance difference between reduced-order state-space models and reduced-length FIR models was also shown in Section 2.2 where batch MOESP and batch least-squares channel estimators were considered.

The convergence rate of the recursive MOESP algorithm is in general slower than that of the RLS algorithm, but the difference stays in a comparable range. The faster convergence of RLS is to be expected because the (spatially

and temporally) white training sequence is optimal for FIR model structure (Ljung, 1999).

The demonstration that subspace system identification methods can deliver a high-performance (low-error) MIMO channel model with adaptation rate comparable to that of FIR-based RLS methods, validates their further consideration for the adaptive MIMO channel equalization problem. In the next section we shall move on to present in more detail the necessary modifications to a standard SSI algorithm to achieve this performance in this signal environment.

### 3. Subspace system identification for MIMO channel estimation

#### 3.1. Subspace system identification

A state-space MIMO channel model is denoted by

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k,$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k + \mathbf{n}_k, \quad (5)$$

where  $\mathbf{u}_k$  ( $m \times 1$ ) is the input vector at time  $k$ ,  $\mathbf{y}_k$  ( $p \times 1$ ), output vector at time  $k$ ,  $\mathbf{x}_k$  ( $n \times 1$ ), state vector at time  $k$ , and  $\mathbf{n}_k$  ( $p \times 1$ ) is the white Gaussian noise vector at time  $k$ .

Subspace system identification (SSI) refers to a class of relatively recent algorithms, such as MOESP (Verhaegen & Dewilde, 1992) and N4SID (Van Overschee & De Moor, 1996), which apply input–output system identification methods to determine directly a state-space realization of a system. The key idea of SSI methods is to estimate the extended observability matrix through the projection of future input–output data onto past input–output data based on the relationship between Hankel matrices of the input and output given by

$$Y_{0,i,t} = \Gamma_i X_{0,t} + H_i U_{0,i,t} + N_{0,i,t}, \quad i > n, \quad (6)$$

where  $U_{0,i,t}$ ,  $Y_{0,i,t}$  and  $N_{0,i,t}$  are Hankel matrices of the input, output and noise, respectively, with the form of

$$Y_{0,i,t} = \begin{bmatrix} \mathbf{y}_0 & \mathbf{y}_1 & \cdots & \mathbf{y}_{t-i+1} \\ \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_{t-i+2} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{y}_{i-1} & \mathbf{y}_i & \cdots & \mathbf{y}_t \end{bmatrix}. \quad (7)$$

$X_{0,t}$  is a matrix containing the state vectors

$$X_{0,t} = [\mathbf{x}_0 \quad \cdots \quad \mathbf{x}_{t-i+1}].$$

$\Gamma_i$  and  $H_i$  are, respectively, the extended observability matrix and the matrix of Markov coefficients:

$$\Gamma_i = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{i-2} \\ CA^{i-1} \end{bmatrix},$$

$$H_i = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & 0 & \cdots & 0 \\ CAB & CB & D & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ CA^{i-2}B & & \cdots & D \end{bmatrix}.$$

The system matrices  $A$ ,  $B$ ,  $C$  and  $D$  are computed based on an estimate of the extended observability matrix,  $\hat{\Gamma}_i$ , which can be found by performing an LQ decomposition on the Hankel input–output data matrices as follows (Verhaegen & Dewilde, 1992).

$$\begin{bmatrix} U_{0,i,t} \\ Y_{0,i,t} \end{bmatrix} = \begin{bmatrix} R_{11} & 0 \\ R_{21} & R_{22} \end{bmatrix} \begin{matrix} mi \times mi \\ pi \times pi \end{matrix} Q, \quad (8)$$

and

$$\hat{\Gamma}_i = R_{22}.$$

See (Verhaegen & Dewilde, 1992; Van Overschee & De Moor, 1996; Ljung, 1999) for in-depth treatments of subspace system identification algorithms. Note that the identified state-space realization is not in a canonical form.

Based on the channel model given in (5), SSI methods require the input to satisfy the following requirements for the channel to be identifiable:

- (1) The input  $\mathbf{u}_k$  is uncorrelated with the additive Gaussian white noise  $\mathbf{n}_k$ .
- (2) The input  $\mathbf{u}_k$  is persistently exciting of order of at least twice the maximum order of the channel.
- (3) The symbols in the input sequence are contiguous and, for consistency, the number of inputs goes to infinity.

The first assumption is usually satisfied for wireless communication systems. The second one requires the training sequence to maintain a certain structure.

Also, notice that the third assumption places limitation on the application of SSI methods to channel estimation in wireless communication systems where the training sequences are usually not contiguous in time. Instead, these symbols are placed in the overhead or mid-amble of a frame and are separated by data symbols, as shown in Fig. 1. Furthermore, realistic channels are often time-varying which suggests the use of a recursive version of SSI algorithms. Therefore, SSI needs to be reformulated before it can be applied to channel estimation.

### 3.2. SSI for non-contiguous data

With non-contiguous training data, the state evolution of the received data must be restarted at the frame boundaries. This is at variance with the standard formulation of SSI.

Consider the evolution of two contiguous  $N_t$ -symbol-long blocks of received data, with the first block commencing at time 0 and the second commencing immediately thereafter at  $N_t$ . For simplicity, we omit the noise and write the blocked state as,

$$Y_{0,i,N_t-1} = \Gamma_i X_{0,N_t-1} + H_i U_{0,i,N_t-1},$$

$$Y_{N_t,i,2N_t-1} = \Gamma_i X_{N_t,2N_t-1} + H_i U_{N_t,i,2N_t-1}.$$

In standard SSI approaches, these are combined to form a new matrix equation

$$Y_{0,i,2N_t-1} = \Gamma_i X_{0,2N_t-1} + H_i U_{0,i,2N_t-1}.$$

This absorbs the data vectors into the Hankel structure of the new  $U$  and  $Y$  matrices. This adds further columns to the equation to be solved for the observability matrix  $\Gamma_i$ .

Next consider the availability of discontinuous  $N_t$ -symbol-long blocks of received data with the first block commencing at time 0 and the second at some later time  $t$  with  $t > N_t$ . Then, we still achieve the relationships

$$Y_{0,i,N_t-1} = \Gamma_i X_{0,N_t-1} + H_i U_{0,i,N_t-1},$$

$$Y_{t,i,t+N_t-1} = \Gamma_i X_{t,t+N_t-1} + H_i U_{t,i,t+N_t-1}.$$

But now the absorption of the data into individual Hankel matrices is no longer possible, because of the non-contiguity of the received data.

We may, however, write an augmented equation composed from the above set:

$$[Y_{0,i,N_t-1} \ Y_{t,i,t+N_t-1}] = \Gamma_i [X_{0,N_t-1} \ X_{t,t+N_t-1}] + H_i [U_{0,i,N_t-1} \ U_{t,i,t+N_t-1}]. \quad (9)$$

This set of equations to be solved for  $\Gamma_i$  is comparable to the contiguous-data set of equations. It has the same number of rows,  $i$ , and has  $i - 1$  fewer columns. When the length of the training sequence,  $N_t$ , is sufficiently large compared to the dimension of the generalized observability matrix,  $\Gamma_i$  (which depends on the state dimension of the model), in estimation power the non-contiguous-data approach is similar to the contiguous-data approach, as shown in Fig. 4.

Fig. 4 shows the mean  $H_2$  channel estimation error of the MOESP algorithms with non-contiguous training sequences of various length,  $N_t$ . The  $x$  axis represents the length of the training data block,  $N_t$ . For each value of  $N_t$ , all the 160-symbol-long training data is divided to  $\lceil 160/N_t \rceil$  non-contiguous blocks based on which the MOESP algorithm for non-contiguous data is run over 100 random  $2 \times 2$  7-tap spatially white MIMO channels. The full order of the channels is 12 and  $i$  is set to 13 in the MOESP algorithm. It can be seen that, as  $N_t$  increases beyond 30, the  $H_2$  error of non-contiguous training becomes very close to the contiguous training case where  $N_t = 160$ .

### 3.3. Recursive SSI for time-varying channels

When the channel is time-varying, channel estimation needs to be carried out in a recursive manner so that any

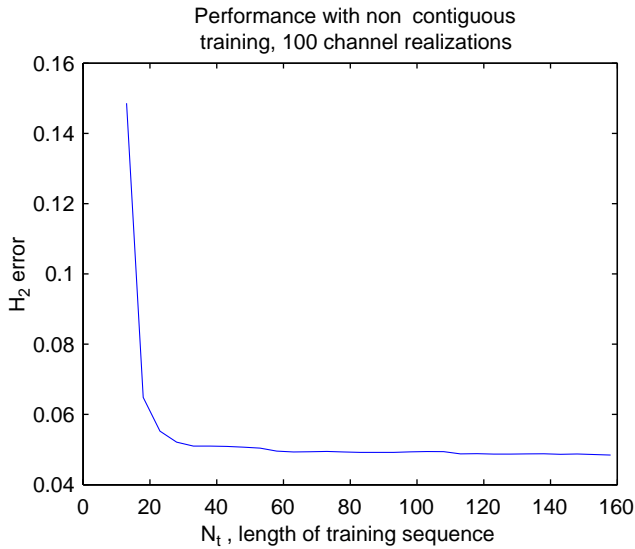


Fig. 4.  $H_2$  Channel estimation error of the MOESP algorithms with non-contiguous training data,  $i = 13$ .

newly received data can be used to produce an up-to-date estimate of the channel. An exponential forgetting factor should also be able to be included to reduce the effect of distant past data.

Recursive SSI algorithms have been developed to meet this need. A recursive MOESP scheme was proposed in Verhaegen and Deprettere (1991) and Lovera et al. (2000), whose key idea is to update the estimate of the extended observability matrix with the newly received data using LQ factorization.

To give a brief summary of the recursive ordinary MOESP algorithm in Verhaegen and Deprettere (1991) and Lovera et al. (2000), consider that at time  $t$  we have an LQ decomposition as in (8):

$$\begin{bmatrix} U_{0,i,t} \\ Y_{0,i,t} \end{bmatrix} = \begin{bmatrix} R_{11}(t) & 0 \\ R_{21}(t) & R_{22}(t) \end{bmatrix} Q(t).$$

It was shown in Verhaegen and Dewilde (1992) that  $R_{22}(t)$  is an estimate of the extended observability matrix at time  $t$ ,  $\Gamma_i(t)$ , in the sense that they share the same column space.

The task of a recursive MOESP scheme would be to compute  $R_{22}(t+1)$ , an estimate of  $\Gamma_i(t+1)$ , based on the LQ factorization at time  $t$  and the newly received data at time  $t+1$ . Suppose at time  $t+1$  a new set of input–output data,  $\mathbf{u}_{t+1}$  and  $\mathbf{y}_{t+1}$ , becomes available. We can form the following data vectors:

$$\phi_u(t+1) = \begin{bmatrix} \mathbf{u}_{t-i+2} \\ \vdots \\ \mathbf{u}_t \\ \mathbf{u}_{t+1} \end{bmatrix}, \quad \phi_y(t+1) = \begin{bmatrix} \mathbf{y}_{t-i+2} \\ \vdots \\ \mathbf{y}_t \\ \mathbf{y}_{t+1} \end{bmatrix}.$$

Then the new input–output Hankel data matrix at time  $t+1$  can be formed by appending the above vectors to the right-hand end of the Hankel data matrix at time  $t$  and applying

an exponential forgetting factor  $\lambda \in (0, 1]$ . We perform an LQ factorization

$$\begin{aligned} \begin{bmatrix} U_{0,i,t+1} \\ Y_{0,i,t+1} \end{bmatrix} &= \begin{bmatrix} \lambda U_{0,i,t} & \phi_u(t+1) \\ \lambda Y_{0,i,t} & \phi_y(t+1) \end{bmatrix} \\ &= \begin{bmatrix} \lambda R_{11}(t) & 0 & \phi_u(t+1) \\ \lambda R_{21}(t) & \lambda R_{22}(t) & \phi_y(t+1) \end{bmatrix} \\ &\quad \times \begin{bmatrix} Q(t) & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} R_{11}(t+1) & 0_{mi \times pi} & 0_{mi \times 1} \\ R_{21}(t+1) & \lambda R_{22}(t) & \phi(t+1) \end{bmatrix} \\ &\quad \times Q(t+1) \\ &= \begin{bmatrix} R_{11}(t+1) & 0 \\ R_{21}(t+1) & R_{22}(t+1) \end{bmatrix} Q(t+1), \end{aligned} \quad (10)$$

where the  $pi \times 1$  vector  $\phi(t+1)$  is obtained through a series of Givens rotations on the large lower triangular matrix.  $Q(t+1)$  is a matrix with orthogonal rows, i.e.,  $Q(t+1)Q(t+1)^H = I$ . Thus  $R_{22}(t+1)$  is obtained through updating the LQ decomposition with the new data.

Notice that, if the aforementioned recursive procedure is carried out from time 0 to time  $t$ , the effective Hankel output data matrix used by the MOESP algorithm is given by

$$\begin{aligned} Y_{0,i,t}^\lambda &= [\lambda^{t-i+1} \phi_y(i-1), \dots, \lambda \phi_y(t-1), \phi_y(t)] \\ &= \begin{bmatrix} \lambda^{t-i+1} \mathbf{y}_0 & \dots & \lambda \mathbf{y}_{t-i} & \mathbf{y}_{t-i+1} \\ \vdots & & \vdots & \vdots \\ \lambda^{t-i+1} \mathbf{y}_1 & \dots & \lambda \mathbf{y}_{t-2} & \mathbf{y}_{t-1} \\ \lambda^{t+i-1} \mathbf{y}_{i-1} & \dots & \lambda \mathbf{y}_{t-1} & \mathbf{y}_t \end{bmatrix}, \end{aligned} \quad (11)$$

with  $U_{0,i,t}^\lambda$  having similar form. The input–output data  $\{\mathbf{y}_\tau\}_{\tau=0}^t$  and  $\{\mathbf{u}_\tau\}_{\tau=0}^t$  are not weighted according to the time instant they are received or transmitted, but according to which column they are positioned in the weighted Hankel matrices  $Y_{0,i,t}^\lambda$  and  $U_{0,i,t}^\lambda$ . Within each column of  $Y_{0,i,t}^\lambda$  or  $U_{0,i,t}^\lambda$ , the newest data enjoys the same weight as the past data. Therefore, the channel estimation algorithms using data Hankel matrices with this weighting structure may not be able to respond promptly to the channel variation represented in the newest input–output data. That is, this weighting approach may result in relatively slow convergence rate of the estimation algorithm.

### 3.3.1. Recursive MOESP with an improved weighting scheme

In order to improve the convergence rate, we propose that the exponential forgetting factor should be applied according to the time at which the data symbols are transmitted or received. This weighting scheme originates from the

weighting approach used by the RLS algorithm. The new effective Hankel output matrix  $\tilde{Y}_{0,i,t}^\lambda$  is given by

$$\tilde{Y}_{0,i,t}^\lambda = \begin{bmatrix} \lambda^t \mathbf{y}_0 & \dots & \lambda^i \mathbf{y}_{t-i} & \lambda^{i-1} \mathbf{y}_{t-i+1} \\ \vdots & & \vdots & \vdots \\ \lambda^{t-1} \mathbf{y}_1 & \dots & \lambda^{i-1} \mathbf{y}_{t-2} & \lambda \mathbf{y}_{t-1} \\ \lambda^{t+i-1} \mathbf{y}_{i-1} & \dots & \lambda \mathbf{y}_{t-1} & \mathbf{y}_t \end{bmatrix}, \quad (12)$$

with the new effective Hankel input data matrix  $\tilde{U}_{0,i,t}^\lambda$  having the same form.

Given the new Hankel data matrices, it is necessary to develop the channel identification procedure accordingly. Note that  $\tilde{Y}_{0,i,t}^\lambda$  and  $\tilde{U}_{0,i,t}^\lambda$  can be written as

$$\tilde{Y}_{0,i,t}^\lambda = A_p Y_{0,i,t} A, \quad (13)$$

$$\tilde{U}_{0,i,t}^\lambda = A_m U_{0,i,t} A, \quad (14)$$

where  $p$  and  $m$  are, respectively, the number of outputs and that of inputs.  $A_p$ ,  $A_m$  and  $A$  are given, respectively by

$$A_p = \text{diag}\{\lambda^{i-1} I_p, \dots, \lambda I_p, I_p\},$$

$$A_m = \text{diag}\{\lambda^{i-1} I_m, \dots, \lambda I_m, I_m\},$$

$$A = \text{diag}\{\lambda^{i-1}, \dots, \lambda, 1\}.$$

With (13) and (14), the original Hankel matrix relationship given in (6) can be rearranged to

$$\begin{aligned} A_p Y_{0,i,t} A &= \underbrace{A_p \Gamma_i}_{\triangleq \Gamma_i^\lambda} \underbrace{X_{0,t} A}_{\triangleq X_{0,t}^\lambda} + \underbrace{A_p H_i A_m^{-1}}_{\triangleq H_i^\lambda} A_m U_{0,i,t} A \\ &\quad + A_p N_{0,i,t} A, \\ \tilde{Y}_{0,i,t}^\lambda &= \Gamma_i^\lambda X_{0,t}^\lambda + H_i^\lambda \tilde{U}_{0,i,t}^\lambda + \tilde{N}_{0,i,t}^\lambda. \end{aligned} \quad (15)$$

Now (15) can be used as the basis of the recursive MOESP algorithm in place of (6). As a comparison, the conventional recursive MOESP is based on the relationship

$$Y_{0,i,t} A = \Gamma_i X_{0,t} A + H_i A_m^{-1} A_m U_{0,i,t} A + N_{0,i,t} A. \quad (16)$$

Since (15) produces estimates of  $\Gamma_i^\lambda$  and  $H_i^\lambda$  as opposed to  $\Gamma_i$  and  $H_i$  of (16), it is necessary to modify the procedure for computing  $A, B, C$  and  $D$  in terms of the new structure of  $\Gamma_i^\lambda$  and  $H_i^\lambda$ .

*Estimating C and A.* Form the following data vectors that contain the newest input–output data at time  $t + 1$

$$\begin{aligned} \tilde{\phi}_u(t+1) &= \begin{bmatrix} \lambda^{i-1} \mathbf{u}_{t-i+2} \\ \vdots \\ \lambda \mathbf{u}_t \\ \mathbf{u}_{t+1} \\ \lambda^{i-1} \mathbf{y}_{t-i+2} \\ \vdots \\ \lambda \mathbf{y}_t \\ \mathbf{y}_{t+1} \end{bmatrix}, \\ \tilde{\phi}_y(t+1) &= \begin{bmatrix} \lambda^{i-1} \mathbf{y}_{t-i+2} \\ \vdots \\ \lambda \mathbf{y}_t \\ \mathbf{y}_{t+1} \end{bmatrix}. \end{aligned}$$

Then, since  $\tilde{Y}_{0,i,t}^\lambda$  and  $\tilde{U}_{0,i,t}^\lambda$  satisfy the relation

$$\begin{bmatrix} \tilde{U}_{0,i,t+1}^\lambda \\ \tilde{Y}_{0,i,t+1}^\lambda \end{bmatrix} = \begin{bmatrix} \lambda \tilde{U}_{0,i,t}^\lambda & \tilde{\phi}_u(t+1) \\ \lambda \tilde{Y}_{0,i,t}^\lambda & \tilde{\phi}_y(t+1) \end{bmatrix},$$

the same procedure as described in (10) can be applied to update the estimate of the extended observability matrix at time  $t + 1$ ,  $\hat{\Gamma}_i^\lambda(t+1)$ , based on which  $\hat{C}(t+1)$  and  $\hat{A}(t+1)$  can be computed using the following relations:

$$C = \lambda^{-i+1} \Gamma_i^\lambda(1 : p, :), \quad (17)$$

$$\Gamma_i^{\lambda(1)} A = \lambda \Gamma_i^{\lambda(2)}, \quad (18)$$

where the matrix index notation of MATLAB is used and

$$\begin{aligned} \Gamma_i^{\lambda(1)} &= \Gamma_i^\lambda(1 : (i-1)p, :) = \begin{bmatrix} \lambda^{i-1} C \\ \lambda^{i-2} C A \\ \vdots \\ \lambda C A^{i-2} \end{bmatrix}, \\ \Gamma_i^{\lambda(2)} &= \Gamma_i^\lambda(p+1 : ip, :) = \begin{bmatrix} \lambda^{i-2} C A \\ \lambda^{i-3} C A^2 \\ \vdots \\ C A^{i-1} \end{bmatrix}. \end{aligned}$$

*Estimating B and D.* In order to estimate  $B$  and  $D$ , we define the orthogonal projection onto the left nullspace of  $\Gamma_i^\lambda$  as

$$P = I - \Gamma_i^\lambda (\Gamma_i^{\lambda H} \Gamma_i^\lambda)^{-1} \Gamma_i^{\lambda H},$$

where  $(\cdot)^H$  represents the complex conjugate transpose of a matrix. By using  $P$  to remove the term  $\Gamma_i^\lambda X_{0,t}^\lambda$  in (15), we obtain

$$\begin{aligned} \underbrace{P \tilde{Y}_{0,i,t}^\lambda}_{\triangleq K} \underbrace{\tilde{U}_{0,i,t}^{\lambda \dagger}}_{\triangleq V} &= P H_i^\lambda + \underbrace{P \tilde{N}_{0,i,t}^\lambda}_{\triangleq V} \underbrace{\tilde{U}_{0,i,t}^{\lambda \dagger}}_{\triangleq V} \\ K &= P H_i^\lambda + V. \end{aligned} \quad (19)$$

By decomposing  $K, P$  and  $V$ , respectively into

$$K = [K_1 \ K_2 \ \dots \ K_i],$$

$$P = [P_1 \ P_2 \ \dots \ P_i],$$

$$V = [V_1 \ V_2 \ \dots \ V_i],$$

where  $\{K_l\}_{l=1}^i, \{P_l\}_{l=1}^i$  and  $\{V_l\}_{l=1}^i$  are  $ip \times p$  matrices, we can re-organize (19) as

$$\begin{bmatrix} K_1 \\ K_2 \\ \vdots \\ K_i \end{bmatrix} = \begin{bmatrix} P_1 & P_2 & P_3 & \dots & P_i \\ P_2 & P_3 & \dots & P_i & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P_i & 0 & \dots & 0 & 0 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \lambda^i \Gamma_i^{\lambda(1)} \end{bmatrix} \begin{bmatrix} D \\ B \end{bmatrix} + \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_i \end{bmatrix}. \quad (20)$$

Therefore, the least-squares solution for  $B$  and  $D$  can be obtained from (20). This procedure can be simplified by

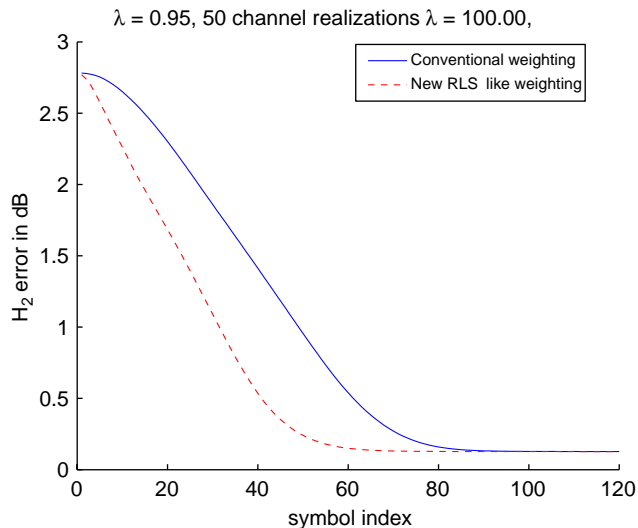


Fig. 5. Comparison between the conventional weighting and the new RLS-like weighting.

applying LQ decomposition as proposed in Verhaegen and Dewilde (1992).

In summary, (17), (18) and (20) are very similar to their counterparts in the original MOESP algorithm (Verhaegen & Dewilde, 1992) except they are modified to accommodate the new RLS-like exponential weighting scheme in (12).

Fig. 5 shows the learning curves of the recursive MOESP algorithms with different weighting scheme applied to 100  $2 \times 2$  6-tap spatially white MIMO channels. The RLS-like weighting approach has a clear advantage over the conventional one since the former exhibits a faster convergence rate and maintains the same steady-state performance as the latter. It is also worthwhile to point out that the recursive MOESP with RLS-like weighting has nearly the same computational complexity as the conventional one.

#### 4. Conclusions

A direct modeling approach for MIMO wireless communication channels using state-space models is proposed to achieve better modeling performance and possible parsimonious parametrization. A recursive subspace system identification (SSI) algorithm for non-contiguous training data is derived for channel estimation in multiple-antenna communication systems. The numerical results show that the state-space-based channel estimation algorithm is capable of providing low-order models of high-quality channel approximation. When the true order of the physical MIMO channel is unknown, the performance of the state-space-based channel estimator is less sensitive, or more robust, to the error in model order selection compared to that of the FIR-based RLS algorithm, while preserving comparable convergence rate.

#### Acknowledgements

The authors would like to thank Dr. Thomas Svantesson and Dr. Bhaskar Rao for the helpful discussion on channel modeling. We would also like to thank the anonymous reviewers for the comments and suggestions that helped improve the quality of this paper. This research was supported by Core Grant No. 02-10109 sponsored by Ericsson.

#### References

- Al-Dhahir, N., & Sayed, A. H. (2000). The finite-length multi-input–multi-output MMSE-DFE. *IEEE Transactions on Signal Processing*, 48(10), 2921–2936.
- Failli, M. (1989). *Digital land mobile radio communications—COST 207*, Final Report. Commission of the European Community, Luxembourg.
- Foschini, G. J., & Gans, M. J. (1998). On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications*, 6(3), 311–335.
- Fragouli, C., Al-Dhahir, N., & Turin, W. (2003). Training-based channel estimation for multiple-antenna broadband transmissions. *IEEE Transactions on Wireless Communications*, 2(2), 384–391.
- Ljung, L. (1999). *System identification, theory for the user*. (2nd ed.), Englewood Cliffs NJ: Prentice-Hall PTR.
- Lovera, M., Gustafsson, T., & Verhaegen, M. (2000). Recursive subspace identification of linear and non-linear wiener state-space models. *Automatica*, 36(11), 1639–1650.
- Pedersen, K. I., Mogensen, P. E., & Fleury, B. H. (2000). A stochastic model of the temporal and azimuthal dispersion seen at the base station in outdoor propagation environments. *IEEE Transactions on Vehicular Technology*, 49(2), 437–447.
- Salz, J., & Winters, J. H. (1994). Effect of fading correlation on adaptive arrays in digital mobile radio. *IEEE Transactions on Vehicular Technology*, 43(4), 1049–1057.
- Steele, R. (1992). *Mobile radio communications*. Piscataway: IEEE Press.
- Tarokh, V., Jafarkhani, H., & Calderbank, A. R. (1999). Space-time block coding for wireless communications: Performance results. *IEEE Journal on Selected Areas in Communications*, 451–460.
- Van Overschee, P., & De Moor, B. (1996). *Subspace identification for linear systems: Theory, implementation, applications*. Dordrecht: Kluwer Academic Publishers.
- Verhaegen, M. (1994). Identification of the deterministic part of mimo state space models given in innovations form from input–output data. *Automatica*, 30(1), 61–74.
- Verhaegen, M., & Deprettere, Ed. (1991). A fast recursive MIMO state space model identification algorithm. In *Proceedings of the 30th conference on decision and control*, Brighton, England.
- Verhaegen, M., & Dewilde, P. (1992). Subspace model identification Part I, the output-error state-space model identification class of algorithms. *International Journal of Control*, 56(5), 1187–1210.
- Yu, K., & Ottersten, B. (2002). Models for MIMO propagation channels: A review. *Wireless Communications and Mobile Computing*, 2, 273–291.



**Chengjin Zhang** is pursuing his Ph.D. degree in the Department of Electrical and Computer Engineering at University of California, San Diego since 2002. He received his M.S. degree in Electrical Engineering from Ohio State University in 2002 and his B.S. degree in Mechanical Engineering from Nanchang University, China in 1997.



**Robert Bitmead** holds the Cymer Corporation Chair at the Department of Mechanical and Aerospace Engineering, University of California, San Diego. He received his PhD from the University of Newcastle in 1980. He was with the Australian National University Department of Systems Engineering from 1982-1999. He is a Fellow of the Australian Academy of Technological Sciences and Engineering and a Fellow of the IEEE. He was member of the IFAC Council from 1996-2002 and a member of IFAC

Technical Board from 1990-1996. He hails originally from Sydney and was General Chair of the IEEE Conference on Decision and Control there from December 12-15, 2000. He is Editor of *Automatica* for Adaptive and Intelligent Control.