MIMO Channel Estimation in Correlated Fading Environments

Yen-Chih Chen and Yu T. Su, *Senior Member, IEEE*

***Abstract*—This paper presents two analytic correlated multiple-input multiple-output (MIMO) block fading channel models and their time-variant extensions that encompass the popular Kronecker model and the more general Weichselberger model as special cases. Both static and time-variant models offer compact representations of spatial- and/or time-correlated channels. When the transmit antenna array is such that the associated MIMO channel has a small angle spread (AS), which occurs quite often in a cellular downlink, our models admit reduced-rank channel representations. They also provide compact channel state information (CSI) descriptions which are needed in feedback systems and in many post channel estimation applications. The latter has the important implication of reduced feedback channel bandwidth requirement and lower post-processing complexity.**

Based on one of the proposed channel models we present novel iterative algorithms for estimating static and time-variant MIMO channels. The proposed models make it natural to decompose each iteration of our algorithms into two successive stages that are responsible for estimating the correlation coefficients and the signal direction, respectively. Using popular industry-approved standard channel models, we verify through simulations that our algorithms yield good MSE performance which, in many practical cases, is better than that achievable by a conventional least-square estimator. The mean-squared error (MSE) performance of our estimators are analyzed and the resulting predictions are consistent with those estimated by simulations.

***Index Terms*—Channel estimation, space-time signal processing, spatial correlation.**

I. Introduction

INCREASING demand for higher wireless system capacity has catalyzed several ground-breaking transmission techniques, among which is the multiple-input/multiple-output (MIMO) technology that has attracted the great part of recent attention. It has been shown that in comparison with conventional single antenna systems, significant capacity gains are achievable when multi-element antennas (MEA) are used at both the transmit and receive sides [1]. Spatial multiplexing techniques, for example, the BLAST (Bell-labs Layered Space-Time) system, were developed to attain very high spectral efficiencies in rich scattering environments.

Ideal rich-scattering environments decorrelate channels between different pairs of transmit and receive antennas so that maximum number of orthogonal subchannels is available. In practice, however, spatial correlations do exist and should be considered when designing a MIMO receiver for evaluating the corresponding system performance [2]. Spatial correlation depends on physical parameters such as antenna spacing, antenna arrangement, and scatters' distributions. Antenna correlations reduce the number of equivalent orthogonal subchannels, decrease spectral efficiency, making it more difficult to detect the transmitted data [1].

Among the many analytic models for spatial-correlated MIMO channels that have been proposed, the Kronecker model [2] is perhaps the most popular one. It assumes separable statistics at transmitter and receiver so that the spatial correlation matrix of the vectorized channel matrix is given by the Kronecker product of those of the transmit and receive antenna arrays. The separable assumption of the Kronecker model, however, has been shown to be inappropriate by some recent experiments; see [3] and the references therein. Hence it has been modified and generalized by Sayeed [4] and, more recently, by Weichselberger *et al.* [3] who considered joint correlation of both link ends.

The analytic models are often used to evaluate the system capacity/performance [5],[6], to design beamformer [7] or training (pilot) sequences [8], [9] and for link level simulations [10]-[12]. These models decompose the channel matrix into terms that respectively characterize various components of the spatial structure thus contains many more parameters than those used in the original channel matrix. As a coherent MIMO receiver requires an accurate channel estimate to perform critical operations and provide satisfactory performance, these models are not suitable for channel estimation application and subsequent post-channel-estimation signal processing.

We propose two general analytic correlated MIMO block fading models and their time-variant extensions that encompass the above-mentioned models as special cases. One of our models results in separable descriptions of channel correlations and mean angle of departure (AoD). Spatial and time covariance (or correlation) functions are described by predetermined nonparametric regressions. Unlike the other analytic correlated MIMO models, the proposed models do not incur additional parameters. On the contrary, when the angle spread (AS) of the antenna array is relatively small, they admit reduced-rank representations and compact channel state information (CSI) representation which have significant implications on complexity and bandwidth reduction in many post-channel estimation operations such as data detection and feedback beamforming.

Manuscript received December 4, 2008; revised May 22, 2009 and October 15, 2009; accepted November 25, 2009. The associate editor coordinating the review of this paper and approving it for publication was M. Morelli.

This work was supported in part by Taiwan's National Science Council under contact NSC93-2218-E-009-016, and by Computer and Communication Laboratories, Industrial Technology Research Institute, Chu-Tung, Taiwan. The material was presented in part at the PIMRC2004, Barcelona, Spain, September 2004.

Y.-C. Chen is with Realtek Inc., Hsinchu, Taiwan (e-mail: joechen@realtek.com.tw).

Y. T. Su is with the Department of Electrical Engineering, National Chiao Tung University, Hsinchu, Taiwan (e-mail: ytsu@mail.nctu.edu.tw).

Digital Object Identifier 10.1109/TWC.2010.03.081603

1536-1276/10$25.00 © 2010 IEEE

Various pilot-assisted MIMO channel estimators have been proposed [8]-[14]. Unfortunately, few estimators are specifically designed for correlated MIMO channels and those few exploited only channel's time and frequency correlation characteristics by approximating the time- and/or frequency-domain responses by an analytic model [13], [14]. Because of the special structures of the proposed models, conventional MIMO channel estimation approaches are not applicable. We present novel pilot-assisted channel estimation schemes based on one of the proposed new general MIMO channel representations which does not require information of second-order channel statistics. This representation enables us to develop efficient algorithms to identify the realistic channel responses. Although a model-based scheme inevitably induces a modelling error [13]-[16], as will be shown in Section VI, our algorithms are capable of describing realistic correlated MIMO channels with negligible modelling errors.

After a brief review of the typical space-time antenna setup and a general received MIMO signal model, we derive two new models for spatial-correlated block-faded narrowband MIMO channels and their relations with some established analytic models in Section II. We then propose single-block based iterative least squares (LS) channel estimators in the following section while the extension that takes the time-correlation into account is given in Section IV. In Section V, we analyze the mean squared error (MSE) of the proposed channel estimation algorithms. Numerical examples using industrial standard approved channel models are given in Section VI to validate the proposed channel models and to demonstrate the effectiveness of our algorithms. Concluding remarks are given in the last section.

II. Modelling of Spatial-Correlated MIMO Channels

A. System Setup

Consider a cellular MIMO system in which the base station (BS) and a mobile station (MS) are equipped with linear arrays of M and N antennas, respectively. Independent data streams $\mathbf{x}(t) = [x_1(t), x_2(t), x_3(t), \cdots, x_M(t)]^T$ are transmitted from the BS at time t, where $x_m(t)$ denotes the source signal of the mth transmit antenna and the superscript T denotes vector (matrix) transposition. At the MS, the received baseband signals are given by $\mathbf{y}(t) = [y_1(t), y_2(t), y_3(t), \cdots, y_N(t)]^T$, where $y_n(t)$ is the signal received by the nth receive antenna at time t. With a sampling interval of $\triangle t$ seconds, the corresponding ith transmit and receive sample vectors are $\mathbf{x}_i = \mathbf{x}(i\triangle t)$, and $\mathbf{y}_i = \mathbf{y}(i\triangle t)$, respectively.

B. Wireless MIMO Channels

A general MIMO channel between BS and MS antennas is modelled as

$$\mathbf{H}(t,\tau) = \sum_{l=1}^{G} \mathbf{H}_l(t)\delta(\tau - \tau_l), \tag{1}$$

where G is the maximum number of paths associated with any sub-channel between a transmit and receive antenna pair, τ_l is the delay of the lth path, and δ denotes the Dirac delta function. The complex channel gain matrix associated with the lth path is given by $\mathbf{H}_l = [h^l_{ij}]$, for $1 \le i \le N, 1 \le j \le M$, where h^l_{ij} is the complex sub-channel gain between the jth transmit and ith receive antennas. For a narrowband fading channel, (1) is reduced to a single-tap fading matrix and the received vector waveform is $\mathbf{y}(t) = \mathbf{H}(t)\mathbf{x}(t) + \mathbf{n}(t)$, where $\mathbf{H}(t)$ is an $N \times M$ complex channel matrix and $\mathbf{n}(t)$ a zero mean additive white Gaussian noise (AWGN) vector with covariance matrix $E\{\mathbf{n}\mathbf{n}^H\} = N_0\mathbf{I}_N$. We first consider the block fading case in which the channel gain matrix remains unchanged within a block of B symbol intervals and eliminate the time parameter t in related expressions. Section IV will discuss the case which takes the time-correlation among blocks into consideration.

C. Spatial-correlated block fading channels

Let $\mathbf{\Phi}$, $\mathbf{\Phi}_{\mathrm{T}}$ and $\mathbf{\Phi}_{\mathrm{R}}$ be the spatial correlation matrices of vec($\mathbf{H}$), vec($\cdot$) being the stacking operator, the transmit and receive antennas, respectively. The separable statistics assumption of the Kronecker model implies $\mathbf{\Phi} = \mathbf{\Phi}_{\mathrm{R}} \otimes \mathbf{\Phi}_{\mathrm{T}} = \mathbf{\Phi}^{\frac{1}{2}}(\mathbf{\Phi}^{\frac{1}{2}})^H$, where the "square root" matrix $\mathbf{\Phi}^{\frac{1}{2}}$ has a similar decomposition $\mathbf{\Phi}^{\frac{1}{2}} = \mathbf{\Phi}_{\mathrm{T}}^{\frac{1}{2}} \otimes \mathbf{\Phi}_{\mathrm{R}}^{\frac{1}{2}}$ and therefore gives

$$\mathbf{H} = \mathbf{\Phi}_{\mathrm{R}}^{\frac{1}{2}}\mathbf{H}_w\mathbf{\Phi}_{\mathrm{T}}^{\frac{1}{2}\ T}, \tag{2}$$

where $\mathbf{H}_w$ is an $N \times M$ channel matrix whose entries are i.i.d. complex zero-mean, unit-variance Gaussian random variables.

Weichselberger *et al.* [3] considered joint correlation of both link ends and suggested the following analytic model

$$\mathbf{H} = \mathbf{U}_{\mathrm{R}}\left(\tilde{\mathbf{\Omega}} \odot \mathbf{H}_w\right)\mathbf{U}_{\mathrm{T}}^T, \tag{3}$$

where $\mathbf{U}_{\mathrm{T}}$ and $\mathbf{U}_{\mathrm{R}}$ are the eigenbases of the one-sided correlation matrices at the transmit and receive sites, respectively. Operator $\odot$ denotes the Hadamard product operation [17]. $\tilde{\mathbf{\Omega}}$ is the element-wise square root of the coupling matrix in which each entry specifies the mean amount of energy coupled with an eigenvector of the transmitter to that of the receiver.

An $N \times M$ matrix $\mathbf{H}$ always admit the singular value decomposition (SVD), $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$, where $\mathbf{U}$ is an $N \times N$ unitary matrix, $\mathbf{V}$ is an $M \times M$ unitary matrix, and the diagonal matrix $\mathbf{\Lambda}$ is $N \times M$ with non-negative entries. When $\mathbf{H}$ is random, its SVD component matrices are random and depend on the sample (matrix) value of $\mathbf{H}$. As $\mathbf{U}$ and $\mathbf{V}$ can be transformed into two predefined unitary matrices $\mathbf{Q}_{\mathrm{R}}$ and $\mathbf{Q}_{\mathrm{T}}$ by $\mathbf{U}\mathbf{P}_1 = \mathbf{Q}_{\mathrm{R}}$ and $\mathbf{V}\mathbf{P}_2 = \mathbf{Q}_{\mathrm{T}}$, with both transforms $\mathbf{P}_1$ and $\mathbf{P}_2$ being unitary, we have

$$\mathbf{H} = \mathbf{Q}_{\mathrm{R}}\mathbf{P}_1^{-1}\mathbf{\Lambda}(\mathbf{P}_2^{-1})^T\mathbf{Q}_{\mathrm{T}}^T = \mathbf{Q}_{\mathrm{R}}\mathbf{C}\mathbf{Q}_{\mathrm{T}}^T \tag{4}$$

and the only random term on the right hand side is $\mathbf{C}$. For the Weichselberger model, the predefined matrices are eigenbases of the one-sided correlation matrices while Sayeed's virtual channel representation uses the DFT bases.

Let $\mathbf{\Phi}_{\mathrm{T}}^{\frac{1}{2}} \stackrel{def}{=} [\phi_{\mathrm{T}}(i,j)]$, where $\phi_{\mathrm{T}}(i,j)$ represents the root spatial correlation between ith and jth transmit antennas. As the M column vectors of $\mathbf{\Phi}_{\mathrm{T}}^{\frac{1}{2}}$ lie in a $K_T(\le M)$ dimensional subspace, we have

$$\mathbf{\Phi}_{\mathrm{T}}^{1/2} = \mathbf{Q}_{\mathrm{T}}\mathbf{\Lambda}_{\mathrm{T}}, \tag{5}$$

where $\mathbf{Q}_{\mathrm{T}}$ is an unitary matrix and the coefficient matrix $\mathbf{\Lambda}_{\mathrm{T}}$ can be obtained by the Gram-Schimdt orthonormalization procedure. The above equation implies $\phi_{\mathrm{T}}(i,j)=\sum_{k=1}^{K_T}\lambda_k^j\mathbf{q}_k(i)$, where $\mathbf{q}_k(i)$ is the ith element of the kth basis vector, λ_k^j is the projection of the jth column on $\mathbf{q}_k$.

Using a similar decomposition for $\mathbf{\Phi}_{\mathrm{R}}^{1/2}$ leads to $\mathbf{\Phi}^{1/2}=(\mathbf{Q}_{\mathrm{T}}\mathbf{\Lambda}_{\mathrm{T}})\otimes(\mathbf{Q}_{\mathrm{R}}\mathbf{\Lambda}_{\mathrm{R}})=(\mathbf{Q}_{\mathrm{T}}\otimes\mathbf{Q}_{\mathrm{R}})(\mathbf{\Lambda}_{\mathrm{T}}\otimes\mathbf{\Lambda}_{\mathrm{R}})$, where we have invoked the identity [17],

$$\begin{aligned}(\mathbf{A}_1\otimes\mathbf{B}_1)(\mathbf{A}_2\otimes\mathbf{B}_2)\cdots(\mathbf{A}_k\otimes\mathbf{B}_k)&\\ =(\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_k)\otimes(\mathbf{B}_1\mathbf{B}_2\cdots\mathbf{B}_k).&\end{aligned}\tag{6}$$

¿From the canonical representation, $\mathrm{vec}(\mathbf{H})=\mathbf{\Phi}^{\frac{1}{2}}\mathrm{vec}(\mathbf{H}_w)$, we obtain

$$\begin{aligned}\mathrm{vec}(\mathbf{H}) &= (\mathbf{Q}_{\mathrm{T}}\otimes\mathbf{Q}_{\mathrm{R}})(\mathbf{\Lambda}_{\mathrm{T}}\otimes\mathbf{\Lambda}_{\mathrm{R}})\,\mathrm{vec}(\mathbf{H}_w)\\ &\stackrel{def}{=} (\mathbf{Q}_{\mathrm{T}}\otimes\mathbf{Q}_{\mathrm{R}})\,\mathrm{vec}(\mathbf{C}).\end{aligned}\tag{7}$$

The identity

$$\mathrm{vec}\,(\mathbf{ABD})=\left(\mathbf{D}^T\otimes\mathbf{A}\right)\mathrm{vec}\,(\mathbf{B})\tag{8}$$

implies $\mathrm{vec}(\mathbf{H})=\mathrm{vec}\left(\mathbf{Q}_{\mathrm{R}}\mathbf{C}\mathbf{Q}_{\mathrm{T}}^T\right)$, and so $\mathbf{H}=\mathbf{Q}_{\mathrm{R}}\mathbf{C}\mathbf{Q}_{\mathrm{T}}^T$, which is the same as (4).

We summarize the above derivation on the relation between the proposed analytic model with the Kronecker, Sayeed, and Weichselberger models in

Proposition 1: An $N\times M$ flat-faded MIMO channel matrix $\mathbf{H}$, can always be expressed as

$$\mathbf{H}=\mathbf{Q}_{\mathrm{R}}\mathbf{C}\mathbf{Q}_{\mathrm{T}}^T\tag{9}$$

where $\mathbf{C}$ is a complex random coefficient matrix, $\mathbf{Q}_{\mathrm{R}}$ and $\mathbf{Q}_{\mathrm{T}}$ are predefined unitary matrices. The above model is equivalent to the Kronecker model if the matrix $\mathbf{C}$ satisfies the separable correlation condition

$$\mathrm{vec}(\mathbf{C})=(\mathbf{\Lambda}_{\mathrm{T}}\otimes\mathbf{\Lambda}_{\mathrm{R}})\,\mathrm{vec}(\mathbf{H}_w)\tag{10}$$

where $\mathbf{\Lambda}_{\mathrm{T}}$ and $\mathbf{\Lambda}_{\mathrm{R}}$ are coefficient matrices that depend on the spatial correlations among the transmit and the receive antenna arrays, respectively. (9) is related to the Weichselberger model via

$$\mathbf{U}_{\mathrm{T}}=\mathbf{Q}_{\mathrm{T}}\mathbf{P}_{\mathrm{T}}^H,\quad \mathbf{U}_{\mathrm{R}}=\mathbf{Q}_{\mathrm{R}}\mathbf{P}_{\mathrm{R}}^H\tag{11}$$

$$\mathbf{P}_{\mathrm{R}}^H\Gamma_{\mathrm{R}}\mathbf{P}_{\mathrm{R}}=E\left\{\mathbf{C}\mathbf{C}^H\right\},\quad \mathbf{P}_{\mathrm{T}}^H\Gamma_{\mathrm{T}}\mathbf{P}_{\mathrm{T}}=E\left\{\mathbf{C}^T\mathbf{C}^*\right\}\tag{12}$$

where $\mathbf{P}_{\mathrm{T}}$, $\mathbf{P}_{\mathrm{R}}$ are unitary matrices and $\Gamma_{\mathrm{R}},\Gamma_{\mathrm{T}}$ have the same eigenvalues of the matrices $E\left\{\mathbf{H}\mathbf{H}^H\right\}$ and $E\left\{\mathbf{H}^T\mathbf{H}^*\right\}$, respectively. When the predefined matrices are $\mathbf{U}_{\mathrm{R}}$ and $\mathbf{U}_{\mathrm{T}}$, $\mathbf{C}$ has the special form $\tilde{\mathbf{\Omega}}\odot\mathbf{H}_w$. Moreover, (9) is equivalent to the virtual representation of Sayeed if columns of $\mathbf{Q}_{\mathrm{R}}$ and $\mathbf{Q}_{\mathrm{T}}$ are DFT basis vectors and entries of $\mathbf{C}$ are independent complex Gaussian random variables.

[10] suggested and [11] verified through field measurements that the mean direction of arrival (DoA) can be embedded in the channel model by post-multiplying the channel matrix $\mathbf{H}$ by a diagonal matrix which is a function of the DoA. We can derive a similar model by invoking the fact that if $\mathbf{W}$ is a diagonal matrix with unit modulus entries and $\mathbf{V}$ is unitary then both $\mathbf{VW}$ and $\mathbf{W}^{-1}\mathbf{V}$ are also unitary, to obtain the alternative representation (13).

Corollary 1: An equivalent channel matrix for stationary frequency-flat MIMO channel is given by

$$\mathbf{H}=\mathbf{Q}_{\mathrm{R}}\mathbf{C}\overline{\mathbf{Q}}_{\mathrm{T}}^T\mathbf{W}\tag{13}$$

where $\overline{\mathbf{Q}}_{\mathrm{T}}^T\mathbf{W}=\mathbf{Q}_{\mathrm{T}}^T$ and $\mathbf{W}$ is a diagonal matrix with unit modulus entries.

Several remarks and observations on the channel models (9) and (13) are given below.

R1. The Kronecker model requires that $\mathbf{C}$ has the special structure (10) while the Weichselberger and Sayeed models demand that the entries of $\mathbf{C}$ be independent (but not identical) Gaussian random variables. In contrast, the proposed model does not impose any constraint on the coefficient matrix $\mathbf{C}$ and is valid for arbitrary block-faded $\mathbf{H}$.

R2. For practical correlated MIMO channels, which are of particular concern to this paper, the entries of $\mathbf{H}$ are not i.i.d. but correlated random variables. Although $\mathbf{H}$ or the corresponding one-sided correlation matrices for a correlated channel is still likely to be of full rank, it often has a large eigen-spread and thus admits a reduced-rank representation by ignoring the weaker eigenmodes. The rank-reduction is most obvious for typical urban macro-cellular environments in which an MS is surrounded by local scatterers, and waveforms impending the receive antennas are richly scattered, while the BS is not obstructed by the local scatterers [2][12]. Appendix A shows that, if the AS Δ is not too large, the diagonal matrix $\mathbf{W}$

$$\mathbf{W}=\mathrm{diag}\left[w_1,w_2,\cdots,w_M\right],\tag{14}$$

has entries of the form $w_i=\exp\left[-j2\pi\frac{(i-1)d}{\lambda}\sin\phi\right]$, d being the inter-element distance, that bear the mean AoD information. As will become clear later, the separability of the channel correlation and angle information characterizations has some useful implications.

R3. Given predefined bases $\mathbf{Q}_{\mathrm{R}}$, $\mathbf{Q}_{\mathrm{T}}$, or $\overline{\mathbf{Q}}_{\mathrm{T}}$, the statistical properties of the corresponding coefficient matrix is completely determined by those of $\mathbf{C}$. Identification of the unknown channel $\mathbf{H}$ is equivalent to the estimation of $\mathbf{C}$ or the pair $(\mathbf{C},\mathbf{W})$, which usually has a lower (dominant) rank and thus much smaller number of entries than those of $\mathbf{H}$ for the link environment of interest. Thus, using model (9) or (13) reduces the number of parameters to be estimated and enhances the performance. Moreover, as the bases in both (9) and (13) are pre-defined, these two models can be easily extended to time-varying block fading and frequency-selective fading environments.

R4. There are several classes of basis functions to choose from. The Taylor and Weierstrass arguments and the results of [21] suggest the use of polynomial bases. If we use polynomials of degree P as basis functions in expanding a spatial correlation function of length P, the corresponding basis matrix $\mathbf{P}_P$ has entries

$$[\mathbf{P}]_{i,j}=(i-1)^{j-1},\ i,j=1,2,\ldots,P,\tag{15}$$

Although the column vectors in (15) form a basis, they are not orthogonal. Furthermore, these vectors have different norms, which might result in numerical instability. By applying the QR decomposition to the corresponding $\mathbf{P}_P$ [22], we obtain an orthonormalized polynomial basis matrix $\mathbf{P}_o$, i.e. $\mathbf{P}_P = \mathbf{P}_o\mathbf{R}_o$, where $\mathbf{R}_o$ is the corresponding upper triangular matrix. Since the columns of $\mathbf{P}_0$ are arranged in ascending order of polynomial degree, (so that its kth column represents an eigenmode describing higher correlation than that described by the lth column if $l > k$), we select the first K_T, K_R or K_L columns to form the basis matrices $\mathbf{Q}_{M,K_T}$ and $\mathbf{Q}_{N,K_R}$ of (17) or $\mathbf{Q}_{L,K_L}$ of (37).

R5. For a fixed base one needs to determine the modelling orders, K_T and K_R. Either the Akaike information criterion (AIC) or the minimum description length (MDL) approach can be used to determine the optimal modelling orders that trade-off the system complexity and performance [23]. Time domain modelling order K_L discussed in Section IV can also be similarly determined. Depending on the application scenario, these order parameter values can be obtained by an one-shot open loop estimate or should be periodically updated.

R6. R2 and R3 indicate that, for a MIMO system with a small-to-medium AS, the model (13) is more useful for post-channel-estimation application, hence our subsequent discourse will focus on this model only. Numerical results reported in Section VI also confirm that this representation leads to significant bandwidth and complexity reduction for systems which have to feedback information about or process a large MIMO channel matrix.

R7. Our simulation experiments indicate that, when the AS $\triangle$ becomes large, the dominant rank of $\mathbf{C}$ increases accordingly and there is no dominant spatial angle. The steering matrix $\mathbf{W}$ becomes an identity matrix which gives no AoD information and (13) degenerates to (9).

III. Single-Block Based Channel Estimation

In this section we consider estimation schemes which are based on a single block of observation without taking into account the (time-)correlation among blocks. We propose two iterative schemes in which an iteration consists of two phases. The first phase is responsible for the estimation of the coefficient matrix, $\mathbf{C}$, while the directional matrix, $\mathbf{W}$ in (13), is estimated in the second phase. Both tentative estimates are updated as one proceeds with each new iteration until the stopping criterion is met. The two schemes differ in the second phase only.

Consider the $M \times B$ matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_B]$ formed by B length-M input symbol vectors, where $B \geq M$. Assuming $\mathbf{H}$ remains static during a B-block period, we express the received sample block, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_B]$ as

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N}, \tag{16}$$

where $\mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2, \cdots, \mathbf{n}_B]$ is the corresponding noise matrix whose entries are i.i.d. zero mean complex Gaussian random variables. In estimating $\mathbf{H}$, $\mathbf{X}$ is assumed to be composed of either the pilot vectors or some decision feedback results. Substituting two known unitary matrices $\mathbf{Q}_{M,K_T}$ and $\mathbf{Q}_{N,K_R}$ with ranks $K_T(\leq M)$ and $K_R(\leq N)$ for $\mathbf{Q}_T$ and $\overline{\mathbf{Q}}_R$ in (13), we want to find the optimal solution $\{\mathbf{W}_{opt}, \mathbf{C}_{opt}\}$ to the problem

$$\arg\min_{\mathbf{W},\mathbf{C}} \|\mathbf{Y} - \mathbf{Q}_{N,K_R}\mathbf{C}\mathbf{Q}_{M,K_T}^T\mathbf{W}\mathbf{X}\|^2 \tag{17}$$

We express the corresponding optimal (least-squares) channel estimate in terms of $\mathbf{W}_{opt}$ and $\mathbf{C}_{opt}$

$$\mathbf{H}_{opt} = \mathbf{Q}_{N,K_R}\mathbf{C}_{opt}\mathbf{Q}_{M,K_T}^T\mathbf{W}_{opt} \tag{18}$$

so that (16) can be rewritten as

$$\mathbf{Y} = \mathbf{H}_{opt}\mathbf{X} + \Delta\mathbf{H}\mathbf{X} + \mathbf{N} \stackrel{def}{=} \mathbf{H}_{opt}\mathbf{X} + \widetilde{\mathbf{N}}, \tag{19}$$

where $\widetilde{\mathbf{N}}$ represents the sum of the modelling error $\Delta\mathbf{H}\mathbf{X}$ due to the reduced rank representation and the AWGN vector, $\mathbf{N}$.

To derive an iterative algorithm for obtaining the joint directional and channel solution $\{\mathbf{W}_{opt}, \mathbf{H}_{opt}\}$, we first notice that, at the $(i-1)$th iteration,

$$\mathbf{Y} = \widehat{\mathbf{H}}_{i-1}\mathbf{X} + \Delta\widehat{\mathbf{H}}_{i-1}\mathbf{X} + \widetilde{\mathbf{N}} \tag{20}$$

where $\Delta\widehat{\mathbf{H}}_{i-1} \stackrel{def}{=} \mathbf{H}_{opt} - \widehat{\mathbf{H}}_{i-1}$ is the residual error at the end of the $(i-1)$th iteration, and consider the estimation of the channel (coefficients) and AoD in two separate phases.

A. Phase I - Coefficient Estimation

Assume that the directional matrix in this phase is optimum, i.e., $\mathbf{W} = \mathbf{W}_{opt}$. From (16) and (18), we have

$$\mathrm{vec}(\mathbf{Y}) = \left\{\left((\mathbf{W}_{opt}\mathbf{X})^T\mathbf{Q}_{M,K_T}\right) \otimes \mathbf{Q}_{N,K_R}\right\}\mathrm{vec}(\mathbf{C}) + \mathrm{vec}(\mathbf{N}). \tag{21}$$

Substituting the definition $\mathbf{Z} \stackrel{def}{=} ((\mathbf{W}_{opt}\mathbf{X})^T\mathbf{Q}_{M,K_T}) \otimes \mathbf{Q}_{N,K_R}$ into (21), we have the LS solution

$$\mathrm{vec}(\widehat{\mathbf{C}}) = (\mathbf{Z}^H\mathbf{Z})^{-1}\mathbf{Z}^H\mathrm{vec}(\mathbf{Y}) \stackrel{def}{=} F(\mathbf{W}_{opt}). \tag{22}$$

While the optimal directional matrix estimate is not available, we replace it by the tentative estimation from the previous iteration, $\widehat{\mathbf{W}}_{i-1}$. $\mathrm{vec}(\widehat{\mathbf{C}})$ is then obtained by computing $F(\widehat{\mathbf{W}}_{i-1})$ instead, and the corresponding tentative estimate is denoted by $\widehat{\mathbf{C}}_i$. Initially, we can arbitrarily set $\widehat{\mathbf{W}}_0$ to be an identity matrix.

B. Phase II - Direction Estimation

Similar to Phase I, we begin with the assumption that the coefficient matrix in this estimation phase is optimum. The directional information is to be obtained by estimating a diagonal matrix $\mathbf{W}$ with unit modulus entries; see (14). Setting

$$\mathbf{G} \stackrel{def}{=} \mathbf{Q}_{N,K_R}\mathbf{C}_{opt}\mathbf{Q}_{M,K_T}^T \tag{23}$$

and invoking (18), we have $\widehat{\mathbf{H}}_{i-1} = \mathbf{G}\widehat{\mathbf{W}}_{i-1}$. As $\mathbf{C}_{opt}$ is unavailable, $\mathbf{C}_{opt}$ is replaced by the previous estimate $\widehat{\mathbf{C}}_{i-1}$ in computing $\mathbf{G}$ during the ith iteration. In the following, we propose two algorithms to estimate the phase of the unit modulus diagonal entries of $\mathbf{W}$.

1) Algorithm A - Maximum Matching Output: To estimate $\widehat{\mathbf{W}}_i$ in diagonal form, we start with the following lemma whose proof is given in Appendix B.

Lemma 1: For two matrices $\mathbf{A}$ *and* $\mathbf{B}$ *of size* $N \times M$ *and* $M \times E$ *respectively, and an arbitrary vector* $\mathbf{c}$ *of size* $M \times 1$*, the following identity holds.*

$$vec\left(\mathbf{A} \cdot diag(\mathbf{c}) \cdot \mathbf{B}\right) = \left[(\mathbf{1}_E \otimes \mathbf{A}) \odot \left(\mathbf{B}^T \otimes \mathbf{1}_N\right)\right] \mathbf{c}, \quad (24)$$

where "diag" denotes the diagonal operation used to translate a vector into a diagonal matrix, with its diagonal terms being the elements of the original vector.

Combined with matrix $\mathbf{G}$ defined in (23), (19) is rewritten as

$$\mathbf{Y} = \mathbf{G}\mathbf{W}_{opt}\mathbf{X} + \widetilde{\mathbf{N}}. \quad (25)$$

Let $\mathbf{w}_{opt}$ be the column vector that consists of the diagonal elements of $\mathbf{W}_{opt}$, i.e., $\mathbf{w}_{opt}(i) = \mathbf{W}_{opt}(i,i)$, for any $1 \leq i \leq M$. Then, by Lemma 1, we have

$$\text{vec}\left(\mathbf{Y}\right) = \left[(\mathbf{1}_B \otimes \mathbf{G}) \odot \left(\mathbf{X}^T \otimes \mathbf{1}_N\right)\right] \mathbf{w}_{opt} + \text{vec}\left(\widetilde{\mathbf{N}}\right) \quad (26)$$

and the LS estimate of $\mathbf{w}_{opt}$ is given by $\widehat{\mathbf{w}}_{LS} = \mathbf{T}^{\dagger} \cdot \text{vec}(\mathbf{Y})$, where $\left[(\mathbf{1}_B \otimes \mathbf{G}) \odot \left(\mathbf{X}^T \otimes \mathbf{1}_N\right)\right] \stackrel{def}{=} \mathbf{T}$.

In order to extract the steering vector $\widehat{\mathbf{w}}$, we introduce $\mathbf{v}(\theta) \stackrel{def}{=} \left[1, v(\theta), \ldots, v^{M-1}(\theta)\right]^T$, where $v(\theta) = \exp\left[-j2\pi\frac{d}{\lambda}\sin(\theta)\right]$. The AoD information $\widehat{\phi}$ is retrieved by maximizing the matching output

$$\widehat{\phi} = \arg \max_{-\pi \leq \theta \leq \pi} \text{Re}\left\{\mathbb{P}\left(\widehat{\mathbf{w}}_{LS}\right)^H \mathbf{v}(\theta)\right\}, \quad (27)$$

where $\mathbb{P}(\cdot)$ is defined by the following phase extraction operator,

$$\begin{aligned} &\mathbb{P}\left([a_0 e^{jb_0}, a_1 e^{jb_1}, \cdots, a_K e^{jb_K}]\right) \\ &\quad \stackrel{def}{=} [1, e^{j(b_1-b_0)}, \cdots, e^{j(b_K-b_0)}], \\ &\qquad\qquad \{a_i\} \in \mathcal{R}_+^{K+1}, \{b_i\} \in [0,\ 2\pi) \end{aligned} \quad (28)$$

Once $\widehat{\phi}$ is available, it is straightforward to obtain $\widehat{\mathbf{W}} = \text{diag}(\mathbf{v}(\widehat{\phi}))$. Solving (27) over $[0, 2\pi)$ can be accomplished by using the conventional line searching algorithm.

Computing $\widehat{\mathbf{w}}_{LS}$ in (27) involves a pseudo-inverse operation of matrix $\mathbf{T}$, and is thus computational expensive. However, the matrix inversion is not needed if an orthogonal training sequence set, which is optimal for LS channel estimator [8], is used. This can be seen by noting that

$$\mathbf{T}^H\mathbf{T} \quad = \quad (\mathbf{G}^H\mathbf{G}) \odot (\mathbf{X}^*\mathbf{X}^T), \quad (29)$$

and the right-hand side of (29) is a diagonal matrix with nonnegative real elements if $\mathbf{X}^*\mathbf{X}^T = B\mathbf{I}$. In this case, we have

$$\begin{aligned} \mathbb{P}\left(\widehat{\mathbf{w}}_{LS}\right) &= \mathbb{P}\left(\mathbf{T}^{\dagger} \cdot \text{vec}(\mathbf{Y})\right) = \mathbb{P}\left(\mathbf{T}^H \cdot \text{vec}(\mathbf{Y})\right) \\ &\stackrel{def}{=} \mathbb{P}\left(\widetilde{\widehat{\mathbf{w}}}_{LS}\right), \end{aligned} \quad (30)$$

The AoD information can thus be obtained simply by substituting $\widetilde{\widehat{\mathbf{w}}}_{LS}$, which is obtained without matrix inversion, for $\widehat{\mathbf{w}}_{LS}$ in (27).

2) Algorithm B - Root Finding Method: An alternative way to find the optimal phase is to convert (27) into a root finding problem. Note that the elements of $\mathbf{w}_{opt}$ are of geometric progression, i.e., they form a row vector of a Vandermonde matrix. Hence if we define the correlation polynomial

$$P(z) \stackrel{def}{=} \mathbb{P}(\widehat{\mathbf{w}}_{LS})^H \mathbf{z} - M, \quad (31)$$

where $\mathbf{z} = [1, z, \ldots, z^{M-1}]$ and let $\mathcal{Z}$ be the set of its zeros in the complex plane, then solving (27) is equivalent to

$$\widehat{z} = \arg \min_{z \in \mathcal{Z}} |(|z| - 1)| \quad \text{and} \quad \widehat{\phi} = \sin^{-1}\left(\frac{-\text{Arg}\{\widehat{z}\}\lambda}{2\pi d}\right) \quad (32)$$

and the directional matrix is reconstructed by $\widehat{\mathbf{W}} = \text{diag}(\widehat{\mathbf{z}})$, where $\widehat{\mathbf{z}} = [1, \widehat{z}, \cdots, \widehat{z}^{M-1}]$. Similar to *Algorithm A*, if the orthogonal training matrix is used, $\widetilde{\widehat{\mathbf{w}}}_{LS}$ of (30) can be substituted for $\widehat{\mathbf{w}}_{LS}$ in (31) to simplify the computation. While the accuracy of *Algorithm A* relies on the resolution the numerical search algorithm used, this algorithm gives the exact analytic solution once (31) is solved.

C. Convergence and Complexity

Since the object function in (17) is jointly convex with respect to $\mathbf{C}$ and $\mathbf{W}$ and the proposed algorithms have the form of a nonlinear Gauss-Seidel algorithm, the convergences of our algorithms are guaranteed [24]. All the simulation examples reported in Section VI converge and achieve the theoretical performance lower bound derived in Section V.

The computation complexity of the proposed algorithm is dominated by the LS operations in *Phase I* and *Phase II*. The flop counts of the LS operation in *Phase I* is $O(BK_T^2)$, $K_T \leq M$ while the conventional LS estimator needs $O(BM^2)$ flops [25]. *Phase II*'s complexity is of the same order as that of the conventional LS estimator, thanks to the special structure of $\mathbf{T}$. By using an orthogonal training matrix, the pseudo-inversion in (22) for *Phase I* can be simplified to a matrix-vector product, which needs no matrix inversion operation as the modelling matrix is unitary and $\widehat{\mathbf{W}}$ has unit modulus terms. The pseudo-inverse operation of *Phase II* can also be replaced by the product of $\mathbf{T}^H$ and $\text{vec}(\mathbf{Y})$, and the complexity is reduced as well. Moreover, except for static channels, the estimates for both $\mathbf{W}$ and $\mathbf{C}$ need to be updated periodically. Let each $B-$symbol interval be called an estimation interval (EI). Since the mean AoD usually change much slower than the channel coefficients (gains) variation, updating frequencies for $\mathbf{W}$ and $\mathbf{C}$ can and should be different, i.e., if the two estimates are updated every T_o^c and T_o^w EIs, respectively, then $T_o^w \gg T_o^c$ (see Fig. 8 of Section VI). This dual updating frequency option is unique to our approach and implies that *Phase II* may be disabled most of the time while *Phase I* needs single iteration per update EI, hence our algorithm may be computational more efficient than the conventional LS approach for many non-static channels.

The major advantage of our channel model and estimator lies not in the computational efficiency of the channel estimator but in the compactness of CSI representation which is needed in a feedback system and that of post-channel-estimation operations. As mentioned in R2 and R3, a small K_T is often sufficient to accurately describe a MIMO channel with high transmit spatial correlation. For any post channel estimation operation associated with $\mathbf{H}$, e.g., taking pseudo-inverse or eigen-decomposition of $\mathbf{H}$, the computing load is reduced as it involves the $K_R \times K_T$ coefficient matrix and the estimated AoD instead of the original $N \times M$ channel matrix.

IV. Channel Estimation with Time Correlation Consideration

We now extend our investigation to the case that the time correlation among blocks has to be taken into account. Like our spatial modelling approach, we use a set of orthonormal basis functions to describe a snap shot of a fading channel's time domain behavior. We assume an equally spaced pilot-block arrangement. The issue of the optimal pilot arrangement that minimizes the MSE or bit error rate (BER) was addressed in [9] and [26].

Assuming the two leading pilot symbol vectors of two consecutive pilot blocks are T symbol intervals away, we express the receive signal block at time nT as

$$\mathbf{Y}_n = \mathbf{H}_n \mathbf{X}_n + \mathbf{N}_n \tag{33}$$

where $\mathbf{Y}_n = \mathbf{Y}(nT)$ and $\mathbf{X}_n = \mathbf{X}(nT)$ are the $N \times B$ receive matrix at time nT and the corresponding $M \times B$ transmit block, respectively. $\mathbf{H}_n$ is the $N \times M$ matrix whose (i,j)th entry represents the link gain between the ith transmit and the jth receive antennas at time nT.

We consider the time-variant behavior of a MIMO channel within a fixed observation window of L blocks (EIs). The received sample blocks from nT to $(n+L-1)T$ can be cascaded into the matrix

$$\mathbf{Y}_{n,L} \stackrel{def}{=} [\mathbf{Y}_n, \mathbf{Y}_{n+1}, \ldots, \mathbf{Y}_{n+L-1}]. \tag{34}$$

Using (8), we obtain

$$\text{vec}(\mathbf{Y}_{n,L}) = \left(\mathbf{X}_{n,L}^T \otimes \mathbf{I}_N\right) \cdot \text{vec}\left(\mathbf{H}_{n,L}\right) + \text{vec}\left(\mathbf{N}_{n,L}\right) \tag{35}$$

where $\text{vec}(\mathbf{H}_{n,L}) \stackrel{def}{=} \left[\text{vec}(\mathbf{H}_n)^T, \ldots \text{vec}(\mathbf{H}_{n+L-1})^T\right]^T$, $\text{vec}(\mathbf{N}_{n,L}) \stackrel{def}{=} \left[\text{vec}(\mathbf{N}_n)^T, \ldots \text{vec}(\mathbf{N}_{n+L-1})^T\right]^T$, and

$$\mathbf{X}_{n,L}^T \stackrel{def}{=} \begin{bmatrix} \mathbf{X}_n^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{n+1}^T & \vdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_{n+L-1}^T \end{bmatrix}.$$

Substituting (9) for each $\mathbf{H}_n$ and assuming the eigenbases $\mathbf{Q}_\text{T}$ and $\mathbf{Q}_\text{R}$ remain invariant during an estimation period, we obtain

$$\text{vec}(\mathbf{H}_{n,L}) = (\mathbf{I}_L \otimes \mathbf{Q}_\text{T} \otimes \mathbf{Q}_\text{R})\, \mathbf{\Gamma}_{n,L}. \tag{36}$$

Each component of the vector $\mathbf{\Gamma}_{n,L} = \left[\gamma_n^T, \gamma_{n+1}^T, \cdots, \gamma_{n+L-1}^T\right]^T$ is itself an $(NM) \times 1$ column vector $\gamma_n = \left[\gamma_{1n}, \gamma_{2n}, \cdots, \gamma_{(NM)n}\right]^T$ that represents the complex fading coefficients for all NM MIMO subchannels at time nT and, γ_{pn}, $1 \le p \le NM$, are independent.

The stacked vector, $\gamma(p) = \left[\gamma_{pn}, \gamma_{p(n+1)}, \cdots, \gamma_{p(n+L-1)}\right]^T$, represents a finite-duration sample of the complex random process associated with the pth subchannel [11]. Such a process can also be expanded by a set of smooth functions [14], [15], and thus its estimation can be obtained by using a method similar to that developed in the previous section. Hence, we can first apply the orthogonal transform $\gamma(p) = \mathbf{Q}_L \mathbf{b}_{pn}$, where $\mathbf{Q}_L$ is an $L \times L$ orthogonal matrix, and $\mathbf{b}_{pn}$ is the transform domain coefficient vector. Then, the time domain channel correlation can be approximated by using the reduced basis matrix $\mathbf{Q}_{L,K_L}$

$$\gamma(p) \approx \mathbf{Q}_{L,K_L} \mathbf{c}_{pn}, \;\; \mathbf{\Gamma}_{n,L} \approx \left(\mathbf{Q}_{L,K_L} \otimes \mathbf{I}_{LMN}\right) \mathbf{c}_\text{coef}, \tag{37}$$

where K_L denotes the time domain modelling order, and $\mathbf{c}_{pn}$ is a $K_L \times 1$ coefficient vector.

By using (13), (36) and the approximation (37), we decouple the signal part of (35) into the product of two modelling domains - space and time domains

$$\begin{aligned} \text{vec}(\bar{\mathbf{Y}}_{n,L}) &\approx \left(\mathbf{X}_{n,L}^T \otimes \mathbf{I}_N\right) \left[\mathbf{Q}_{L,K_L} \otimes (\mathbf{W}^T \overline{\mathbf{Q}}_\text{T}) \otimes \mathbf{Q}_\text{R}\right] \mathbf{c}_\text{coef} \\ &\approx \left(\mathbf{X}_{n,L}^T \otimes \mathbf{I}_N\right) \left[\mathbf{Q}_{L,K_L} \otimes (\mathbf{W}^T \mathbf{Q}_{\text{T},K_T}) \otimes \mathbf{Q}_{\text{R},K_R}\right] \tilde{\mathbf{c}}_\text{coef} \\ &\stackrel{def}{=} \left(((\mathbf{W}_L \mathbf{X}_{n,L})^T \widetilde{\mathbf{Q}}_{\text{T},K_T}) \otimes \mathbf{Q}_{\text{R},K_R}\right) \tilde{\mathbf{c}}_\text{coef} \end{aligned} \tag{38}$$

where $\mathbf{W}_L \stackrel{def}{=} (\mathbf{I}_L \otimes \mathbf{W})$, $\widetilde{\mathbf{Q}}_{\text{T},K_T} \stackrel{def}{=} \mathbf{Q}_{L,K_L} \otimes \mathbf{Q}_{\text{T},K_T}$ and $\mathbf{Q}_{\text{T},K_T}$ and $\mathbf{Q}_{\text{R},K_R}$ are composed of K_T and K_R column vectors of $\overline{\mathbf{Q}}_\text{T}$ and $\mathbf{Q}_\text{R}$, respectively. $\mathbf{W}$ is the steering matrix defined in (14). Since the mean AoD usually varies slowly with respect to a sub-channel's coherent time, we assume that $\mathbf{W}$ remains the same during a period of L data blocks. Just as the narrowband case (13), we do not impose the implicit Kronecker structure and Gaussian assumption on $\tilde{\mathbf{c}}_\text{coef}$.

As (38) can be obtained by replacing $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{W}$, vec$(\mathbf{C})$, $\mathbf{Q}_{M,K_T}$, and $\mathbf{Q}_{N,K_R}$ in (21) by $\mathbf{X}_{n,L}$, $\mathbf{Y}_{n,L}$, $\mathbf{W}_L$, $\tilde{\mathbf{c}}_\text{coef}$, $\widetilde{\mathbf{Q}}_{\text{T},K_T}$, and $\mathbf{Q}_{\text{R},K_R}$, we conclude that both spatial and time correlations can be described by similar models. Hence, the two-phase iterative estimation scheme developed in Section III can be extended to estimate the coefficient vector $\tilde{\mathbf{c}}_\text{coef}$, and the directional matrix $\mathbf{W}_L$ in (38). In the following, we describe two-phase channel estimation schemes with time correlation consideration.

A. Phase I - Coefficient Estimation

Following an argument similar to that used in Section III, we assume that the directional matrix $\mathbf{W}_L$ is optimal in the coefficient estimation phase and define

$$\widetilde{\mathbf{Z}} \stackrel{def}{=} \left((\mathbf{W}_{L,opt} \mathbf{X}_{n,L})^T \widetilde{\mathbf{Q}}_{\text{T},K_T}\right) \otimes \mathbf{Q}_{\text{R},K_R}. \tag{39}$$

The LS estimate of $\tilde{\mathbf{c}}_\text{coef}$ is

$$\widehat{\tilde{\mathbf{c}}}_\text{coef} = (\widetilde{\mathbf{Z}}^H \widetilde{\mathbf{Z}})^{-1} \widetilde{\mathbf{Z}}^H \text{vec}\left(\mathbf{Y}_{n,L}\right) \stackrel{def}{=} \widetilde{F}(\mathbf{W}_{L,opt}), \tag{40}$$

which is a function of the optimal directional matrix $\mathbf{W}_{L,opt}$. At the ith iteration, since the optimal directional matrix is not available, the tentative estimation, $\widehat{\mathbf{W}}_{L,i-1}$, replaces $\mathbf{W}_{L,opt}$.

B. Phase II - Direction Estimation

Similar to the single-block based case, we propose two AoD estimation algorithms. Again, we assume the optimal coefficient vector is available, i.e., $\tilde{\mathbf{c}}_\text{coef} = \tilde{\mathbf{c}}_{\text{coef},opt}$, when estimating the directional information.

Define a new matrix $\widetilde{\mathbf{G}} \overset{def}{=} \mathbf{Q}_{\mathrm{R},K_R}\widetilde{\mathbf{C}}_{\mathrm{coef},opt}\widetilde{\mathbf{Q}}^T_{\mathrm{T},K_T}$, where $\widetilde{\mathbf{C}}_{\mathrm{coef},opt}$ is a $K_R \times K_L K_T$ matrix derived from $\tilde{\mathbf{c}}_{\mathrm{coef},opt}$ by $\widetilde{\mathbf{C}}_{\mathrm{coef},opt}(i,j) = \tilde{\mathbf{c}}_{\mathrm{coef},opt}\,(K_R(j-1)+i)$, $1 \le i \le K_R, 1 \le j \le K_L K_T$. We rewrite the received matrix in vector form

$$\begin{aligned}\mathrm{vec}(\mathbf{Y}_{n,L}) &= \mathrm{vec}\left(\widetilde{\mathbf{G}}\mathbf{W}_L\mathbf{X}_{n,L}\right) + \widetilde{\mathbf{N}}_{n,L} \\ &= \left(\mathbf{X}^T_{n,L} \otimes \widetilde{\mathbf{G}}\right)\mathrm{vec}(\mathbf{I}_L \otimes \mathbf{W}) + \widetilde{\mathbf{N}}_{n,L},\end{aligned} \quad (41)$$

where $\widetilde{\mathbf{N}}_{n,L}$ represents the sum of the modelling error associated with $\widetilde{\mathbf{G}}$ and the AWGN term $\mathbf{N}_{n,L}$.

1) Algorithm A - Maximum Matching Output: If $\mathbf{W}$ is constrained to be a diagonal matrix, i.e., $\mathbf{W} = \mathrm{diag}(\mathbf{w})$, then $\mathbf{I}_L \otimes \mathbf{W} = \mathrm{diag}(\mathbf{1}_L \otimes \mathbf{w})$ and therefore

$$\mathrm{vec}(\mathbf{Y}_{n,L}) = \mathrm{vec}\left(\widetilde{\mathbf{G}} \cdot \mathrm{diag}(\mathbf{1}_L \otimes \mathbf{w}) \cdot \mathbf{X}_{n,L}\right) + \widetilde{\mathbf{N}}_{n,L}. \quad (42)$$

¿From Lemma 1, we have

$$\begin{aligned}&\mathrm{vec}\left(\widetilde{\mathbf{G}} \cdot \mathrm{diag}(\mathbf{1}_L \otimes \mathbf{w})) \cdot \mathbf{X}_{n,L}\right) \\ &= \left(\left(\mathbf{1}_{LE} \otimes \widetilde{\mathbf{G}}\right) \odot \left(\mathbf{X}^T_{n,L} \otimes \mathbf{1}_N\right)\right)(\mathbf{1}_L \otimes \mathbf{I}_M)\mathbf{w} \\ &\overset{def}{=} \widetilde{\mathbf{T}}\mathbf{w}.\end{aligned} \quad (43)$$

The LS estimate of $\mathbf{w}_{opt}$, like its counterpart in *Algorithm A* of the previous subsection, is given by $\widehat{\mathbf{w}}_{LS} = \widetilde{\mathbf{T}}^\dagger \cdot \mathrm{vec}(\mathbf{Y}_{n,L})$. To improve the estimate and reconstruct a steering vector $\widehat{\mathbf{w}}$, we analogously define a steering vector $\mathbf{v}(\theta) \overset{def}{=} \left[1, v(\theta), \cdots, v^{M-1}(\theta)\right]^T$, where $v(\theta) = \exp(-j2\pi\frac{d}{\lambda}\sin(\theta))$. The AoD information $\widehat{\phi}$ can be retrieved by

$$\widehat{\phi} = \arg\max_{-\pi \le \theta \le \pi} \mathrm{Re}\left\{\mathbb{P}(\widehat{\mathbf{w}}_{LS})^H\mathbf{v}(\theta)\right\}, \quad (44)$$

where $\mathbb{P}$ denotes the phase extraction operator defined by (28). Having obtained $\widehat{\phi}$, we then proceed to compute $\widehat{\mathbf{W}}_L = \mathbf{I}_L \otimes \mathbf{V}(\widehat{\phi})$, where $\mathbf{V}(\widehat{\phi}) = \mathrm{diag}(\mathbf{v}(\widehat{\phi}))$.

The pseudo-inverse operation $\widetilde{\mathbf{T}}^\dagger$ is not necessary if the orthogonal training matrix is used for $\mathbf{X}_n$, i.e., $\mathbf{X}_n\mathbf{X}^H_n = B\mathbf{I}$ for each n. We then have

$$\mathbb{P}(\ \widetilde{\mathbf{T}}^\dagger \cdot \mathrm{vec}\left(\mathbf{Y}_{n,L}\right)\ ) = \mathbb{P}(\ \overbrace{\widetilde{\mathbf{T}}^H \cdot \mathrm{vec}\left(\mathbf{Y}_{n,L}\right)}^{\overset{def}{=}\widetilde{\widehat{\mathbf{w}}}_{LS}}\ ), \quad (45)$$

and $\widehat{\mathbf{w}}_{LS}$ is to be replaced by $\widetilde{\widehat{\mathbf{w}}}_{LS}$ defined in the above equation.

2) Algorithm B - Root-Finding Method: The root-finding approach for the block fading case can be used as well. It is easy to see that (44) is equivalent to searching for the root of the correlation polynomial $P(z)$ which is the closest to the unit circle, i.e.,

$$\widehat{z} = \arg\min_z\ ||z| - 1|, \qquad \text{subject to } P(z) \overset{def}{=} \mathbb{P}(\widehat{\mathbf{w}}_{LS})^H\mathbf{z} - M = 0 \quad (46)$$

and then retrieving the AoD information from $\widehat{z} = \exp\left[-j2\pi\frac{d}{\lambda}\sin(\widehat{\phi})\right]$. The directional matrix is to be reconstructed by $\widehat{\mathbf{W}}_L = \mathbf{I}_L \otimes \mathrm{diag}(\widehat{\mathbf{z}})$, where $\widehat{\mathbf{z}} = [1, \widehat{z}, \ldots, \widehat{z}^{M-1}]$. Similarly, using orthogonal training matrices, we replace $\widehat{\mathbf{w}}_{LS}$ by $\widetilde{\widehat{\mathbf{w}}}_{LS}$ to avoid pseudo-matrix inversion.

The total complexity per block of the proposed algorithm, like the single-block based case, is still larger than the conventional LS estimator when the dual updating frequencies option is not used. However, if the operating scenario allows the use of the latter option, the complexity can be asymptotically reduced to $\frac{K_T^2}{M^2}$ of that of the conventional LS method if $T^w_o \ll T^c_o$. Furthermore, by using an orthogonal training matrix, we need no matrix inversion in both *Phase I* and *Phase II* and significantly reduce the computing complexity. For slowing time-variant channels, the required time domain modelling order, K_L, is small, the number of channel representation parameters is reduced from LMN to $K_L K_T K_R + 1$. Such a reduction yields compact CSI representation and benefits many post channel estimation operations involving $\mathbf{H}$, as was discussed at the end of last section.

V. Performance Analysis

In analyzing the MSE performance

$$\epsilon \overset{def}{=} E\left\{\|\mathbf{H} - \widehat{\mathbf{H}}\|^2_F\right\} = E\left\{\|\mathrm{vec}(\mathbf{H}) - \mathrm{vec}(\widehat{\mathbf{H}})\|^2_2\right\}. \quad (47)$$

of the proposed $\widehat{\mathbf{H}}$, we first make the optimistic assumptions that the optimal orthogonal training matrix [8] for conventional LS channel estimator is used and the directional matrix estimate $\widehat{\mathbf{W}}$ is perfect.

Notations

For notational simplicity and when there is no danger of ambiguity, $\mathbf{H}$ and $\mathbf{W}$ in this section denote the channel and directional matrices of (16)/(17) or (35)/ (38) for single-block based or time-correlated based estimators, and $\mathbf{X}_p$ represents $\mathbf{X}$ in (21) or $\mathbf{X}_{n,L}$ in (38). Furthermore, $\mathbf{Q}_{\mathrm{T}}$ and $\mathbf{Q}_{\mathrm{R}}$ denote either the modelling bases $\mathbf{Q}_{M,K_T}$ and $\mathbf{Q}_{N,K_R}$ in (21), or $\widetilde{\mathbf{Q}}_{\mathrm{T},K_T}$ and $\mathbf{Q}_{\mathrm{R},K_R}$ in (38).

Then (47) can be expressed as

$$\begin{aligned}\epsilon(\mathbf{X}_p;\mathbf{W}) &= E\left\{\|\mathrm{vec}(\mathbf{H}) - \mathrm{vec}(\mathbf{Q}_{\mathrm{R}}\widehat{\mathbf{C}}\mathbf{Q}^T_{\mathrm{T}}\mathbf{W})\|^2_2\right\} \\ &= E\left\{\|\mathrm{vec}(\mathbf{H}) - \mathbf{\Psi}\mathbf{\Omega}_z\mathrm{vec}(\mathbf{H}\mathbf{X}_p + \mathbf{N})\|^2_2\right\}\end{aligned} \quad (48)$$

where $\mathbf{\Psi} \overset{def}{=} (\mathbf{W}^T\mathbf{Q}_{\mathrm{T}}) \otimes \mathbf{Q}_{\mathrm{R}}$ and $\mathbf{\Omega}_z \overset{def}{=} (\widehat{\mathbf{Z}}^H\widehat{\mathbf{Z}})^{-1}\widehat{\mathbf{Z}}^H$, $\widehat{\mathbf{Z}}$ being the LS estimate of $\mathbf{Z}$ defined in III-A, i.e.,

$$\widehat{\mathbf{Z}} \overset{def}{=} ((\mathbf{W}\mathbf{X}_p)^T\mathbf{Q}_{\mathrm{T}}) \otimes \mathbf{Q}_{\mathrm{R}} \quad (49)$$

As $\mathbf{H}\mathbf{X}_p$ and $\mathbf{N}$ are statistically independent, the MSE can be separated into two terms which are contributed by modelling error (reduced-rank basis matrices) and AWGN, respectively.

$$\begin{aligned}\epsilon(\mathbf{X}_p;\mathbf{W}) &= E\left\{\|\mathrm{vec}(\mathbf{H}) - \mathbf{\Psi}\mathbf{\Omega}_z\mathrm{vec}(\mathbf{H}\mathbf{X}_p)\|^2_2\right\} \\ &\quad + E\left\{\|\mathbf{\Psi}\mathbf{\Omega}_z\mathrm{vec}(\mathbf{N})\|^2_2\right\} \\ &\overset{def}{=} \epsilon_h(\mathbf{X}_p,\mathbf{W}) + \epsilon_n(\mathbf{X}_p,\mathbf{W})\end{aligned} \quad (50)$$

By defining the projections $\mathbf{P}_{\mathbf{W}} \overset{def}{=} \left[\mathbf{W}^T\mathbf{Q}_{\mathrm{T}}(\mathbf{Q}^T_{\mathrm{T}}\mathbf{W}^*\mathbf{X}^*_p\mathbf{X}^T_p\mathbf{W}^T\mathbf{Q}_{\mathrm{T}})^{-1}\mathbf{Q}^T_{\mathrm{T}}\mathbf{W}^*\mathbf{X}^*_p\mathbf{X}^T_p\right] \otimes \mathbf{Q}_{\mathrm{R}}\mathbf{Q}^T_{\mathrm{R}}$ and $\tilde{\mathbf{P}}_{\mathbf{W}} \overset{def}{=} \left[\mathbf{W}^T\mathbf{Q}_{\mathrm{T}}(\mathbf{Q}^T_{\mathrm{T}}\mathbf{Q}_{\mathrm{T}})^{-1}\mathbf{Q}^T_{\mathrm{T}}\mathbf{W}^*\right] \otimes \mathbf{Q}_{\mathrm{R}}\mathbf{Q}^T_{\mathrm{R}}$. we rewrite the first term on the RHS of (50) as

$$\begin{aligned}\epsilon_h(\mathbf{X}_p;\mathbf{W}) &= E\|(\mathbf{I} - \mathbf{P}_{\mathbf{W}})\,\mathrm{vec}(\mathbf{H})\|^2_2 \\ &= \mathrm{tr}\left((\mathbf{I} - \tilde{\mathbf{P}}^H_{\mathbf{W}})(\mathbf{I} - \tilde{\mathbf{P}}_{\mathbf{W}})\mathbf{R}_h\right) \\ &= \sum_{k=1}^{\chi}\lambda_k\|(\mathbf{I} - \tilde{\mathbf{P}}_{\mathbf{W}})\mathbf{f}_k\|^2_2\end{aligned} \quad (51)$$

where $\mathbf{R}_h = E\left\{\text{vec}(\mathbf{H})\text{vec}(\mathbf{H})^H\right\}$ is the channel correlation matrix and $\mathbf{f}_k$ is $\mathbf{R}_h$'s eigenvector associated with the eigenvalue λ_k, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_\chi$; χ being the degree of freedom of $\mathbf{H}$. For the single-block based case, $\chi = NM$ and it is equal to NML when the estimator considers the time correlation effect. (51) is valid since the orthogonal training matrix $\mathbf{X}_p$ is used. Let $1 < K \leq \chi$ be the rank of the dominant signal subspace of the channel covariance matrix. Then $\mathbf{R}_h = \sum_{k=1}^{\chi} \lambda_k \mathbf{f}_k \mathbf{f}_k^H \simeq \sum_{k=1}^{K} \lambda_k \mathbf{f}_k \mathbf{f}_k^H$, with $\lambda_k \ll 1$ for $K < k \leq \chi$. Since $\|(\mathbf{I} - \tilde{\mathbf{P}}_\mathbf{W})\mathbf{f}_k\|_2^2 \leq 1$, we have $\sum_{k=K+1}^{\chi} \lambda_k \|(\mathbf{I} - \tilde{\mathbf{P}}_\mathbf{W})\mathbf{f}_k\|_2^2 \leq \sum_{k=K+1}^{K_s} \lambda_k \ll 1$. Let the compound modelling order K_s be equal to $K_T K_R$ and $K_T K_R K_L$ for the two cases under investigation. If K_s is chosen to be larger than K, the rank of $\mathbf{R}_h$, i.e., $K < K_s \leq \chi$, and the basis matrices $\mathbf{Q}_\text{T}$ and $\mathbf{Q}_\text{R}$ span the dominant signal subspace of $\mathbf{R}_h$, then the matrix $\tilde{\mathbf{P}}_\mathbf{W}$ is a projection operator whose range lies mostly in the space spanned by $\{\mathbf{f}_k\}, 1 \leq k \leq K$ and we conclude that $\|(\mathbf{I} - \tilde{\mathbf{P}}_\mathbf{W})\mathbf{f}_k\|_2^2 \overset{def}{=} |\tilde{\mathbf{P}}_\mathbf{W}^\perp \mathbf{f}_k\|_2^2 \ll 1$, for $1 \leq k \leq K$. Therefore, the modelling error ϵ_h is negligible in this case. On the other hand, if the modelling order is not enough to span the signal subspace, there is under-modelling error contributed by those non-negligible terms $\lambda_k \|(\mathbf{I} - \tilde{\mathbf{P}}_\mathbf{W})\mathbf{f}_k\|_2^2$ which will dominate the mean squared error when the AWGN is small (high SNR region).

As for the MSE due to thermal noise–the second term on the RHS of (50), we can show that

$$\begin{aligned} \epsilon_n(\mathbf{X}_p, \mathbf{W}) &= E\left\{\|\mathbf{\Psi}\mathbf{P}_z \text{vec}(\mathbf{N})\|_2^2\right\} \\ &= \text{tr}\left(\frac{N_0}{B}\tilde{\mathbf{P}}_\mathbf{W}\right) = \frac{N_0}{B}K_s, \end{aligned} \tag{52}$$

where we have invoked the facts that (i) the training signal $\mathbf{X}_p$ and the noise $\mathbf{N}$ are independent, (ii) unitary training matrix is used, and (iii) elements of $\mathbf{N}$ is i.i.d. complex white Gaussian noise with variance $\sigma_n^2 = N_0$. (52) implies that thermal noise induced MSE can be reduced by using a small modelling order. In Section VI (Figs. 3-5), we find that this noise-reduction effect is significant in low SNR environments where thermal noise dominates the MSE performance while the modelling error of (51) dominates in high SNR region.

If $\widehat{\mathbf{W}}$ is not perfect and $\mathbf{W} = \widehat{\mathbf{W}} + \Delta\mathbf{W}$, then

$$\begin{aligned} \widehat{\mathbf{Z}} &\overset{def}{=} \mathbf{Z} + \Delta\mathbf{Z} \\ &= ((\mathbf{W}\mathbf{X}_p)^T \mathbf{Q}_\text{T}) \otimes \mathbf{Q}_\text{R} + ((\Delta\mathbf{W}\mathbf{X}_p)^T \mathbf{Q}_\text{T}) \otimes \mathbf{Q}_\text{R}. \end{aligned} \tag{53}$$

The coefficient vector estimation $\text{vec}(\widehat{\mathbf{C}})$ can be approximated up to the first order of $\Delta\mathbf{Z}$ as [27]

$$\begin{aligned} \text{vec}(\widehat{\mathbf{C}}) &\simeq \text{vec}(\mathbf{C}) - \mathbf{Z}^\dagger \Delta\mathbf{Z}\text{vec}(\mathbf{C}) + \mathbf{Z}^\dagger \text{vec}(\mathbf{N}) \\ &\quad + (\mathbf{Z}^H\mathbf{Z})^{-1}\Delta\mathbf{Z}^H P_\mathbf{Z}^\perp \text{vec}(\mathbf{N}) - \mathbf{Z}^\dagger \Delta\mathbf{Z}\mathbf{Z}^\dagger \text{vec}(\mathbf{N}), \end{aligned} \tag{54}$$

where $P_\mathbf{Z}^\perp = \mathbf{I} - \mathbf{Z}(\mathbf{Z}^H\mathbf{Z})^{-1}\mathbf{Z}^H$. The above equation indicates that, besides the terms that have to do with the noise $\mathbf{N}$, the coefficient vector estimation error is determined by the projection error $\Delta\mathbf{Z}$. Hence, when the projection error $\Delta\mathbf{W}$ is small (and thus $\Delta\mathbf{Z}$ is small), $\text{vec}(\widehat{\mathbf{C}})$ is a good approximation of $\text{vec}(\mathbf{C})$ at high SNR region.

VI. Numerical Results and Discussion

Simulation results reported here use the reference MIMO channel model [28], the IEEE 802.11 TGn channel model [29], and the SCM model [30]. The former two are stochastic models whose spatial correlation matrices are generated by the power azimuth spectrum (PAS) at the BS and MS, respectively. The SCM model generates the channel coefficients according to a set of selected parameters (e.g., AS, AoD, AoA, etc.). It is a popular parametric stochastic model whose spatial cross correlations are functions of the joint distribution of the AoD at the transmit side and the AoA at the receive side. We assume that the environment surrounding MS is rich scattering with negligible spatial correlations. Hence, a full rank basis matrix is used to characterize the spatial correlation at the receive side. Other assumptions and conditions used in our simulation are: (i) the antenna spacings at transmit and receive arrays are both 0.5λ, (ii) an orthogonal training matrix is used, (iii) 10 iterations are used for all simulations (although in most cases convergence occurs in less than 3 iterations), and (iv) SNR (E_b/N_0) is defined as the average signal to noise power ratio at the input of each receive antenna, (v) orthonormal polynomial basis matrices are used. Both algorithms compute $\widehat{\mathbf{H}}$ by substituting the final result of *Phase I*–estimated coefficient matrix $\widehat{\mathbf{C}}$–and that of *Phase II*–$\widehat{\mathbf{W}}$–into (18).

Solid curves in Fig. 1 are the MSE performance of *Algorithm B* in Section III for an 8×8 MIMO system with $\Delta = 2°$ and are based on the channel model of [28]. The channel is a block fading channel with an approximated rank of two. Since the BS spatial correlations are high, the corresponding correlation function lies in a low-dimension subspace so that a small K_T is sufficient to describe the channel. Dotted curves in Fig. 1 show the system performance when $\Delta = 15°$. It is obvious that as Δ increases, the spatial correlations among the transmit antennas elements decrease and a higher modelling order is necessary to describe rapid-changing spatial waveforms at the transmitter side. However, as can be seen from Figs. 2-5, an optimal K_T exists for any given SNR and Δ; increasing the modelling order does not necessary improves the channel estimator's performance. As expected, we find that modelling errors dominate the MSE performance when SNR is high. Such a behavior is consistent with what the performance analysis given in Section V has predicted and is similar to those observed in other model-based approaches [13]-[16].

The MSE performance of *Algorithm B* of Section IV for a time-correlated fading channel [28] are depicted in Fig. 2 and Fig. 3 using an observation window of 12 EIs and $f_d T_s = 0.031772$ or 0.015886. Similar to the single-blocked based case (Fig. 1), the processing dimension (K_T) can be drastically reduced provided that either the spatial or time domain correlation is high enough. Performance degradation occurs when the modelling order is not large enough to capture the channel characteristics. In Fig. 4, we compare the theoretical MSE derived in Section V with the simulated performance and find that the latter is very close to the theoretical bound which assumes a perfect $\widehat{\mathbf{W}}$. When used for estimating other reference channels, the proposed estimators exhibit similar performance behaviors. Fig. 5 depicts the

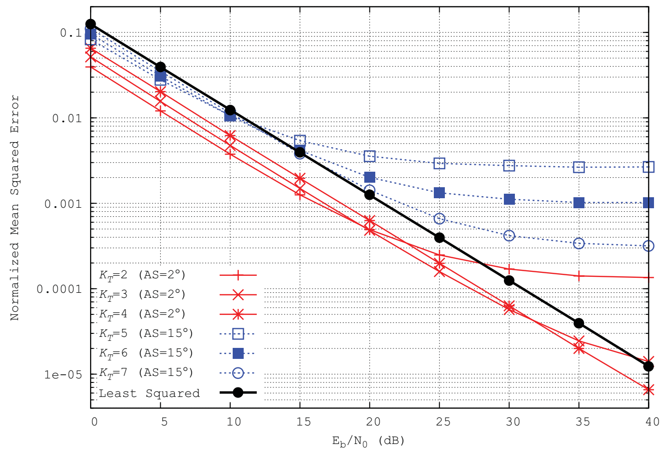

Fig. 1. MSE performance of *Algorithm B* as a function of SNR with different modelling orders; solid curves: AS=2°, dotted curves: AS=15°.

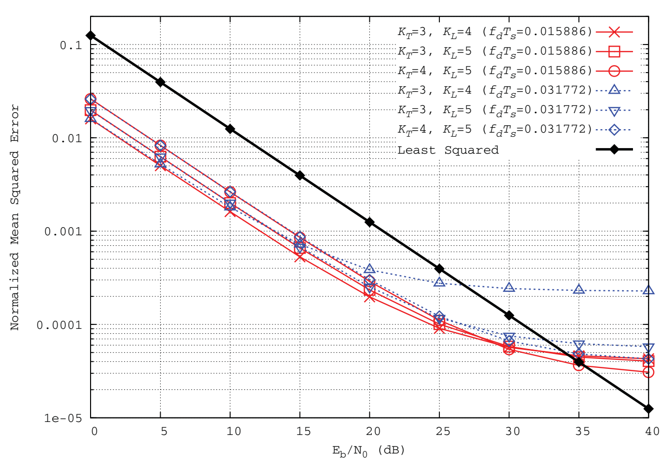

Fig. 2. The effect of the modelling order on *Algorithm B*'s MSE performance in a channel generated by the model described in [10] with AS=2°.

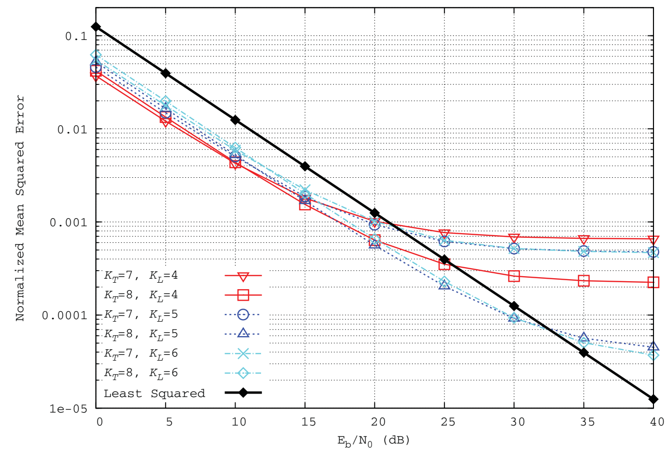

Fig. 3. The effect of modelling order on *Algorithm B*'s MSE performance in a channel generated by the model described in [10] with AS=15° and f_dT_s=0.031772.

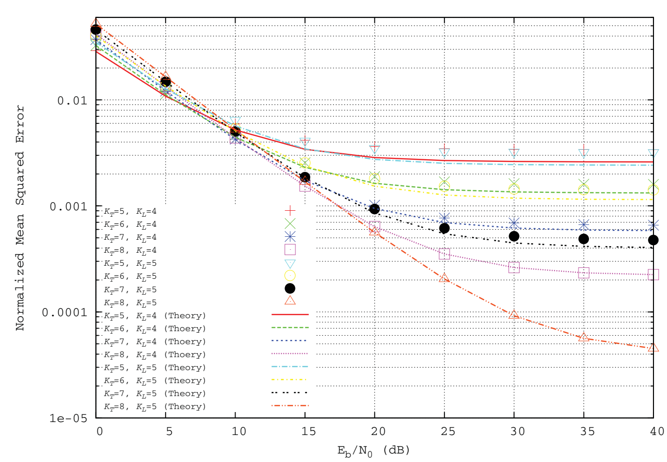

Fig. 4. Comparison of theoretical and simulated MSE performance of *Algorithm B* in a channel generated by the model described in [10]; AS=15° and f_dT_s=0.031772.

MSE performance in an IEEE 802.11 TGn channel [29] with $L = 12$, $\Delta = 15°$, and $f_dT_s = 0.0022$, while Fig. 6 shows the MSE performance in a 3GPP-SCM channel [30] with $L = 12$, $\Delta = 15°$ and $f_dT_s = 0.02844$. When K_T is large enough, the time-domain modelling order needed to characterize a slow fading channel like the IEEE 802.11 TGn channel is smaller than that for a fast fading SCM channel. Note that in all cases, the performance becomes independent of the AS when the full modelling order is used (i.e., $K_T = 8$) and is equivalent to that of the conventional LS approach.

The remaining numerical results assume that the algorithms developed in Section III are used and, except for Fig. 8, the same channel model as that used for Fig. 1. Fig. 7 compares the MSE performance of *Algorithms A* and *B* developed in Section III when $\Delta = 15°$. If the maximum matching output is obtained by selecting the best one from the outputs using 100 candidate phases uniformly distributed within $[-\pi, \pi)$, *Algorithm A* and *Algorithm B* give almost identical performance. However, if only 20 candidate phases are used, *Algorithm A* results in a little performance degradation with respect to that obtained by *Algorithm B* when SNR is high. Fig. 8 examines the MSE performance when $\widehat{\mathbf{W}}$ is updated with different EI lengths for various channel settings. Smaller performance loss results if the channel is more static or less dynamic (smaller f_dT_s). When $K_T \geq 3$ for Channel-1 [28] and $KT \geq 2$ for Channel-2 [29], the performance loss is negligible for all the update frequencies. Recall that more computation saving is obtained by a larger T_o^w. It is clear that our reduced-order modelling approach outperform the conventional LS estimator for most $\frac{E_b}{N_0}$ when a proper K_T is used.

To show the advantage of the proposed schemes in post processing, we demonstrate here a feedback eigen-beamforming scheme. This scheme can adapt to the feedback estimated CSI to optimize the reception performance in a spatial correlated environment [31]. To use the optimal beamforming scheme of [31] which minimizes the mean squared error between the transmit symbol and equalized received sample, we need to substitute $\mathbf{Q}_{\mathrm{R}}\widehat{\mathbf{C}}\mathbf{Q}_{\mathrm{T}}^T\widehat{\mathbf{W}}$ for $\mathbf{H}$ in (7), (16) and (22) of [31]. The flops of the eigen-decomposition operation needed by the beamforming system decrease from $O(M^3)$ to $O(K_T^3)$. Moreover, the total number of feedback floating-point variables in our approach is $K_T^2 + K_T + 1$ while that of the conventional CSI is $M^2 + M$. MSE in Fig. 9 is defined as the mean power of the error vector between a transmit data vector and the corresponding received/equalized vector. Simulation

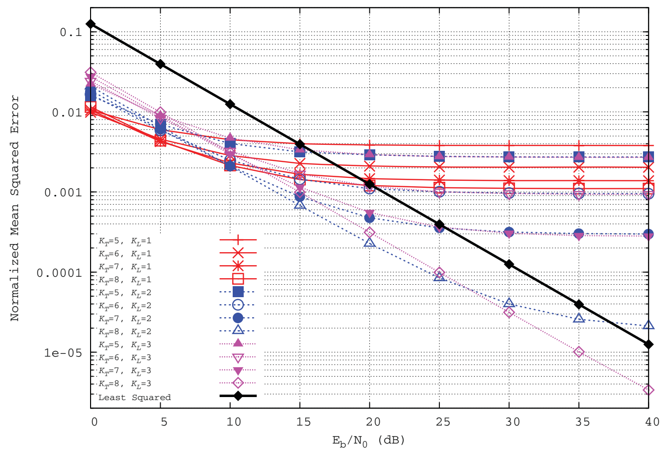

Fig. 5. The effect of the modelling order on the MSE performance of *Algorithm B* in a channel generated by IEEE 802.11 TGn channel model A; AS=15°, and f_dT_s =0.0022.

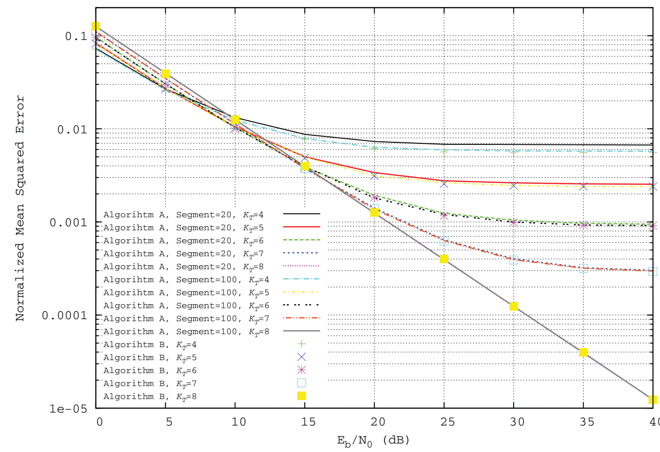

Fig. 7. MSE performance comparison of *Algorithm A* and *Algorithm B*; AS=15°.

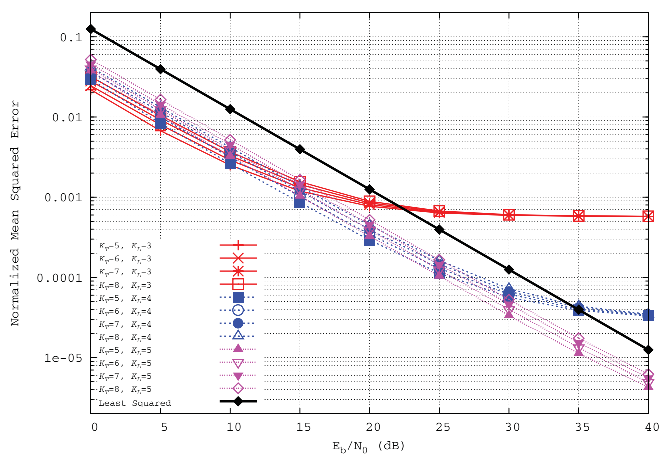

Fig. 6. The effect of the modelling order (K_T) on the MSE performance of *Algorithm B* in a 3GPP-SCM channel; AS=15° and f_dT_s=0.02844.

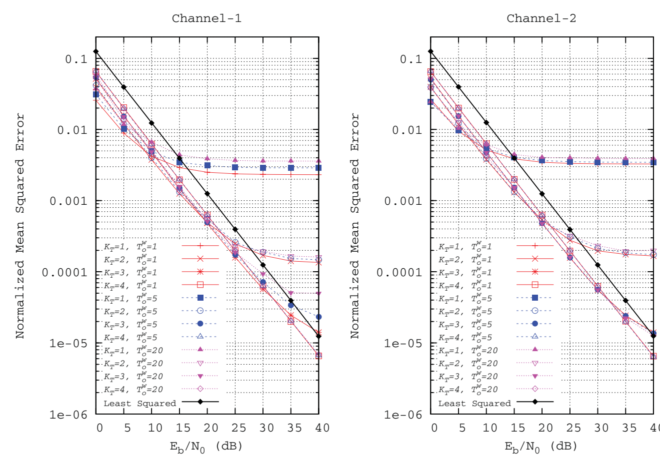

Fig. 8. The effect of the update period on the MSE performance of *Algorithm B*. Channel-1 is based on [28] with f_dT_s=0.015886 while Channel-2 is based on [29] with $f_dT_s = 0.0022$. AS=2°, $T_o^c = 1$; both T_o^c and T_o^w are measured in EIs.

results shown in Fig. 9 reveal that even for a weakly correlated environment, i.e. AS=15°, the performance degradation is negligible for $K_T \geq 4$ when $\frac{E_b}{N_0} \geq 12$ dB while there is performance improvement over the conventional LS approach when $\frac{E_b}{N_0} \leq 12$ dB. More performance and complexity-reduction improvements are achievable for channels with higher spatial correlation, e.g., when AS=4°.

VII. Conclusion

This paper presents a novel analytic model which spans the spatial and/or time correlation functions over the dominant signal subspace and provides additional directional information. Iterative algorithms are proposed for estimating spatial-correlated MIMO channels. These estimators are extended to time-varying cases in which the time-correlation has to be taken into account. We simulate the estimators' performance in various popular industry-approved and standardized channels to validate the accuracy of our model and the usefulness of our channel estimators. Numerical results show that in many instants the proposed algorithms give superior MSE performance. Our estimators offer tradeoffs between performance and complexity. They are easily extendable for use in wide-band MIMO systems and are most effective when the channel's AS is small, i.e., when the dimension of the dominant subspace is much smaller than full channel correlation rank. Not only do they offer fast and accurate estimates, give MSE performance improvement due to the noise reduction effect but, more importantly, also provide compact and useful CSI that lead to significant feedback channel bandwidth reduction and other potential post processing complexity cutbacks.

Appendix A
AoD Information Extraction

For small Δ, the correlation between two transmit antennas i, j can be approximated by [2]

$$E\left\{h_{mi}h_{mj}^*\right\} \approx \exp\left\{-j\frac{2\pi}{\lambda}(i-j)d\sin\phi\right\} J_0\left(\Delta\frac{2\pi}{\lambda}(i-j)d\cos\phi\right). \quad \text{(A.1)}$$

In addition, correlation between two receive antennas p, q can be approximated by $E\left\{h_{pi}h_{qi}^*\right\} \approx J_0\left(\frac{2\pi}{\lambda}(p-q)d\right)$, for $\frac{d}{R} \ll 1$. By using the $\mathbf{W}$ defined in (14), the definition $\tilde{\Delta} \stackrel{def}{=} \frac{2\pi d}{\lambda}\Delta\cos\phi$ and (A.1) implies $\mathbf{\Phi}_{\mathrm{T}} \stackrel{def}{=} \mathbf{W}\bar{\mathbf{\Phi}}_{\mathrm{T}}\mathbf{W}^H =$

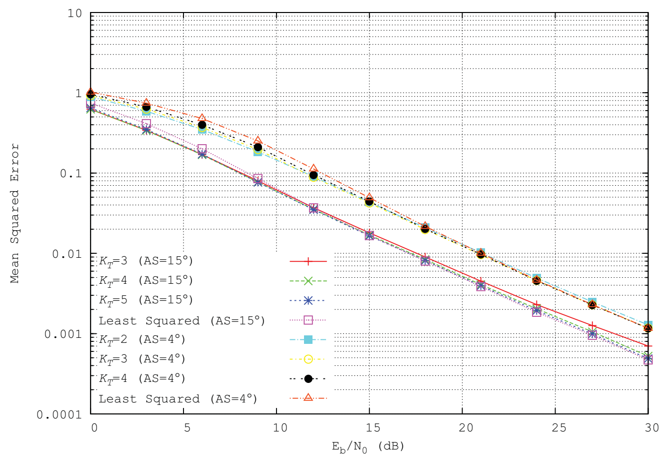

Fig. 9. MSE performance of a feedback beamforming system using the estimates produced by *Algorithms B*; 10 iterations, AS=15° and AS=4°.

$\mathbf{W}\bar{\mathbf{\Phi}}_{\mathrm{T}}^{\frac{1}{2}}\bar{\mathbf{\Phi}}_{\mathrm{T}}^{\frac{1}{2}H}\mathbf{W}^H$, where $\bar{\mathbf{\Phi}}_{\mathrm{T}} = \left[\left\{J_0\left(|i-j|\tilde{\Delta}\right)\right\}\right]$, for $1 \leq i,j \leq M$, and thus $\mathbf{\Phi}_{\mathrm{T}}^{\frac{1}{2}} = \mathbf{W}\bar{\mathbf{\Phi}}_{\mathrm{T}}^{\frac{1}{2}}$. The correlation matrix at the receive site can also be decomposed as $\mathbf{\Phi}_{\mathrm{R}} = \left[\left\{J_0\left(|i-j|\tilde{d}\right)\right\}\right]$, for $1 \leq i,j \leq N$ and $\tilde{d} \overset{def}{=} \frac{2\pi d}{\lambda}$. The above two equations immediately lead to (A.2). Hence, (2) is equivalent to

$$\mathbf{H} = \bar{\mathbf{\Phi}}_{\mathrm{R}}^{\frac{1}{2}}\mathbf{H}_w\bar{\mathbf{\Phi}}_{\mathrm{T}}^{\frac{1}{2}\ T}\mathbf{W}, \tag{A.2}$$

where $\bar{\mathbf{\Phi}}_{\mathrm{T}}$ and $\bar{\mathbf{\Phi}}_{\mathrm{R}}$ denote the power correlation matrices at the transmit and the receive sites, respectively. Using $\mathbf{\Phi}_{\mathrm{T}}^{1/2} = \mathbf{W}\bar{\mathbf{\Phi}}_{\mathrm{T}}^{1/2}$ and following a procedure similar to (9)–(12), we obtain (13) of the main text. Note that Forenza *et al.* [18] have recently showed that, for a clustered MIMO channel with uniform linear or circular array, the cross-correlation coefficients also have a regression form similar to (A.1). Hence if we assume a similar environment, we will obtain an analytical model of the same form as (13).

In the above single-directional model, the AoD from the transmitting antennas at the transmitter can be captured by a mean AoD. In contrast, the principle of maximum entropy [19] assumes i.i.d. uniformly distributed AoA angles over $[0, 2\pi]$ and leaves no mean arriving direction being modelled at the mobile side. It models the separate power azimuthal spectra (PAS) of AoA and AoD, with a common direction being described by the mean AoD at the base station [20].

Appendix B
Proof of Lemma 1

According to Lemma 5.1.3 of [17], the ith entry of the vector $\left[(\mathbf{1}_E \otimes \mathbf{A}) \odot (\mathbf{B}^T \otimes \mathbf{1}_N)\right]\mathbf{c}$ is identical to the (i,i)th diagonal entry of the square matrix $\left[(\mathbf{1}_E \otimes \mathbf{A})\mathrm{diag}(\mathbf{c})(\mathbf{B} \otimes \mathbf{1}_N^T)\right]$, for $i = 1, 2, \ldots, NE$. Define $\tilde{\mathbf{A}} = [\tilde{a}_{m,n}] \overset{def}{=} (\mathbf{1}_E \otimes \mathbf{A})$ and $\tilde{\mathbf{B}} = [\tilde{b}_{m,n}] \overset{def}{=} (\mathbf{B}^T \otimes \mathbf{1}_N)$. Then, for $i = N(p-1)+q$, $p \in \{1, \ldots, E\}$ and $q \in \{1, \ldots, N\}$, we have

$$\begin{aligned}\left[(\mathbf{1}_E \otimes \mathbf{A})\mathrm{diag}(\mathbf{c})(\mathbf{B} \otimes \mathbf{1}_N^T)\right]_{i,i} &= \sum_{j=1}^{M} \tilde{a}_{i,j}c_j\tilde{b}_{i,j} \\ &= \sum_{j=1}^{M} a_{q,j}c_j b_{j,p} = \left[\mathrm{vec}\left(\mathbf{A}\ \mathrm{diag}(\mathbf{c})\ \mathbf{B}\right)\right]_i, \end{aligned} \tag{B.1}$$

where $a_{q,j}$ is the (q,j)th entry of $\mathbf{A}$ and $b_{j,p}$ is the (j,p)th entry of $\mathbf{B}$. $[\mathbf{D}]_{i,i}$ denotes the (i,i)th entry of the matrix $\mathbf{D}$, while $[\mathbf{e}]_i$ denotes the ith entry of the vector $\mathbf{e}$. Hence we conclude that

$$\left[\left((\mathbf{1}_E \otimes \mathbf{A}) \odot (\mathbf{B}^T \otimes \mathbf{1}_N)\right)\mathbf{c}\right]_i = \left[\mathrm{vec}\left(\mathbf{A}\ \mathrm{diag}(\mathbf{c})\ \mathbf{B}\right)\right]_i, \quad \forall\ i = 1, \ldots, NE, \tag{B.2}$$

which proves *Lemma 1*.

References

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communication in a fading environment when using multiple antennas," *Wireless Personal Commun.,* vol. 6, pp. 311-335, Mar. 1998.
[2] D. S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.,* vol. 48, no. 3, pp. 502-513, Mar. 2000.
[3] W. Weichselberger, M. Herdin, H. Özcelik, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 90-100, Jan. 2006.
[4] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp.2563-2579, Oct. 2002.
[5] V. V. Veeravalli, Y. Liang, and A. M. Sayeed, "Correlated MIMO wireless channels: capacity, optimal signaling, and asymptotics," *IEEE Trans. Inf. Theory,* vol. 51, no. 6, pp.2058-2072, June 2005.
[6] X. Gao, B. Jiang, X. Li, A. B. Gershman, and M. R. McKay, "Statistical eigenmode transmission over jointly correlated MIMO channels," *IEEE Trans. Inf. Theory,* vol. 55, no. 8, pp.3735-3750, Aug. 2009.
[7] P. Nagvanshi, E. Masry, and L. Milstein, "Optimum transmit-receive beamforming with noisy channel estimates for correlated MIMO Rayleigh channels," *IEEE Trans. Commun.*, vol. 56, no. 11, pp. 1869-1880, Nov. 2008.
[8] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884-893, Mar. 2006.
[9] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. Wireless Commun.,* vol. 4, no. 2, pp. 453-466, Mar. 2005.
[10] K. I. Pedersen, J. B. Andersen, J. P. Kermoal, and P. E. Mogensen, "A stochastic multiple-input multiple-output radio channel model for evaluation of space-time coding algorithms," in *Proc. Vehicular Technology Conf.,* Boston, MA, Sept. 2000, pp. 893-897.
[11] L. Schumacher, J. P. Kermoal, F. Frederiksen, K. I. Pedersen, A. Algans, and P. Mogensen, "MIMO channel characterisation," IST Project IST-1999-11729 METRA Deliverable D2, Feb. 2001 (http://www.ist-imetra.org/metra/index_deliverables.html).
[12] A. Abdi and M. Kaveh, "A space-time correlation model for multielement antenna systems in mobile fading channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 550-560, Apr. 2002.
[13] X. Wang and K. J. R. Liu, "Model-based channel estimation framework for MIMO multicarrier communication systems," *IEEE Trans. Wireless Commun.,* vol. 4, no. 3, pp. 1050-1063, May 2005.
[14] Y. H. Kho and D. P. Taylor, "MIMO channel estimation and tracking based on polynomial prediction with application to equalization," *IEEE Trans. Veh. Technol.,* vol. 57, no. 3, pp. 1585-1595, May 2008.
[15] D. K. Borah and B. D. Hart, "Frequency-selective fading channel estimation with a polynomial time-varying channel model," *IEEE Trans. Commun.,* vol. 47, no. 6, pp. 862-873, Jun. 1999.
[16] M.-X. Chang and Y. T. Su, "Model-based channel estimation for OFDM signals in Rayleigh fading," *IEEE Trans. Commun.*, vol. 50, no. 4, pp. 540-544, Apr. 2002.
[17] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis.* Cambridge, U.K.: Cambridge Univ. Press, 1991.

[18] A. Forenza, D. J. Love, and R. W. Heath, Jr., "Simplified spatial correlation models for clustered MIMO channels with different array configurations," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 1924-1933, July 2007.
[19] M. Debbah and R. R. Müller, "MIMO channel modeling and the principle of maximum entropy," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1667-1690, May 2005.
[20] H. Xu, D. Chizhik, H. Huang, and R. Valenzuela, "A generalized space-time multiple-input multiple-output (MIMO) channel model," *IEEE Trans. Wireless Commun.,* vol. 3, no. 3, pp. 966-975, May 2004.
[21] F. B. Gross, "New approximation to J_0 and J_1 Bessel functions," *IEEE Trans. Antennas and Propag.,* vol. 43, no. 8, pp. 904-907, Aug. 1995.
[22] L. S. Scharf, *Statistical Signal Processing.* Reading, MA: Addison-Wesley, 1991.
[23] A. D. R. McQuarrie and C.-L. Tsai, *Regression and Time Series Model Selection.* World Scientific, 1998.
[24] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computations.* Englewood Cliffs, NJ: Prentice-Hall, 1989.
[25] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Johns Hopkins, 3rd edition, 1996.
[26] W. Zhang, X.-G. Xia, and P. C. Ching, "Optimal training and pilot pattern design for OFDM systems in Rayleigh fading," *IEEE Trans. Broadcast.*, vol. 52, no. 4, pp. 505-514, Dec. 2006.
[27] A. Swindlehurst and J. Yang, "Using least squares to improve blind signal copy performance," *IEEE Signal Process. Lett.*, vol. 1, no. 5, pp. 80-82, May 1994.
[28] J. P. Kermoal, L. Schumacher, and P. Mogensen, "MIMO channel characterisation," IST Project IST-2000-30148 I-METRA Deliverable D2, Oct. 2002 (http://www.ist-imetra.org/index_deliverables.html).
[29] V. Erceg, L. Schumacher, P. Kyritsi, *et al.*, "TGn channel models," IEEE 802.11 document 03/940r4, IEEE, New York, NY, USA, Jan. 2004.
[30] 3GPP, TR 25.996, "Spatial channel model for multiple input multiple output (MIMO) simulations (Rel. 6)," 2003 (http://www.3gpp.org/ftp/Specs/html-info/25996.htm).
[31] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. Signal Process.,* vol. 52, no. 5, pp. 1179-1197, May 2004.

Yen-Chih Chen received the B.S. degree in control engineering from the National Chiao Tung University, Hsinchu, Taiwan, in 1996, the M.S. degree in electrical engineering from the National Taiwan University of Science and Technology, Taipei, Taiwan, in 1999, and the Ph.D. degree in communications engineering from the National Chiao Tung University, Hsinchu, Taiwan, in 2009. He was with Computer and Communications Laboratories, Industrial Technology Research Institute during 1999-2000 and has been with Realtek Inc., Taiwan since 2006. His main research interests include MIMO communications, coding theory and statistical signal processing.

Yu T. Su received the B.S.E.E. degree from Tatung Institute of Technology, Taipei, Taiwan and Ph.D. degree from the University of Southern California, Los Angeles, USA, in 1974 and 1983, respectively. From 1983 to 1989, he was with LinCom Corporation, Los Angeles, USA, where his last position was a Corporate Scientist. Since September 1989, he has been with the National Chiao Tung University, Hsinchu, Taiwan and is currently a professor in the Department of Electrical Engineering. He was an Associative Dean of the College of Electrical and Computer Engineering from 2004 to 2007, Head of the Communications Engineering Department from 2001 to 2003. He is also affiliated with the Microelectronic and Information Systems Research Center of the same university and served as a Deputy Director during 1997-2000. From 2005 to 2008, he was the Area Coordinator of Taiwan National Science Council's Telecommunications Programme. His main research interests include communication theory and statistical signal processing.