

**MASTER COPY:** PLEASE KEEP THIS "MEMORANDUM OF TRANSMITTAL" BLANK FOR REPRODUCTION PURPOSES. WHEN REPORTS ARE GENERATED UNDER THE ARO SPONSORSHIP, FORWARD A COMPLETED COPY OF THIS FORM WITH EACH REPORT SHIPMENT TO THE ARO. THIS WILL ASSURE PROPER IDENTIFICATION. NOT TO BE USED FOR INTERIM PROGRESS REPORTS; SEE PAGE 2 FOR INTERIM PROGRESS REPORT INSTRUCTIONS.

**MEMORANDUM OF TRANSMITTAL**

U.S. Army Research Office  
ATTN: AMSRL-RO-BI (TR)  
P.O. Box 12211  
Research Triangle Park, NC 27709-2211

Reprint (Orig + 2 copies)

Technical Report (Orig + 2 copies)

Manuscript (1 copy)

Final Progress Report (Orig + 2 copies)

Related Materials, Abstracts, Theses (1 copy)

CONTRACT/GRANT NUMBER: **W911NF0410224 (46637-CI-MUR)**

REPORT TITLE:

**A Performance Bound for Prediction of MIMO Channels**

T. Svantesson, A. Swindlehurst;

IEEE Transactions on Signal Processing, Vol. 54, No. 2, pp. 520-529 - February, 2006

is forwarded for your information.

SUBMITTED FOR PUBLICATION TO (applicable only if report is manuscript):

**Sincerely,**

**Dr. Lee Swindlehurst**

**Enclosure 3**

<b>REPORT DOCUMENTATION PAGE</b>			Form Approved OMB NO. 0704-0188
Public Reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comment regarding this burden estimates or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188,) Washington, DC 20503.			
1. AGENCY USE ONLY ( Leave Blank)	2. REPORT DATE <b>February 2006</b>	3. REPORT TYPE AND DATES COVERED <b>Reprint 01 June 2004 - 31 April 2006</b>	
4. TITLE AND SUBTITLE <b>A Performance Bound for Prediction of MIMO Channels</b>		5. FUNDING NUMBERS  <b>W911NF0410224</b>	
6. AUTHOR(S) <b>T. Svantesson, A. Swindlehurst;</b>			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>University of California. San Diego 9500 Gilman Drive, La Jolla, CA 92093-0407</b>		8. PERFORMING ORGANIZATION REPORT NUMBER  <b>N/A</b>	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)  U. S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211		10. SPONSORING / MONITORING AGENCY REPORT NUMBER  <b>N/A</b>	
11. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.			
12 a. DISTRIBUTION / AVAILABILITY STATEMENT  Approved for public release; distribution unlimited.		12 b. DISTRIBUTION CODE  <b>N/A</b>	
13. ABSTRACT (Maximum 200 words)  <b>Knowledge of future channel conditions can increase the performance of many types of wireless systems. This is especially true for radio channels with multiple transmit and receive antennas, i.e., multiple-input multiple-output (MIMO) systems. This paper derives a performance bound for MIMO channel prediction. It is assumed that prediction is based upon estimating a model for the channel and then extrapolating that model to predict future values of the channel. A vector formulation of the Cramér–Rao bound for functions of parameters is used to find a lower bound on the prediction error. Numerical evaluation of this bound shows that substantially longer prediction lengths are possible for MIMO channels than for single antenna channels. An intuitive interpretation of this result is that more of the channel structure is revealed when using multiple antennas at both ends. Finally, the longer prediction lengths for MIMO channels are confirmed by numerical results obtained by implementing a MIMO extension of a single-antenna prediction scheme.</b>			
14. SUBJECT TERMS  <b>Autoregressive processes, MIMO systems, parameter estimation, prediction methods.</b>		15. NUMBER OF PAGES <b>10</b>	16. PRICE CODE <b>N/A</b>
17. SECURITY CLASSIFICATION OR REPORT <b>UNCLASSIFIED</b>	18. SECURITY CLASSIFICATION ON THIS PAGE <b>UNCLASSIFIED</b>	19. SECURITY CLASSIFICATION OF ABSTRACT <b>UNCLASSIFIED</b>	20. LIMITATION OF ABSTRACT  <b>UL</b>

# A Performance Bound for Prediction of MIMO Channels

Thomas Svantesson, *Member, IEEE*, and A. Lee Swindlehurst, *Fellow, IEEE*

**Abstract**—Knowledge of future channel conditions can increase the performance of many types of wireless systems. This is especially true for radio channels with multiple transmit and receive antennas, i.e., multiple-input multiple-output (MIMO) systems. This paper derives a performance bound for MIMO channel prediction. It is assumed that prediction is based upon estimating a model for the channel and then extrapolating that model to predict future values of the channel. A vector formulation of the Cramér–Rao bound for functions of parameters is used to find a lower bound on the prediction error. Numerical evaluation of this bound shows that substantially longer prediction lengths are possible for MIMO channels than for single antenna channels. An intuitive interpretation of this result is that more of the channel structure is revealed when using multiple antennas at both ends. Finally, the longer prediction lengths for MIMO channels are confirmed by numerical results obtained by implementing a MIMO extension of a single-antenna prediction scheme.

**Index Terms**—Autoregressive processes, MIMO systems, parameter estimation, prediction methods.

## I. INTRODUCTION

THERE are many instances in wireless communications where knowledge of the future radio channel can be beneficial. For instance, if a fading dip can be predicted in advance, precautions such as power control or adaptive modulation can provide a significant performance increase [5], [15]. Performance gains can also be achieved by using prediction to bridge the time gap between channel estimates and the current state of the channel [12], [13]. Several researchers have addressed the problem of predicting the future channel for systems with a single transmitting and a single receiving antenna, i.e., a single-input single-output (SISO) channel [1], [4]–[6], [9], [20]. Unfortunately, their results show that for channels with dense multipath, the SISO channel can only be accurately predicted for short distances, on the order of a few tenths of a wavelength, depending on the scenario. Analysis

of a bound on the prediction error for SISO channels in [20] suggests that the channel predictor requires channel knowledge over many wavelengths in order to provide accurate predictions further into the future. In practice, the underlying structure of the radio channel due to scattering and reflections may not remain stationary long enough to provide the channel knowledge necessary for SISO prediction. Hence, for many environments, the practical use of SISO channel prediction appears limited.

Recently, the possibilities of predicting the channel when using an array at the receiver end of a wireless link have been investigated in [2]. A performance gain was observed in a downlink beamforming application by employing channel prediction. A plausible explanation for this is that more of the channel structure is revealed when several antennas sample the wavefield. Using arrays at both the transmitter and receiver, it seems likely that even more of the channel structure is revealed and better prediction performance may be expected. Another advantage of using multiple antennas at both ends of a wireless link is that the achievable data rate can be significantly increased since the spatial dimension can be exploited more efficiently [7], [21]. Since these large performance gains are reached without requiring additional bandwidth, such multiple-input multiple-output (MIMO) systems have attracted considerable attention. It has also been found that channel knowledge is important in order to realize the full potential of these systems [16]. Therefore, channel prediction may be particularly useful in MIMO systems since better prediction performance may be possible and improved performance is more critical than in the SISO case.

This paper will study the theoretical performance of MIMO channel prediction and show that longer *prediction lengths* than for SISO systems are possible. In this paper, we define *prediction length* to be the interval (measured in units of space or time) over which the channel can be predicted to within a certain level of accuracy (a more precise definition will be given later). A bound for the prediction error of MIMO channels will be derived and analyzed in a manner similar to the bound for SISO channels in [20]. A common ray-based channel model for SISO and MIMO channels is used in the analysis throughout the paper. Prediction is based upon estimating the model for the channel using a segment where the channel is known and then simply extrapolating this model to predict future values of the channel. Note that the same channel model is used both for generating data and when estimating/predicting the channel. A lower bound on the prediction error is derived by using a vector formulation of the Cramér–Rao lower bound (CRB) for functions of the parameters. The prediction performance is then studied using this bound, and it is found that it is possible to predict the channel

Manuscript received November 15, 2003; revised December 19, 2004. This work was supported in part by the National Science Foundation under Information Technology Research Grant CCR-0313056, and by the U. S. Army Research Office under the Multi-University Research Initiative (MURI) Grant W911NF-04-1-0224. Parts of this manuscript have been presented at the 11th IFAC Symposium on System Identification, Rotterdam, The Netherlands, 2003, and at the 37th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, 2003. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Helmut Boelskei.

T. Svantesson was with the Department of Electrical and Computer Engineering, University of California, San Diego, CA 92093 USA. He is now with ArrayCom, Inc., San Jose, CA 95131 USA (e-mail: tomaso@arraycom.com).

A. L. Swindlehurst is with the Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT 84604 USA (e-mail: swindle@ee.byu.edu).

Digital Object Identifier 10.1109/TSP.2005.861105

further into the future when multiple antennas are used at both the transmitter and receiver. Finally, the performance of a practical implementation of a simple MIMO prediction scheme is investigated. It is found that the channel indeed can be predicted further ahead when using several antennas.

The paper is organized as follows. Section II introduces the channel model and the notation used throughout the paper. The bound on the variance of the prediction error, i.e., the CRB, is derived in Section III. Numerical evaluations of the bound for different scenarios are presented in Section IV. Initial experiments with channel prediction algorithms are reported in Section V, and some conclusions are discussed in Section VI.

## II. CHANNEL MODEL

A ray-based MIMO channel model, which several recent measurement campaigns have used to describe and analyze measured data [14], [17], [22], will be used to study prediction performance. The model is essentially an extension of the common narrow-band (frequency flat fading) SISO channel model [10]

$$h(t) = \sum_{l=1}^L \alpha_l e^{j\omega_l t} \quad (1)$$

where  $\alpha_l$  is the scattering coefficient of path  $l$ ,  $\omega_l$  is the Doppler frequency in radians per second (rad/s) of path  $l$  at time  $t$ , and  $L$  denotes the total number of paths. A simple way of extending this model to a MIMO system with  $M_t$  transmit and  $M_r$  receive antennas is to add the spatial dimension of each ray as

$$\mathbf{H}(t) = \sum_{l=1}^L \alpha_l \mathbf{a}_{r,l} \mathbf{a}_{t,l}^T e^{j\omega_l t} \quad (2)$$

where  $\mathbf{a}_{r,l}$  is the array response vector at the receiver associated with the  $l$ th path and  $\mathbf{a}_{t,l}$  is the corresponding vector for the transmitter. This model is valid for all types of antenna arrangements, but if a uniform linear array (ULA) is used, the array response vector exhibits a Vandermonde structure

$$\mathbf{a}_{t,l}^T = [1 \quad e^{-j\Omega_{t,l}} \quad e^{-j2\Omega_{t,l}} \quad \dots \quad e^{-j(M_t-1)\Omega_{t,l}}] \quad (3)$$

where the angular frequency of path  $l$  is modeled as  $\Omega_{t,l} = kd_t \sin \phi_{t,l}$ ,  $k$  is the wave number,<sup>1</sup>  $d_t$  is the element separation distance, and  $\phi_{t,l}$  is the direction of departure (DOD) of path  $l$  at the transmitter. At the receiver, the direction of arrival (DOA)  $\Omega_{r,l}$  is defined in a similar manner.

The summation in (2) can also be written using a matrix formulation as

$$\mathbf{H}(t) = \sum_{l=1}^L \alpha_l \mathbf{a}_{r,l} \mathbf{a}_{t,l}^T e^{j\omega_l t} = \mathbf{A}_r \mathbf{P} \mathbf{A}_t^T = \mathbf{A}_r \mathbf{X} \mathbf{W}(t) \mathbf{A}_t^T \quad (4)$$

where

$$\mathbf{P}(t) = \text{diag}(\alpha_1 e^{j\omega_1 t}, \alpha_2 e^{j\omega_2 t}, \dots, \alpha_L e^{j\omega_L t}) \quad (5)$$

$$\mathbf{X} = \text{diag}(\alpha_1, \alpha_2, \dots, \alpha_L) \quad (6)$$

$$\mathbf{W}(t) = \text{diag}(e^{j\omega_1 t}, e^{j\omega_2 t}, \dots, e^{j\omega_L t}). \quad (7)$$

<sup>1</sup>The wave number  $k$  is defined as  $k = 2\pi/\lambda$ , where  $\lambda$  denotes the wavelength at the frequency of operation.

Here,  $\mathbf{A}_r$  and  $\mathbf{A}_t$  are  $M_r \times L$  and  $M_t \times L$  matrices whose columns correspond, respectively, to the receive and transmit array response vectors for all paths. For the following analysis, it is convenient to vectorize<sup>2</sup> the above expression:

$$\begin{aligned} \mathbf{h}(t) &= \text{vec}[\mathbf{H}(t)] = (\mathbf{A}_t \otimes \mathbf{A}_r)(\mathbf{I}_L \otimes \mathbf{X})\text{vec}[\mathbf{W}(t)] \\ &= (\mathbf{A}_t \otimes \mathbf{A}_r \mathbf{X})\text{vec}[\mathbf{W}(t)]. \end{aligned} \quad (8)$$

Here,  $\otimes$  denotes the Kronecker product.<sup>3</sup>

For the purpose of studying channel prediction, it is assumed that a series of previous channel matrices is known. Transmitting a training sequence or using joint channel and symbol detection are two different ways of obtaining previous channel realizations. In practice, of course, the channel estimates will be imperfect due to the effects of noise and interference. We will model the observed or estimated channel at time  $t$ ,  $\hat{\mathbf{h}}(t)$ , as the sum of the true channel and a Gaussian “noise” term due to the estimation error

$$\tilde{\mathbf{h}}(t) = \mathbf{h}(t) + \mathbf{n}(t) \quad (9)$$

where the  $M_t M_r \times 1$  vector  $\mathbf{n}(t)$  denotes the Gaussian noise. Hence,  $\tilde{\mathbf{h}}(t) \sim \mathcal{CN}(\mathbf{h}(t), \mathbf{C})$ . It is further assumed that the noise due to estimation error is spatially and temporally white, i.e.,  $\mathbf{C} = \sigma \mathbf{I}$ , where  $\sigma$  denotes the variance. This model holds when unitary training signals are used to estimate the channel, a reasonable assumption since it has been shown that such a training scheme maximizes capacity [8].

To further justify our model for  $\mathbf{n}(t)$  and other assumptions that will be made later, assume that during training interval  $t$ , the transmitter sends an  $M_t \times T$  matrix of data  $\mathbf{S}(t)$ , where  $T \geq M_t$  and  $\mathbf{S}(t)\mathbf{S}(t)^H = a\mathbf{I}_{M_t}$  for some  $a > 0$ . Typically, the average power transmitted per unit time is constrained to be less than some value  $\rho$ . Thus

$$\frac{1}{T} \text{Tr}[\mathbf{S}\mathbf{S}^H] = \frac{aM_t}{T} \leq \rho \quad (10)$$

and assuming we transmit the training data at full power, we have  $a = \rho T/M_t$ . If  $T$  is small enough so that the channel can be modeled as constant over the training interval, then the received data can be expressed as

$$\mathbf{Y}(t) = \mathbf{H}(t)\mathbf{S}(t) + \mathbf{Z}(t) \quad (11)$$

where  $\mathbf{Z}(t)$  is additive noise present during training and is assumed to be composed of zero-mean independent identically

<sup>2</sup>The vectorization operator  $\text{vec}(\mathbf{A})$  stacks the columns of  $\mathbf{A}$  into a vector of length  $mn$  if  $\mathbf{A}$  is  $m \times n$ .

<sup>3</sup>The Kronecker product  $\mathbf{A} \otimes \mathbf{B}$  is defined as

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \mathbf{A}_{1,1}\mathbf{B} & \mathbf{A}_{1,2}\mathbf{B} & \dots & \mathbf{A}_{1,n}\mathbf{B} \\ \mathbf{A}_{2,1}\mathbf{B} & \mathbf{A}_{2,2}\mathbf{B} & \dots & \mathbf{A}_{2,n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{m,1}\mathbf{B} & \mathbf{A}_{m,2}\mathbf{B} & \dots & \mathbf{A}_{m,n}\mathbf{B} \end{bmatrix}.$$

distributed (i.i.d.) elements of variance  $\sigma_z$ . The maximum-likelihood channel estimate is then simply given by

$$\hat{\mathbf{H}}(t) = \frac{1}{\sqrt{a}} \mathbf{Y}(t) \mathbf{S}(t)^H = \mathbf{H}(t) + \frac{1}{\sqrt{a}} \mathbf{Z}(t) \mathbf{S}(t)^H \quad (12)$$

and hence  $\mathbf{n}(t) = \text{vec}[\mathbf{Z}(t) \mathbf{S}(t)^H / \sqrt{a}]$ . It is straightforward to show that the elements of  $\mathbf{n}(t)$  are uncorrelated, with equal variance given by

$$\sigma = \frac{\sigma_z}{a} = \frac{M_t \sigma_z}{\rho T}. \quad (13)$$

An important observation to be made here is that, if the number of training symbols  $T$  is proportional to the number of transmit antennas  $M_t$  (a reasonable assumption), then  $\sigma$  is independent of the dimension of  $\mathbf{H}$ .

Sampling or estimating the channel matrix  $N_M$  times gives the following stacked matrix model:

$$\tilde{\mathbf{h}} = \begin{bmatrix} \tilde{\mathbf{h}}(t_1) \\ \tilde{\mathbf{h}}(t_2) \\ \vdots \\ \tilde{\mathbf{h}}(t_{N_M}) \end{bmatrix} = \begin{bmatrix} (\mathbf{A}_t \otimes \mathbf{A}_r \mathbf{X}) \text{vec}[\mathbf{W}(t_1)] \\ (\mathbf{A}_t \otimes \mathbf{A}_r \mathbf{X}) \text{vec}[\mathbf{W}(t_2)] \\ \vdots \\ (\mathbf{A}_t \otimes \mathbf{A}_r \mathbf{X}) \text{vec}[\mathbf{W}(t_{N_M})] \end{bmatrix} + \mathbf{n} \\ = (\mathbf{I}_{N_M} \otimes (\mathbf{A}_t \otimes \mathbf{A}_r \mathbf{X})) \mathbf{W} + \mathbf{n} = \mathbf{h} + \mathbf{n} \quad (14)$$

where  $\mathbf{h} = (\mathbf{A}_t \otimes \mathbf{A}_r \mathbf{X})$ , the noise term  $\mathbf{n}$  represents stacked vector samples of the channel estimation error (formed similarly to  $\tilde{\mathbf{h}}$ ), and

$$\mathbf{W} = \begin{bmatrix} \text{vec}[\mathbf{W}(t_1)] \\ \text{vec}[\mathbf{W}(t_2)] \\ \vdots \\ \text{vec}[\mathbf{W}(t_{N_M})] \end{bmatrix}. \quad (15)$$

Note that  $\mathbf{h}$  is an  $M_t M_r N_M \times 1$  vector and that the model in (14) represents a parameterized channel model with parameters  $\boldsymbol{\alpha}, \boldsymbol{\omega}, \boldsymbol{\Omega}_t, \boldsymbol{\Omega}_r$ , and noise variance  $\sigma$ . We define  $\boldsymbol{\alpha}^T = [\alpha_1, \alpha_2, \dots, \alpha_L]$  and  $\boldsymbol{\omega}, \boldsymbol{\Omega}_t$ , and  $\boldsymbol{\Omega}_r$  similarly. For analysis purposes, all the parameters will be stacked into one real-valued parameter vector  $\boldsymbol{\theta}^T = [\sigma, \text{Re}[\boldsymbol{\alpha}^T], \text{Im}[\boldsymbol{\alpha}^T], \boldsymbol{\omega}^T, \boldsymbol{\Omega}_t^T, \boldsymbol{\Omega}_r^T]$ . Except for the scalar  $\sigma$ , all the parameters are  $L \times 1$  vectors, making the parameter vector  $\boldsymbol{\theta}$  a  $(5L + 1) \times 1$  vector.

### III. LOWER BOUND ON PREDICTION ERROR

We assume that  $N_M$  samples from (9) are available for use in predicting the MIMO channel. The channel samples  $\{\tilde{\mathbf{h}}(t_1), \dots, \tilde{\mathbf{h}}(t_{N_M})\}$  are called the measurement segment and used to estimate the channel in the prediction segment  $\{\hat{\mathbf{h}}(t_{N_M+1}), \dots, \hat{\mathbf{h}}(t_{N_M+N_P})\}$ , where we use  $\hat{\cdot}$  to denote an estimate. For purposes of deriving a lower bound on the prediction error, we will assume that the predicted channel estimates are obtained first by estimating the model parameters  $\boldsymbol{\theta}$  using data from the measurement segment and second by using the estimated parameters to extrapolate the model in (2). In practice, it is unlikely that such a method would be used for prediction since, unless  $L$  is small, estimating all of the parameters in  $\boldsymbol{\theta}$  is computationally problematic. Our primary

goal is not to suggest a particular algorithm for prediction; rather, we are interested in the best possible performance any predictor might hope to achieve. The ray-based model we are using simply serves as a convenient way of describing the time evolution of the MIMO channel.

The estimation/prediction error at any time  $t$ , given estimated parameters  $\hat{\boldsymbol{\theta}}$ , can be expressed using (2) as

$$\mathbf{e}(t) = \hat{\mathbf{h}}(t) - \mathbf{h}(t) = \mathbf{h}(t, \hat{\boldsymbol{\theta}}) - \mathbf{h}(t, \boldsymbol{\theta}) \\ = \sum_{l=1}^L (\text{vec}(\hat{\mathbf{a}}_{r,l} \hat{\mathbf{a}}_{t,l}^T) \hat{\alpha}_l e^{j\hat{\omega}_l t} - \text{vec}(\mathbf{a}_{r,l} \mathbf{a}_{t,l}^T) \alpha_l e^{j\omega_l t}) \\ = \sum_{l=1}^L (\hat{\mathbf{a}}_{t,l} \otimes \hat{\mathbf{a}}_{r,l} \hat{\alpha}_l e^{j\hat{\omega}_l t} - \mathbf{a}_{t,l} \otimes \mathbf{a}_{r,l} \alpha_l e^{j\omega_l t}). \quad (16)$$

Observing that the channel simply represents a nonlinear function of the parameters  $\boldsymbol{\theta}$ , a lower bound on the covariance matrix of any unbiased estimator  $\hat{\mathbf{h}}$  can be found using the Cramér–Rao lower bound (CRB). Using a vector formulation of the CRB for functions of parameters [11], the bound can be written as

$$E[(\hat{\mathbf{h}}(t) - \mathbf{h}(t))(\hat{\mathbf{h}}(t) - \mathbf{h}(t))^H] \geq \mathbf{H}' \mathbf{B} \mathbf{H}'^H \quad (17)$$

where the matrix inequality  $\mathbf{A} - \mathbf{B} > \mathbf{0}$  means that the matrix difference  $\mathbf{A} - \mathbf{B}$  is positive definite,  $\mathbf{B}$  denotes the matrix defining the CRB of the parameters  $\boldsymbol{\theta}$ , and  $\mathbf{H}'$  is the  $(M_t M_r) \times (5L + 1)$  Jacobian matrix

$$\mathbf{H}' = \begin{bmatrix} \frac{\partial \mathbf{h}(t)}{\partial \theta_1} & \frac{\partial \mathbf{h}(t)}{\partial \theta_2} & \dots & \frac{\partial \mathbf{h}(t)}{\partial \theta_{5L+1}} \end{bmatrix}. \quad (18)$$

It is straightforward but tedious to evaluate all the derivatives used in forming the above matrix (see Appendix A for the details). The CRB for  $\boldsymbol{\theta}$  can be calculated using Bang's formula [3], which results in a compact derivation given the above model and parameters. Using Bang's formula, the CRB matrix can be expressed as

$$[\mathbf{B}^{-1}]_{ij} = \text{Tr} \left[ \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_i} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_j} \right] \\ + 2 \text{Re} \left[ \frac{\partial \mathbf{h}^H}{\partial \theta_i} \mathbf{C}^{-1} \frac{\partial \mathbf{h}}{\partial \theta_j} \right] = \mathbf{J}(\boldsymbol{\theta})_{ij} \quad (19)$$

where  $\mathbf{J}(\boldsymbol{\theta})$  is the Fisher matrix. Inserting expressions for the derivatives, a block structure for the Fisher matrix  $\mathbf{J}(\boldsymbol{\theta})$  results, as shown in Appendix B. By collecting the expressions for the individual blocks of the Fisher matrix, the block structure of the total Fisher matrix can be written as (20), shown at the bottom of the next page. The expressions for  $\mathbf{P}_1$ – $\mathbf{P}_{10}$  are given by

$$\mathbf{P}_1 = \frac{2}{\sigma} \check{\mathbf{W}}^H (\mathbf{I}_{N_M} \otimes \mathbf{A}_t^H \mathbf{A}_t \odot \mathbf{A}_r^H \mathbf{A}_r) \check{\mathbf{W}} \quad (21)$$

$$\mathbf{P}_2 = \frac{2}{\sigma} \check{\mathbf{W}}^H (\mathbf{I}_{N_M} \otimes \mathbf{A}_t^H \mathbf{A}_t \odot \mathbf{A}_r^H \mathbf{A}_r \mathbf{X}) \check{\mathbf{D}}_{\boldsymbol{\omega}} \quad (22)$$

$$\mathbf{P}_3 = \frac{2}{\sigma} \check{\mathbf{W}}^H (\mathbf{I}_{N_M} \otimes \mathbf{A}_t^H \mathbf{D}_t \odot \mathbf{A}_r^H \mathbf{A}_r \mathbf{X}) \check{\mathbf{W}} \quad (23)$$

$$\mathbf{P}_4 = \frac{2}{\sigma} \check{\mathbf{W}}^H (\mathbf{I}_{N_M} \otimes \mathbf{A}_t^H \mathbf{A}_t \odot \mathbf{A}_r^H \mathbf{D}_r \mathbf{X}) \check{\mathbf{W}} \quad (24)$$

$$\mathbf{P}_5 = \frac{2}{\sigma} \check{\mathbf{D}}_\omega^H (\mathbf{I}_{N_M} \otimes \mathbf{A}_t^H \mathbf{A}_t \odot \mathbf{X}^H \mathbf{A}_r^H \mathbf{A}_r \mathbf{X}) \check{\mathbf{D}}_\omega \quad (25)$$

$$\mathbf{P}_6 = \frac{2}{\sigma} \check{\mathbf{D}}_\omega^H (\mathbf{I}_{N_M} \otimes \mathbf{A}_t^H \mathbf{D}_t \odot \mathbf{X}^H \mathbf{A}_r^H \mathbf{A}_r \mathbf{X}) \check{\mathbf{W}} \quad (26)$$

$$\mathbf{P}_7 = \frac{2}{\sigma} \check{\mathbf{D}}_\omega^H (\mathbf{I}_{N_M} \otimes \mathbf{A}_t^H \mathbf{A}_t \odot \mathbf{X}^H \mathbf{A}_r^H \mathbf{D}_r \mathbf{X}) \check{\mathbf{W}} \quad (27)$$

$$\mathbf{P}_8 = \frac{2}{\sigma} \check{\mathbf{W}}^H (\mathbf{I}_{N_M} \otimes \mathbf{D}_t^H \mathbf{D}_t \odot \mathbf{X}^H \mathbf{A}_r^H \mathbf{A}_r \mathbf{X}) \check{\mathbf{W}} \quad (28)$$

$$\mathbf{P}_9 = \frac{2}{\sigma} \check{\mathbf{W}}^H (\mathbf{I}_{N_M} \otimes \mathbf{D}_t^H \mathbf{A}_t \odot \mathbf{X}^H \mathbf{A}_r^H \mathbf{D}_r \mathbf{X}) \check{\mathbf{W}} \quad (29)$$

$$\mathbf{P}_{10} = \frac{2}{\sigma} \check{\mathbf{W}}^H (\mathbf{I}_{N_M} \otimes \mathbf{A}_t^H \mathbf{A}_t \odot \mathbf{X}^H \mathbf{D}_r^H \mathbf{D}_r \mathbf{X}) \check{\mathbf{W}} \quad (30)$$

where  $\odot$  denotes an element-wise matrix product (Hadamard product),  $\text{Re}[\cdot]$  and  $\text{Im}[\cdot]$  denote real and imaginary parts, respectively, and the matrices  $\check{\mathbf{W}}$ ,  $\check{\mathbf{D}}_\omega$ ,  $\mathbf{D}_t$ , and  $\mathbf{D}_r$  are defined in Appendices A and B.

Finally, the sum of the variances of the estimation/prediction errors for the entries of the channel matrix can be bounded as

$$E[\|\mathbf{e}(t)\|_F^2] \geq \text{Tr}[\mathbf{H}'\mathbf{B}\mathbf{H}'^H] = \varepsilon_0(t) \quad (31)$$

where  $\|\cdot\|_F$  denotes the Frobenius norm,  $\mathbf{B}$  is the CRB expression obtained from (20), and  $\mathbf{H}'$  is given in (18). Note that the error expression in (31) is not limited to the prediction segment but can be applied to any  $t$  once the model parameters  $\boldsymbol{\theta}$  are estimated (for example, a bound on a ‘‘smoothed’’ estimate of the channel during the measurement segment can be calculated). Hence, the above expression can also be used to quantify the estimation error as a function of any model parameter such as observation length or number of receive antennas.

#### IV. NUMERICAL EVALUATION

In this section, the bound on channel prediction derived above is evaluated for several different scenarios. While several performance measures have been proposed in the literature for SISO channel prediction [1], [20], the introduction of multiple sub-channels in the MIMO case requires a new performance measure. A natural criterion in the MIMO case is the following root-mean-square-error (RMSE) measure

$$\sqrt{\frac{E[\|\mathbf{e}(t+t_Q)\|_F^2]}{E[\|\mathbf{H}(t+t_Q)\|_F^2]}} \geq \sqrt{\frac{\varepsilon_0(t+t_Q)}{E[\|\mathbf{H}(t+t_Q)\|_F^2]}} = \varepsilon(Q) \quad (32)$$

where  $t_Q$  denotes the time at which the channel is predicted. The error measure essentially represents a bound on the average error of all the elements of the channel matrix  $\mathbf{H}$  so that the

error levels for the MIMO and SISO cases can be fairly compared. Our numerical evaluations of the prediction length will be given in units of distance (wavelengths) rather than time, since this allows for direct comparisons to earlier work in SISO channel prediction and makes our results independent of the mobile velocity. In other words, the fundamental question of how far ahead the channel can be predicted will be expressed in wavelengths using the bound and the channel model of the previous sections. Assuming a given mobile velocity, it is straightforward to translate the prediction horizon  $Q$  into time.

#### A. Channel Parameters

To reduce the complexity of the model, it is assumed that the different parameters

$$\boldsymbol{\theta}^T = [\sigma, \text{Re}[\boldsymbol{\alpha}^T], \text{Im}[\boldsymbol{\alpha}^T], \boldsymbol{\omega}^T, \boldsymbol{\Omega}_t^T, \boldsymbol{\Omega}_r^T]$$

are independent, which is a reasonable assumption. Furthermore, the scattering parameters  $\boldsymbol{\alpha}$  are assumed to be Gaussian distributed:  $\alpha_i \sim \mathcal{CN}(0, 1)$ . The Doppler frequency of path  $l$  can be derived from a physical viewpoint to be  $\omega_l = k\Delta_x \sin \phi_l = 2\pi(\Delta_x/\lambda) \sin \phi_l$ , where  $\Delta_x$  is the distance separating consecutive channel samples. Since the numerical evaluations are performed in the spatial domain, the separation distance  $\Delta_x$  is specified instead of defining the Doppler frequencies. A uniform distribution for the angle between propagation path  $l$  and the direction of travel  $\phi_l \sim U[0, 2\pi)$  is further assumed. Finally, ULAs are assumed at both the transmitter and receiver with element separation distances  $d_t$  and  $d_r$ . The DOD and DOA for path  $l$  are defined as  $\Omega_{t,l} = kd_t \sin \phi_{t,l}$  and  $\Omega_{r,l} = kd_r \sin \phi_{r,l}$ , where the angles are assumed to be uniformly distributed as  $\phi_{t,l}, \phi_{r,l} \sim U[0, 2\pi)$ . Of course, other angular densities can also be explored but a uniform distribution is assumed here to reduce the number of scenario parameters.

With the above parameter distributions, the average channel power becomes  $E[\|\mathbf{H}(t)\|_F^2] = LM_t M_r$ , reducing the expression for the normalized prediction error to

$$\varepsilon(Q) = \sqrt{\frac{\text{Tr}[\mathbf{H}'\mathbf{B}\mathbf{H}'^H]}{LM_t M_r}}. \quad (33)$$

Furthermore, the SNR in the following simulations is defined from (9) to be

$$\text{SNR} = \frac{E[\|\mathbf{h}(t)\|^2]}{E[\|\mathbf{n}(t)\|^2]} = \frac{L}{\sigma}. \quad (34)$$

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{M_t M_r N_M}{\sigma^2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \text{Re}[\mathbf{P}_1] & -\text{Im}[\mathbf{P}_1] & \text{Re}[\mathbf{P}_2] & \text{Re}[\mathbf{P}_3] & \text{Re}[\mathbf{P}_4] \\ 0 & \text{Im}[\mathbf{P}_1] & \text{Re}[\mathbf{P}_1] & \text{Im}[\mathbf{P}_2] & \text{Im}[\mathbf{P}_3] & \text{Im}[\mathbf{P}_4] \\ 0 & \text{Re}[\mathbf{P}_2^T] & \text{Im}[\mathbf{P}_2^T] & \mathbf{P}_5 & \mathbf{P}_6 & \mathbf{P}_7 \\ 0 & \text{Re}[\mathbf{P}_3^T] & \text{Im}[\mathbf{P}_3^T] & \mathbf{P}_6^T & \mathbf{P}_8 & \mathbf{P}_9 \\ 0 & \text{Re}[\mathbf{P}_4^T] & \text{Im}[\mathbf{P}_4^T] & \mathbf{P}_7^T & \mathbf{P}_9^T & \mathbf{P}_{10} \end{bmatrix}. \quad (20)$$

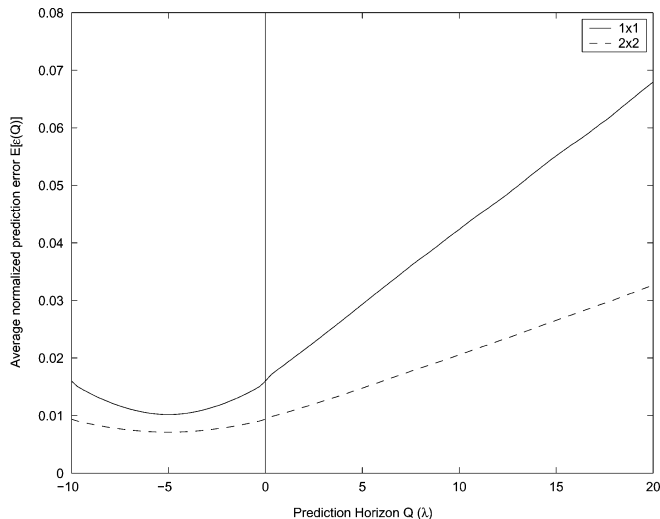


Fig. 1. Average normalized prediction error  $\epsilon(Q)$  for a SISO system and a  $M_t = M_r = 2$  MIMO system versus prediction horizon.

However, to keep a fixed signal-to-noise ratio (SNR) independent of the number of paths in the channel, the additive noise would need to be changed accordingly. Therefore, to avoid this, the noise power will be held fixed instead of the SNR.

### B. A Two Path Scenario

To display the fundamental behavior of the prediction bound, a channel scenario with only two propagation paths will be considered first. A noise power of  $-20$  dB is assumed, and the element separation distance at both transmitter and receiver is one half-wavelength, i.e.,  $d_t = d_r = \lambda/2$ . The parameter values were generated randomly according to the distributions given above and the results shown represent the average of 500 channel realizations. Furthermore, the measurement length consists of  $N_M = 100$  samples taken uniformly over a distance of  $10\lambda$ .

The average normalized prediction error  $E[\epsilon(Q)]$  for a SISO system and a  $M_t = M_r = 2$  MIMO system is shown in Fig. 1 for various prediction horizons  $Q$  between 0 and  $20\lambda$ . Note that the averaging is over the 500 different channel realizations and that negative  $Q$  values correspond to estimation errors instead of prediction errors. It is clear that the bound grows rapidly with the prediction horizon. Furthermore, it is reasonable that the smallest error occurs in the middle of the measurement segment since both future and past channel samples contribute to a lower error in that case. Increasing the length of the measurement segment or decreasing the measurement noise  $\sigma$  will lower the prediction error bound. It is also clear that the bound is significantly reduced for the  $2 \times 2$  system. This indicates that the drawback of having to estimate the angular directions for the  $2 \times 2$  system (a reasonable task for  $L = 2$ ) is compensated for by the fact that the  $2 \times 2$  system has three more data points per channel snapshot than the  $1 \times 1$  system.

### C. Ten-Path Scenario

A more realistic channel model will consist of more multipath components, i.e., a larger value for  $L$ . Previous investigations [10] have shown that a limited number of complex sinusoids

is all that is required to reproduce a Rayleigh fading channel. Furthermore, it was reported in [2] that outdoor measurements in downtown Austin, TX, were in general well described with three to eight multipath components. Here, a channel with  $L = 10$  will be considered for channel prediction.

Generating Doppler frequencies, DODs and DOAs at random give rise to an identifiability problem. If two paths have very similar parameters, the CRB matrix will be poorly conditioned, and the estimation procedure will fail. Finding identifiability conditions for this problem is in general fairly difficult and is still an open research problem. However, for practical purposes, two paths with similar parameters may be combined into one with reasonable prediction performance. The analysis of the performance bound for SISO systems in [20] essentially handled this problem by removing paths until the CRB matrix was full rank. Fortunately, identifiability is less of a problem in the MIMO case since more observables are available. For instance, for each matrix channel sample, four scalar samples are obtained for a  $2 \times 2$  system and nine for a  $3 \times 3$  case, but in both situations the number of unknown channel parameters remain the same. Simulation results using ten paths indicate that identifiability is only a problem for SISO systems, and even for  $2 \times 2$  systems, it is very rare to obtain a channel realization with a rank-deficient CRB matrix. In this paper, identifiability issues are avoided by removing 5% of the channel realizations whose CRB matrix had the highest condition number.

The prediction length  $Q_\epsilon$  is defined as the maximum prediction horizon  $Q$  with an average normalized prediction error less than 0.05, i.e.,  $E[\epsilon(Q_\epsilon)] < 0.05$ , and is shown in Fig. 2 versus measurement length. Here, the measurement length is increased while keeping the number of samples constant, corresponding to increasing the distance  $\Delta_x$  between channel samples from 0.02–0.10  $\lambda$ . Again, the results represent the average of 500 channel realizations with a noise power of  $-20$  dB and  $N_M = 100$ .

It is clear that increasing the measurement length decreases the prediction error, and hence also increases the prediction length  $Q_\epsilon$ . However, the more interesting result is that the prediction length is substantially longer for MIMO systems than for the corresponding SISO system. This is especially important for practical MIMO channel prediction where the channel parameters are usually time varying, limiting the measurement length. For example, SISO prediction is possible only 0.07  $\lambda$  ahead, while for a  $3 \times 3$  system the corresponding distance is  $12\lambda$  for a measurement length of  $6\lambda$ . This performance gain is achieved since more of the spatial structure of the channel is revealed by using more antennas. In fact, going from a SISO to a  $3 \times 3$  system does not require more than  $2L$  extra parameters to be estimated, while at the same time nine samples instead of one is obtained for each channel snapshot. Hence, it is expected that the prediction performance should improve. Similar effects are also visible when using an array only at the receiver [single-input multi-output (SIMO) systems], as shown in Fig. 2. However, the performance gains are obviously smaller since less observations are available.

The prediction performance comparison in Fig. 2 is based on having access to  $N_M$  channel estimates for each system, regardless of the number of available antennas. Although this

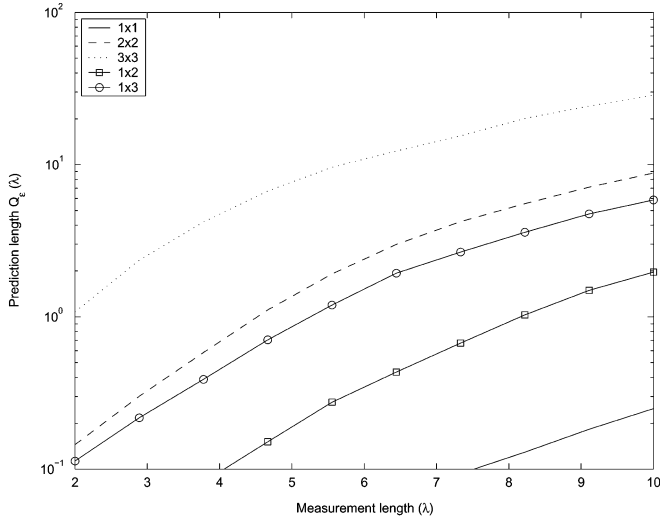


Fig. 2. Prediction length  $Q_\varepsilon$  for various MIMO systems versus measurement length.

is sufficient for studying the fundamental distance that the channel can be predicted, the number of available channel estimates may vary depending on the number of antennas. For instance, longer training sequences are needed as the number of transmit antennas is increased. Fortunately, for the purpose of channel prediction, the channel needs to be sampled fairly infrequently compared to the data rate of the system. Thus, for many systems, the channel can easily be sampled frequently enough for the prediction to work properly. However, to investigate the impact of the longer training sequences limiting the number of channel estimates, the scenario corresponding to Fig. 2 is repeated assuming the number of available channel estimates is proportional to  $1/M_t$ . Hence, the SISO system has 100 measurement samples, the  $2 \times 2$  has 50 samples, and the  $3 \times 3$  system has 33 samples. The results are presented in Fig. 3. Although the prediction lengths for the  $2 \times 2$  and  $3 \times 3$  systems are reduced in Fig. 3 compared with Fig. 2, they are still substantially longer than the corresponding prediction lengths for the SISO and SIMO systems. An intuitive explanation for this is that once the measurement length is long enough for prediction to work, sampling more densely only reduces the impact of the noise. However, the larger impact of the noise is more than compensated for by the additional channel structure revealed by the availability of multiple antennas. Thus, even when the number of channel estimates are reduced proportional to the number of transmit antennas, evaluation of the prediction bound indicates longer prediction lengths for MIMO systems.

The ability to predict the channel also depends strongly on the number of paths since the channel structure becomes more complicated and resolving the individual paths is more difficult. The prediction length  $Q_\varepsilon$  versus the number of paths is shown in Fig. 4 for various numbers of antennas. One hundred channel samples were collected over a measurement length of  $5\lambda$ , and the results represent the average of 500 channel realizations with a noise power of  $-20$  dB. It is evident that performance drops rapidly as  $L$  is increased. Hence, environments with significant multipath such as indoor environments will likely be harder to predict than outdoor environments with less multipath. Again, it

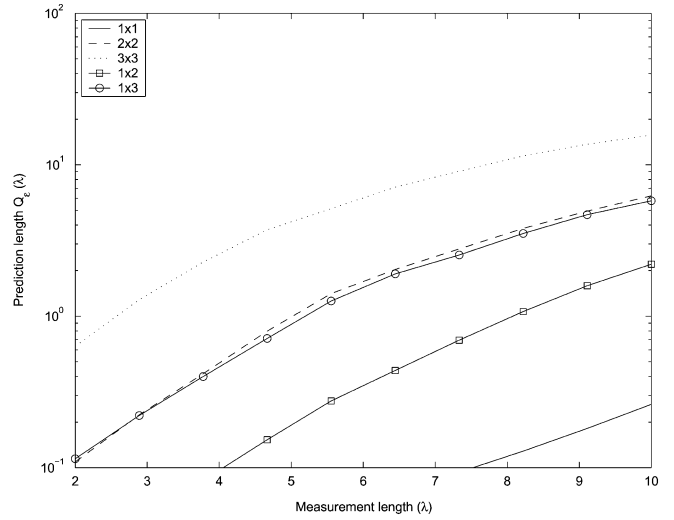


Fig. 3. Prediction length  $Q_\varepsilon$  for various MIMO systems versus measurement length with sample sizes proportional to  $1/M_t$ .

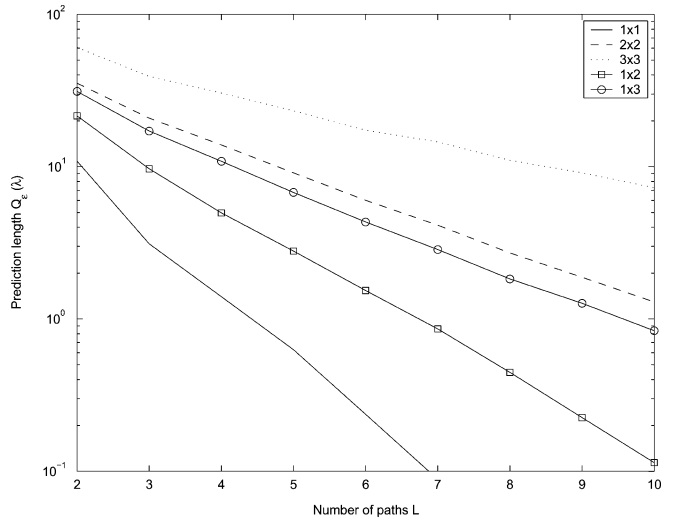


Fig. 4. Prediction length  $Q_\varepsilon$  for various MIMO systems versus number of paths.

is clear that employing antennas at both transmit and receive can substantially increase prediction performance. Just employing an array at one end provides some performance gain but less than the dual-array case. Similar results were obtained when the number of channel estimates were inversely proportional to the number of transmit antennas, as shown in Fig. 5.

## V. PREDICTION EXAMPLE

The above analysis of the prediction bound has indicated that the prediction performance for MIMO systems may exceed that of SISO systems. However, finding predictors that achieve the bound at a reasonable cost is still an open research problem. In principle, a maximum-likelihood (ML) detector that achieves the bound could be designed, but due to the nonlinear dependence of the model on  $\theta$ , a large dimensional search is necessary. Lower complexity predictors that achieve the bulk of the ML predictor performance at a much lower computational cost are therefore of significant interest. However, this is beyond the scope of the current paper. Instead, a simple predictor that has

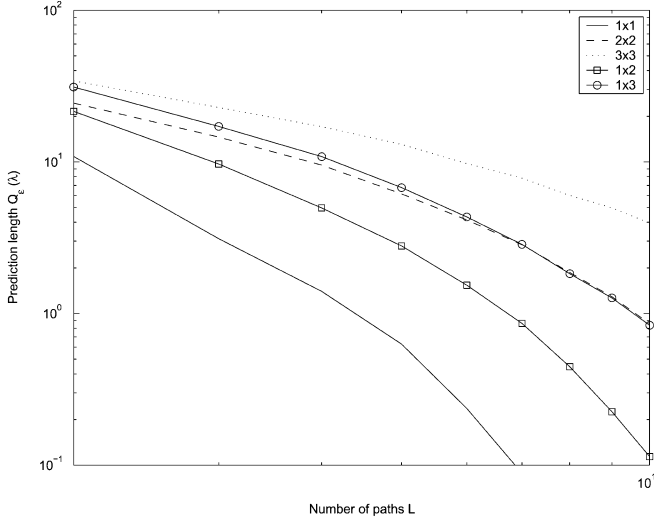


Fig. 5. Prediction length  $Q_\varepsilon$  for various MIMO systems versus number of paths with sample sizes proportional to  $1/M_t$ .

commonly been used for SISO prediction will be investigated and extended to the MIMO case to indicate that the prediction performance indeed can be increased using multiple antennas.

A common approach for channel prediction is to assume that each channel coefficient obeys the following autoregressive (AR) model [5], [6]:

$$h_i(t) + \sum_{n=1}^{N_a} a_{i,n} h_i(t-n) = e_i(t) \quad (35)$$

where the  $N_a$  coefficients  $a_{i,n}$  define the filter for the  $i$ th channel coefficient, and  $e_i(t)$  is the corresponding residual white noise. The prediction is then easily calculated by extrapolating an AR model that is fit to the training data. The AR approach has been found to be a reasonable model for SISO channels with single clusters of scatterers [6]. As a simple direct extension to the MIMO case, a separate AR filter could be estimated for each channel coefficient. However, the benefit of having multiple channels is not exploited in this case, and the prediction performance is thus the same as for the SISO case.

In order to try to exploit the additional information available by sampling the ray structure of the channel using multiple antennas, the above predictor can be applied to a transformed version of the channel matrix  $\mathbf{H}$ . The transformation should be defined so as to maximize the prediction performance. It is beyond the scope of this paper to find the optimal transformation, but a simple transformation will be used that is based on the important observation that each channel coefficient represents contributions from many rays. If a transformation could be defined that reduced the number of rays “present” in each coefficient, prediction performance would improve.

One reasonable solution is obtained by projecting the channel matrix onto the subspace spanned by a set of outer products of array responses that form a set of beams spanning the entire space. Essentially, the received data is just projected into a number of angular regions. Within each angular region, there

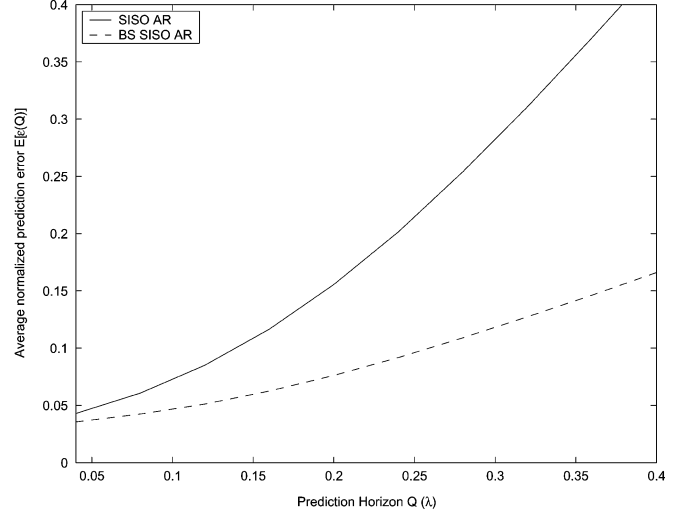


Fig. 6. Average normalized prediction error  $E[\varepsilon(Q)]$  versus prediction horizon  $Q$  using a SISO AR predictor and a beamspace version for a ten-path scenario.

are now only a few rays, and prediction improves since the transformed signal is less complicated. The transformed signal  $\mathbf{Y}(t)$  can be written as

$$\mathbf{Y}(t) = \mathbf{W}_r^H \mathbf{H}(t) \mathbf{W}_t \quad (36)$$

where, for example, the beamspace transformation matrix for a ULA can be written as

$$\mathbf{W}_r^T = \begin{bmatrix} 1 & e^{j\Omega_{r,1}} & e^{j2\Omega_{r,1}} & \dots & e^{j(M_r-1)\Omega_{r,1}} \\ 1 & e^{j\Omega_{r,2}} & e^{j2\Omega_{r,2}} & \dots & e^{j(M_r-1)\Omega_{r,2}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{j\Omega_{r,M_r}} & e^{j2\Omega_{r,M_r}} & \dots & e^{j(M_r-1)\Omega_{r,M_r}} \end{bmatrix}. \quad (37)$$

$\Omega_{r,i} = kd_r \sin \phi_{r,i}$  where  $\phi_{r,i} = 2\pi(i-1)/M_r$ ,  $i \in [1, M_r]$ , and  $\mathbf{W}_t$  is defined similarly. Hence, the prediction step is applied directly to  $\mathbf{Y}$  instead of  $\mathbf{H}$ , and once the predictions of  $\mathbf{Y}$  are calculated, the inverse transform of (36) is applied to obtain the predictions of  $\mathbf{H}$ .

Fig. 6 shows the results using the standard AR approach (a separate AR model for each channel coefficient) and the projection or beamspace version averaged over 100 channel realizations. The predictor is trained using data over  $8\lambda$  with  $N_M = 200$  ( $\Delta_x/\lambda = 0.04$ ), the noise power is  $-20$  dB, ten antennas are used at both ends, and the order of the AR predictor is  $N_a = 20$ . Prediction performance is then evaluated over the following  $2\lambda$  distance, and much better performance is obtained for the projection or beamspace version than the standard AR version. Hence, there is indeed a performance gain in exploiting the structure of the  $\mathbf{H}$  matrix. At the 0.05 error level, the prediction length essentially doubles when using a MIMO system. Still the performance is significantly lower than the performance promised by the analysis of the CRB bound, indicating that much more efficient prediction schemes exist.

## VI. DISCUSSION

Several authors have addressed channel prediction for SISO channels and found that only relatively short prediction lengths

are possible. Hence, the potential gain of employing channel prediction has been limited. In this paper, a performance bound has been derived for MIMO channel prediction based on a common channel model, and analysis of that bound indicates that MIMO channels can offer longer prediction lengths than traditional SISO channels. An intuitive interpretation of this result is that more of the channel structure is revealed when using multiple antennas. Contributing to the performance gain is the fact that more observables are available with only a modest increase in the number of unknowns. A simple SISO prediction scheme was extended to the MIMO case and a longer prediction length was observed but less than that predicted by the performance bound.

A significantly more complex prediction scheme that achieves the performance bound is to use a maximum-likelihood estimator to estimate the channel parameters and then simply interpolate the channel. However, more work is needed to find computationally efficient schemes that achieve or almost achieve the performance bound. Furthermore, initial experiments with measured data in [18] and [19] have shown that there is indeed a spatial structure to the measured data, but it appears that the beamspace AR scheme fails to capture it since limited performance gains over SISO systems have been observed. However, the ray-based model in (2) is widely accepted as a reasonable model for the spatial properties of the channel, so it seems likely that an appropriately designed predictor should be able to exploit the spatial structure.

There are several plausible reasons for the limited performance gains with measured data. The indoor environment is a complicated propagation environment with significant multipath. It is suspected that MIMO prediction would yield larger performance gains in more stable outdoor channels with less multipath. The outdoor environment is at the same time also more interesting for prediction applications due to the increased mobility associated with vehicular communications. Another reason for discrepancies with measured data is that the beamforming approach assumes ideal antennas and planar waves that may not be the case in indoor environments with walls and ceilings close to the antennas. Hence, developing robust prediction schemes that exploit the spatial structure of the channel is an interesting topic of future research. In addition, the performance bound derived in this paper can be used to investigate the gain of using prediction with current space-time coding schemes and its impact on channel capacity.

#### APPENDIX A DERIVATIVES FOR THE FISHER MATRIX

This section gives expressions for each derivative, i.e.,  $(\partial\mathbf{C})/(\partial\theta_i)$  and  $(\partial\mathbf{h})/(\partial\theta_j)$ , that are used to calculate the Fisher matrix in (19).

1) *Derivatives With Respect to  $\sigma$* : Only the covariance  $\mathbf{C}$  depends on  $\sigma$ , so all derivatives are zero except for  $(\partial\mathbf{C})/(\partial\theta_1) = (\partial\mathbf{C})/(\partial\sigma) = \mathbf{I}$ . Hence

$$\begin{aligned}\mathbf{J}(\boldsymbol{\theta})_{1,1} &= \text{Tr} \left[ \mathbf{C}^{-1} \frac{\partial\mathbf{C}}{\partial\theta_1} \mathbf{C}^{-1} \frac{\partial\mathbf{C}}{\partial\theta_1} \right] \\ &= \text{Tr}[\mathbf{I}/\sigma^2] = \frac{M_t M_r N_M}{\sigma^2}\end{aligned}\quad (38)$$

and  $\mathbf{J}(\boldsymbol{\theta})_{1,\cdot} = \mathbf{J}(\boldsymbol{\theta})_{\cdot,1} = 0$ .

2) *Derivatives With Respect to  $\text{Re}[\alpha_i]$* :

$$\begin{aligned}\frac{\partial\mathbf{h}}{\partial\text{Re}[\alpha_i]} &= \frac{\partial(\mathbf{I}_{N_M} \otimes (\mathbf{A}_t \otimes \mathbf{A}_r \mathbf{X})) \mathbf{W}}{\partial\text{Re}[\alpha_i]} \\ &= \left( \mathbf{I}_{N_M} \otimes \left( \mathbf{A}_t \otimes \mathbf{A}_r \frac{\partial\mathbf{X}}{\partial\text{Re}[\alpha_i]} \right) \right) \mathbf{W} \\ &= (\mathbf{I}_{N_M} \otimes \mathbf{A}_t \otimes \mathbf{A}_r \mathbf{O}_{j,j}) \mathbf{W}\end{aligned}\quad (39)$$

where  $\mathbf{O}_{j,j}$  denotes a matrix of all zeros except for a one at the  $j$ th element along the diagonal, and  $\mathbf{W}$  is given by (15).

3) *Derivatives With Respect to  $\text{Im}[\alpha_i]$* :

$$\begin{aligned}\frac{\partial\mathbf{h}}{\partial\text{Im}[\alpha_i]} &= \left( \mathbf{I}_{N_M} \otimes \left( \mathbf{A}_t \otimes \mathbf{A}_r \frac{\partial\mathbf{X}}{\partial\text{Im}[\alpha_i]} \right) \right) \mathbf{W} \\ &= j (\mathbf{I}_{N_M} \otimes \mathbf{A}_t \otimes \mathbf{A}_r \mathbf{O}_{j,j}) \mathbf{W} = j \frac{\partial\mathbf{h}}{\partial\text{Re}[\alpha_i]}\end{aligned}\quad (40)$$

4) *Derivatives With Respect to  $\omega$* :

$$\begin{aligned}\frac{\partial\mathbf{h}}{\partial\omega_j} &= (\mathbf{I}_{N_M} \otimes \mathbf{A}_t \otimes \mathbf{A}_r \mathbf{X}) \frac{\partial\mathbf{W}}{\partial\omega_j} \\ &= (\mathbf{I}_{N_M} \otimes \mathbf{A}_t \otimes \mathbf{A}_r \mathbf{X}) \mathbf{D}_{\omega_j}\end{aligned}\quad (41)$$

where

$$\begin{aligned}\mathbf{D}_{\omega_j} &= \begin{bmatrix} \text{vec}(\text{diag}(0, \dots, 0, jt_1 e^{j\omega_j t_1}, 0, \dots, 0)) \\ \text{vec}(\text{diag}(0, \dots, 0, jt_2 e^{j\omega_j t_2}, 0, \dots, 0)) \\ \vdots \\ \text{vec}(\text{diag}(0, \dots, 0, jt_{N_M} e^{j\omega_j t_{N_M}}, 0, \dots, 0)) \end{bmatrix} \\ &= \begin{bmatrix} \text{vec}(\text{diag}(\mathbf{D}_{\omega_j}(t_1) \mathbf{O}_{j,j})) \\ \text{vec}(\text{diag}(\mathbf{D}_{\omega_j}(t_2) \mathbf{O}_{j,j})) \\ \vdots \\ \text{vec}(\text{diag}(\mathbf{D}_{\omega_j}(t_{N_M}) \mathbf{O}_{j,j})) \end{bmatrix}\end{aligned}\quad (42)$$

and  $[\mathbf{D}_{\omega_j}(t_k)]_{m,n} = [(\partial\mathbf{W}(t_k))/(\partial\omega_j)]_{m,n}$ .

5) *Derivatives With Respect to  $\Omega_t$* :

$$\begin{aligned}\frac{\partial\mathbf{h}}{\partial\Omega_{t,j}} &= \left( \mathbf{I}_{N_M} \otimes \frac{\partial\mathbf{A}_t}{\partial\Omega_{t,j}} \otimes \mathbf{A}_r \mathbf{X} \right) \mathbf{W} \\ &= (\mathbf{I}_{N_M} \otimes \mathbf{D}_{t,j} \otimes \mathbf{A}_r \mathbf{X}) \mathbf{W}\end{aligned}\quad (43)$$

where

$$\mathbf{D}_{t,j} = \frac{\partial \mathbf{A}_t}{\partial \Omega_{t,j}} = \begin{bmatrix} \mathbf{0} \cdots \mathbf{0} & \frac{\partial [\mathbf{A}_t]_{:,j}}{\partial \Omega_{t,j}} & \mathbf{0} \cdots \mathbf{0} \end{bmatrix}. \quad (44)$$

Note that all columns except for the  $j$ th column are zero, and that the above expression is valid for all arrays. For the special case of a uniform linear array (ULA), an explicit expression for the derivative can be found using (3) as

$$\mathbf{D}_{t,j} = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ 0 & & j e^{j \Omega_{t,j}} & & 0 \\ 0 & & 2j e^{j 2 \Omega_{t,j}} & & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & j(M_t - 1) e^{j(M_t - 1) \Omega_{t,j}} & \cdots & 0 \end{bmatrix}. \quad (45)$$

6) Derivatives With Respect to  $\Omega_r$ :

$$\begin{aligned} \frac{\partial \mathbf{h}}{\partial \Omega_{r,j}} &= \left( \mathbf{I}_{N_M} \otimes \mathbf{A}_t \otimes \frac{\partial \mathbf{A}_r}{\partial \Omega_{r,j}} \mathbf{X} \right) \mathbf{W} \\ &= (\mathbf{I}_{N_M} \otimes \mathbf{A}_t \otimes \mathbf{D}_{r,j} \mathbf{X}) \mathbf{W} \end{aligned} \quad (46)$$

where  $\mathbf{D}_{r,j}$  is defined similarly to (44).

## APPENDIX B

### EVALUATION OF THE FISHER MATRIX BLOCKS

Explicit expressions for the Fisher matrix and the Cramér–Rao bound (CRB) are obtained by inserting the derivatives calculated in Appendix A into the CRB formula in (19). Furthermore, the derivatives with respect to the vector parameters  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\omega}$ ,  $\Omega_t$ , and  $\Omega_r$  will give rise to a block structure in the Fisher matrix. Expressions for the Fisher block corresponding to derivatives with respect to  $\text{Re}[\boldsymbol{\alpha}]$  will be calculated next.

Using (39), the following expression results:

$$\begin{aligned} 2\text{Re} \left[ \frac{\partial \mathbf{h}^H}{\partial \text{Re}[\alpha_i]} \mathbf{C}^{-1} \frac{\partial \mathbf{h}}{\partial \text{Re}[\alpha_i]} \right] \\ &= \frac{2}{\sigma} \text{Re} \left[ \mathbf{W}^H (\mathbf{I}_{N_M} \otimes \mathbf{A}_t^H \otimes \mathbf{O}_{i,i} \mathbf{A}_r^H) \right. \\ &\quad \cdot (\mathbf{I}_{N_M} \otimes \mathbf{A}_t \otimes \mathbf{A}_r \mathbf{O}_{j,j}) \mathbf{W} \left. \right] \\ &= \frac{2}{\sigma} \text{Re} \left[ \mathbf{W}^H (\mathbf{I}_{N_M} \otimes \mathbf{A}_t^H \mathbf{A}_t \otimes \mathbf{O}_{i,i} \mathbf{A}_r^H \mathbf{A}_r \mathbf{O}_{j,j}) \mathbf{W} \right]. \end{aligned} \quad (47)$$

The dimensions of the kernel matrix in (47) is  $N_M L^2 \times N_M L^2$ , which is difficult to evaluate numerically even for systems with moderate numbers of samples and multipath components.

Noting that  $\mathbf{W}$  is only a stacking of  $N_M$  diagonal matrices, (47) can easily be formulated as a sum

$$\begin{aligned} \frac{2}{\sigma} \text{Re} \left[ \mathbf{W}^H (\mathbf{I}_{N_M} \otimes \mathbf{A}_t^H \mathbf{A}_t \otimes \mathbf{O}_{i,i} \mathbf{A}_r^H \mathbf{A}_r \mathbf{O}_{j,j}) \mathbf{W} \right] \\ = \frac{2}{\sigma} \text{Re} \left[ \sum_{n=1}^{N_M} \mathbf{w}_n^H (\mathbf{A}_t^H \mathbf{A}_t \otimes \mathbf{O}_{i,i} \mathbf{A}_r^H \mathbf{A}_r \mathbf{O}_{j,j}) \mathbf{w}_n \right] \end{aligned} \quad (48)$$

where  $\mathbf{w}_n = \text{vec}[\mathbf{W}(t)]$ . Hence,  $\mathbf{w}_n$  is a vectorized version of a diagonal matrix, and each block of  $L$  elements  $\check{\mathbf{w}}_k^H = \mathbf{w}_n((k-1)L+1 : kL)$  will only contain a single nonzero element. Furthermore, the product  $\mathbf{O}_{i,i} \mathbf{A}_r^H \mathbf{A}_r \mathbf{O}_{j,j}$  only contains a single nonzero element (the  $i, j$  element)  $[\mathbf{A}_t^H \mathbf{A}_t]_{i,j}$ . Based on these observations, the product in (48) can be reformulated as (49), shown at the bottom of the page, where  $\check{\mathbf{w}}^H = [\check{\mathbf{w}}_1^H \cdots \check{\mathbf{w}}_L^H]$ . Observing that  $\check{\mathbf{w}}_k$  only contains a single nonzero element at position  $k$

$$\mathbf{O}_{j,j} \check{\mathbf{w}}_k = \begin{cases} 0 & k \neq j \\ \check{\mathbf{w}}_k(k) & k = j \end{cases} \quad (50)$$

simplifies the product expression to

$$\begin{aligned} \mathbf{w}_n^H (\mathbf{A}_t^H \mathbf{A}_t \otimes \mathbf{O}_{i,i} \mathbf{A}_r^H \mathbf{A}_r \mathbf{O}_{j,j}) \mathbf{w}_n \\ = \check{\mathbf{w}}_i(i) [\mathbf{A}_t^H \mathbf{A}_t]_{i,j} [\mathbf{A}_r^H \mathbf{A}_r]_{i,j} \check{\mathbf{w}}_j(j) \\ = e^{-j\omega_i t} [\mathbf{A}_t^H \mathbf{A}_t]_{i,j} [\mathbf{A}_r^H \mathbf{A}_r]_{i,j} e^{j\omega_j t}. \end{aligned} \quad (51)$$

Thus

$$\begin{aligned} \frac{2}{\sigma} \text{Re} \left[ \frac{\partial \mathbf{h}^H}{\partial \text{Re}[\alpha_i]} \frac{\partial \mathbf{h}}{\partial \text{Re}[\alpha_j]} \right] \\ = \frac{2}{\sigma} \text{Re} \left[ \sum_{n=1}^{N_M} \mathbf{w}_n^H (\mathbf{A}_t^H \mathbf{A}_t \otimes \mathbf{O}_{i,i} \mathbf{A}_r^H \mathbf{A}_r \mathbf{O}_{j,j}) \mathbf{w}_n \right] \\ = \frac{2}{\sigma} \text{Re} \left[ \sum_{n=1}^{N_M} e^{-j\omega_i t} [\mathbf{A}_t^H \mathbf{A}_t]_{i,j} [\mathbf{A}_r^H \mathbf{A}_r]_{i,j} e^{j\omega_j t} \right]. \end{aligned} \quad (52)$$

Once the individual derivative terms are found for each  $i, j$  combination, it is straightforward to obtain a matrix formulation of the entire block of derivatives as

$$\begin{aligned} \frac{2}{\sigma} \text{Re} \left[ \frac{\partial \mathbf{h}^H}{\partial \text{Re}[\boldsymbol{\alpha}]} \frac{\partial \mathbf{h}}{\partial \text{Re}[\boldsymbol{\alpha}]} \right] \\ = \frac{2}{\sigma} \text{Re} \left[ \sum_{n=1}^{N_M} \mathbf{W}^H(t) (\mathbf{A}_t^H \mathbf{A}_t) \odot (\mathbf{A}_r^H \mathbf{A}_r) \mathbf{W}(t) \right] \\ = \text{Re}[\mathbf{P}_1]. \end{aligned} \quad (53)$$

$$\check{\mathbf{w}}^H \begin{bmatrix} [\mathbf{A}_t^H \mathbf{A}_t]_{1,1} \mathbf{O}_{i,i} \mathbf{A}_r^H \mathbf{A}_r \mathbf{O}_{j,j} & [\mathbf{A}_t^H \mathbf{A}_t]_{1,2} \mathbf{O}_{i,i} \mathbf{A}_r^H \mathbf{A}_r \mathbf{O}_{j,j} & \cdots & [\mathbf{A}_t^H \mathbf{A}_t]_{1,L} \mathbf{O}_{i,i} \mathbf{A}_r^H \mathbf{A}_r \mathbf{O}_{j,j} \\ [\mathbf{A}_t^H \mathbf{A}_t]_{2,1} \mathbf{O}_{i,i} \mathbf{A}_r^H \mathbf{A}_r \mathbf{O}_{j,j} & \cdots & & \\ \vdots & \vdots & \vdots & \vdots \\ [\mathbf{A}_t^H \mathbf{A}_t]_{L,1} \mathbf{O}_{i,i} \mathbf{A}_r^H \mathbf{A}_r \mathbf{O}_{j,j} & [\mathbf{A}_t^H \mathbf{A}_t]_{L,2} \mathbf{O}_{i,i} \mathbf{A}_r^H \mathbf{A}_r \mathbf{O}_{j,j} & \cdots & [\mathbf{A}_t^H \mathbf{A}_t]_{L,L} \mathbf{O}_{i,i} \mathbf{A}_r^H \mathbf{A}_r \mathbf{O}_{j,j} \end{bmatrix} \check{\mathbf{w}}. \quad (49)$$

This can also be expressed without the sum by stacking all  $\mathbf{W}(t)$  as

$$\tilde{\mathbf{W}} = [\mathbf{W}(t_1)^H, \dots, \mathbf{W}(t_{N_M})^H]^H.$$

With this definition, we have

$$\frac{2}{\sigma} \text{Re} \left[ \frac{\partial \mathbf{h}^H}{\partial \text{Re}[\boldsymbol{\alpha}]} \frac{\partial \mathbf{h}}{\partial \text{Re}[\boldsymbol{\alpha}]} \right] = \frac{2}{\sigma} \text{Re} [\tilde{\mathbf{W}}^H (\mathbf{I}_{N_M} \otimes \mathbf{A}_t^H \mathbf{A}_t \odot \mathbf{A}_r^H \mathbf{A}_r) \tilde{\mathbf{W}}]. \quad (54)$$

The remaining blocks of the Fisher matrix are found in a similar manner and the final expression for the Fisher matrix is given by (20)–(30) in Section III.

## REFERENCES

- [1] J. B. Andersen, J. Jensen, S. H. Jensen, and F. Fredriksen, "Prediction of future fading based on past measurements," in *Proc. IEEE 50th Vehicular Technology Conf. Fall*, Amsterdam, The Netherlands, Sep. 1999, pp. 151–155.
- [2] A. Arredondo, K. R. Dandekar, and G. Xu, "Vector channel modeling and prediction for the improvement of downlink received power," *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1121–1129, Jul. 2002.
- [3] W. J. Bangs, "Array processing with generalized beamformers," Ph.D. dissertation, Electrical Engineering, Dept., Yale Univ., New Haven, CT, 1971.
- [4] L. Dong, G. Xu, and H. Ling, "Prediction of fast fading mobile radio channels in wide-band communication systems," in *Proc. IEEE GLOBECOM'01*, vol. 6, San Antonio, TX, Nov. 2001, pp. 3287–3291.
- [5] A. Duel-Hallen, S. Hu, and H. Hallen, "Long-range prediction of fading signals," *IEEE Signal Process. Mag.*, vol. 17, no. 3, pp. 62–75, May 2000.
- [6] T. Ekman, "Prediction of mobile radio channels: modeling and design," Ph.D. dissertation, Signals and Systems Dept., Uppsala Univ., Uppsala, Sweden, 2002.
- [7] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, Mar. 1998.
- [8] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [9] J.-K. Hwang and J. H. Winters, "Sinusoidal modeling and prediction of fast fading processes," in *Proc. IEEE GLOBECOM'98*, Sidney, OH, Nov. 1998, pp. 892–897.
- [10] W. Jakes, *Microwave Mobile Communications*. New York: Wiley-Interscience, 1974.
- [11] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [12] C. Komninakis, C. Fragouli, A. H. Sayed, and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1065–1076, May 2002.
- [13] Z. Liu, X. Ma, and G. B. Giannakis, "Space-time coding and Kalman filtering for time-selective fading channels," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 183–186, Feb. 2002.
- [14] A. F. Molisch, M. Steinbauer, M. Toeltsch, E. Bonek, and R. S. Thomä, "Capacity of MIMO systems based on measured wireless channels," *IEEE J. Select. Areas Commun.*, vol. 20, no. 3, pp. 561–569, Apr. 2002.
- [15] B. Narendran, J. Sienicki, S. Yajnik, and P. Agrawal, "Evaluation of an adaptive power and error control algorithm for wireless systems," in *Proc. IEEE Int. Conf. Communications (ICC'97)*, vol. 1, 1997, pp. 349–355.
- [16] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [17] A. Richter, D. Hampicke, G. Sommerkorn, and R. S. Thomä, "MIMO measurement and joint M-D parameter estimation of mobile radio channels," in *Proc. IEEE Vehicular Technology Conf. (VTC) 2001 Spring*, vol. 1, May 2001, pp. 214–218.
- [18] T. Svantesson and J. Wallace, "Statistical characterization of the indoor MIMO channel based on LOS/NLOS measurements," presented at the 36th Asilomar Conf. Signals, Systems, Computers, Monterey, CA, Nov. 2002.
- [19] —, "Performance evaluation of MIMO channel prediction algorithms using measurements (Invited Paper)," in *Proc. IFAC Symp. on Systems Identification (SYSID) 2003*, Rotterdam, The Netherlands, Aug. 2003.
- [20] P. D. Teal and R. G. Vaughan, "Simulation and performance bounds for real-time prediction of the mobile multipath channel," in *Proc. IEEE Workshop on Statistical Signal Processing (SSP)*, Singapore, Aug. 2001, pp. 548–551.
- [21] E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Laboratories, Tech. Rep. BL0112170-950615-07TM, 1995.
- [22] J. W. Wallace, M. A. Jensen, A. L. Swindlehurst, and B. D. Jeffs, "Experimental characterization of the MIMO wireless channel: Data acquisition and analysis," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 335–343, Mar. 2003.



**Thomas Svantesson** (S'98–M'01) was born in Trollhättan, Sweden, in 1972. He received the M.S. degree in electrical engineering and the Ph.D. degree in signal processing from Chalmers University of Technology, Gvteborg, Sweden, in 1996 and 2001, respectively.

From 2001 to 2004, he conducted research on adaptive antennas in wireless communications, channel modeling, and probing of systems employing transmit and receive antenna arrays at Brigham Young University (BYU) and University of California, San Diego (UCSD). Currently, he is with ArrayComm, San Jose, CA, developing adaptive antenna algorithms for cellular networks employing transmit and receive antenna arrays.



**A. Lee Swindlehurst** (S'83–M'84–SM'99–F'04) received the B.S. degree (*summa cum laude*) and the M.S. degree, both in electrical engineering, from Brigham Young University, Provo, UT, in 1985 and 1986, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1991.

From 1983 to 1984, he was a Scientific Programmer with Eyring Research Institute, Provo, UT. From 1984 to 1986, he was a Research Assistant in the Department of Electrical Engineering at Brigham Young University, working on various problems in signal processing and estimation theory. He was awarded an Office of Naval Research Graduate Fellowship for 1985–1988, and during most of that time, he was affiliated with the Information Systems Laboratory at Stanford University. From 1986 to 1990, he was also employed at ESL, Inc., Sunnyvale, CA, where he was involved in the design of algorithms and architectures for several radar and sonar signal processing systems. He joined the faculty of the Department of Electrical and Computer Engineering at Brigham Young University in 1990, where he currently holds the position of Full Professor and is the department chair. From 1996 to 1997, he held a joint appointment as a Visiting Scholar at both Uppsala University, Uppsala, Sweden, and at the Royal Institute of Technology, Stockholm, Sweden. His research interests include sensor array signal processing for radar and wireless communications, detection and estimation theory, and system identification, and he has over 130 publications in these areas.

Dr. Swindlehurst is a past Secretary of the IEEE Signal Processing Society. He is currently serving as a member of the Sensor Array and Multichannel Signal Processing Technical Committee in the same society and as a member of the Editorial Board for the *EURASIP Journal on Wireless Communications and Networking*. He is a past Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was the Technical Program Chair for the 1998 IEEE Digital Signal Processing Workshop and for the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. He also received the 2000 IEEE W. R. G. Baker Prize Paper Award (for a paper coauthored with P. Stoica titled "Maximum Likelihood Methods in Radar Array Signal Processing," appearing in the IEEE PROCEEDINGS, Feb. 1998) and the IEEE Signal Processing Society Young Author Best Paper Award in 2001 (for a paper he coauthored with D. Astely and B. Ottersten titled "Spatial Signature Estimation for Uniform Linear Arrays with Unknown Gains and Phases," appearing in IEEE TRANSACTIONS ON SIGNAL PROCESSING, August 1999).