View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Drexel Libraries E-Repository and Archives

Blind MIMO System Estimation Based on PARAFAC Decomposition of Higher Order Output Tensors

Turev Acar, *Member, IEEE*, Yuanning Yu, *Student Member, IEEE*, and Athina P. Petropulu, *Senior Member, IEEE*

***Abstract*—We present a novel framework for the identification of a multiple-input multiple-output (MIMO) system driven by white, mutually independent unobservable inputs. Samples of the system frequency response are obtained based on parallel factorization (PARAFAC) of three- or four-way tensors constructed based on, respectively, third- or fourth-order cross spectra of the system outputs. The main difficulties in frequency-domain methods are frequency-dependent permutation and filtering ambiguities. We show that the information available in the higher order spectra allows for the ambiguities to be resolved up to a constant scaling and permutation ambiguities and a linear phase ambiguity. Important features of the proposed approach are that it does not require channel length information, needs no phase unwrapping, and unlike the majority of existing methods, needs no prewhitening of the system outputs.**

***Index Terms*—Blind system identification, convolutive MIMO, higher order statistics, multiple-input multiple-output (MIMO), parallel factorization (PARAFAC).**

I. Introduction

We consider a linear time-invariant (LTI) multiple-input multiple-output (MIMO) system driven by unobservable inputs. Our goal is to identify the system function based on the system outputs. MIMO models arise frequently in speech processing, multiaccess communication, multitrack digital magnetic recording, and biomedical applications, [26], [34], [45].

The case of a memoryless (or scalar) system excited by white inputs has been studied under the name of independent component analysis (ICA) [14]. In this paper, we focus on the identification of convolutive MIMO systems. Among the possible approaches, [15], [27], [39], [46], frequency-domain methods offer certain advantages over time domain ones [5], [9], [12], [16], [17]; they do not require system length information, and also, their formulation can take advantage of existing results for the memoryless MIMO problem. Indeed, in the frequency domain, at each frequency, the convolutive problem is transformed into a scalar one. However, an additional step is required to resolve the frequency-dependent permutation, scaling and phase ambiguities.

Manuscript received July 15, 2004; revised October 31, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Andrew C. Singer. The work was supported by ONR under grant N00014-02-1-0137, and NSF under Grant CNS-0435052.

T. Acar was with the Electrical and Computer Engineering Department, Drexel University, Philadelphia, PA 19104 USA. He is now with Maxim Integrated Products, Sunnyvale, CA 94086 USA (email: Turev_Acar@maximhq.com).

Y. Yu and A. P. Petropulu are with the Electrical and Computer Engineering Department, Drexel University, Philadelphia, PA 19104 USA (e-mail: Yuanning@cbis.ece.drexel.edu; athina@ece.drexel.edu).

Digital Object Identifier 10.1109/TSP.2006.879327

Most of the blind convolutive MIMO system identification methods in the literature exploit either second-order statistics (SOS) or higher order statistics (HOS). SOS-based methods, as opposed to HOS based ones, do not require long data in order to obtain good estimates and involve low complexity. Examples of SOS-based methods can be found in [2], [3], [5], [17], and [48]. All these methods require channel diversity, and they apply to nonwhite inputs only. On the other hand, HOS-based methods provide system information without requiring channel diversity and also can deal with white inputs as long as they are non-Gaussian. Examples of HOS MIMO methods can be found in [12], [16], [20], [27], and [46].

In [16], a frequency-domain approach was proposed for the convolutive MIMO case with white inputs. The system response was obtained based on joint diagonalization of output HOS slices. The result of the joint diagonalization contained a frequency-domain scaling ambiguity, which could be reduced to a constant scaling ambiguity by exploiting the redundancy in the higher order cumulant domain. In order for joint diagonalization to be applicable, the system frequency response at each frequency would need to be a unitary matrix. To guarantee this, a prewhitening operation was applied to the system output. Similar prewhitening is employed in the majority of HOS-based blind MIMO estimation methods [3], [11], [12], [33]. However, prewhitening is a sensitive process as it tends to lengthen the global system response, and as a result increases complexity and estimation errors. The obtained estimate contains the contribution of both the whitening filter and the MIMO system, and extracting the MIMO system response from this estimate is a sensitive task.

The need for whitening can be obviated by another decomposition that does not require unitary matrices. One such approach is the PARAFAC decomposition, which is a low-rank decomposition of three- or higher way arrays. It was first was developed in [13] in order to overcome the rotation problem which exists in two-dimensional (2-D) arrays, and later generalized in [10], [23]–[25], and [30]. The PARAFAC decomposition can be thought as an extension of singular value decomposition to multiway arrays, where uniqueness is guaranteed even if the nondiagonal matrices involved are nonunitary.

In this paper, we show how the PARAFAC ideas can be used in the frequency-domain framework of [16] to avoid the need for prewhitening. The decomposition is applied on one or more tensors that are formed based on HOS of the system output. Immediately following the PARAFAC decomposition the system frequency response matrix can be computed within a constant permutation ambiguity and a frequency dependent diagonal scaling ambiguity. The latter ambiguity can be reduced to a diagonal

1053-587X/$20.00 © 2006 IEEE

constant scaling ambiguity and a diagonal linear phase ambiguity via an iterative scheme.

The paper is organized as follows. In Section II, we summarize the results on PARAFAC decomposition that we will need. In Section III, we formulate the problem and list all required assumptions. In Section IV, we present the main results based on third-order statistics. In Section V, we extend the previous results to employ fourth-order statistics. Implementation issues are discussed in Section VI, and potential applications are discussed in Section VII. Simulation results on estimation performance are given in Section VIII. Finally, concluding remarks are made in Section IX.

Notation:

- Superscripts T, H, and * denote transpose, Hermitian transpose, and complex conjugate operations, respectively.
- $\mathrm{Cum}[x_1, x_2, \ldots, x_n]$ denotes the nth-order cumulant of the random variables $x_1, \ldots, x_n$.
- Boldface symbols denote matrices. Capital calligraphic symbols denote tensors.
- $\mathbf{D}_l[\mathbf{X}]$ denotes a diagonal matrix, whose diagonal contains the lth row of matrix $\mathbf{X}$.
- $\mathrm{diag}(\mathbf{X})$ denotes a vector formed based on the diagonal elements of matrix $\mathbf{X}$.
- $\mathrm{Diag}(x_1, \ldots x_n)$ is a diagonal matrix whose diagonal elements are $x_1, \ldots x_n$.
- $((.))_k$ denotes modulo k.

II. Parallel Factorization

Let us consider a three-way tensor $\mathcal{X}$ with dimensions $I \times J \times K$, given that its element x is indexed by i, j, k, and the F-component decomposition [28]

$$x_{i,j,k} = \sum_{f=1}^{F} a_{i,f} b_{j,f} c_{k,f}. \tag{1}$$

Equation (1) expresses the three-way array $\mathcal{X}$ as a sum of F rank-one three-way factors, each one of which is the outer product of three factors.

In a compact form, $\mathcal{X}$ can be expressed in terms of its slices $\mathbf{X}_i$ $(J \times K)$, $i = 1, \ldots, I$ as

$$\mathbf{X}_i = \mathbf{B}\mathbf{D}_i[\mathbf{A}]\mathbf{C}^T \tag{2}$$

where $\mathbf{A}$ is an $I \times F$ matrix with entries $a_{i,f}$; $\mathbf{B}$ is a $J \times F$ matrix with entries $b_{j,f}$; $\mathbf{C}$ is a $K \times F$ matrix with entries $c_{k,f}$.

Under certain conditions, the tensor $\mathcal{X}$ can be decomposed uniquely into matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$. These conditions are based on the notion of k-rank [21], [25].

Definition 1: Consider an $I \times J$ matrix $\mathbf{A}$. If $\mathrm{rank}(\mathbf{A}) = r$ then $\mathbf{A}$ contains r linearly independent columns. The matrix $\mathbf{A}$ has k-rank $k_A = l$ if every $l \leq J$ columns of $\mathbf{A}$ are linearly independent, but either $l = J$, or there exist a collection of $l+1$ linearly dependent columns in $\mathbf{A}$.

Note that the $k_A \leq \mathrm{rank}(\mathbf{A}) \leq \min(I, J)$.

Theorem 1 [25], [30], [40]: Let $\mathcal{X}$ be a tensor whose slice $\mathbf{X}_i$ is given as in (2). $\mathcal{X}$ can be decomposed into $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ uniquely up to permutation and scaling ambiguities if

$$k_A + k_B + k_C \geq 2F + 2. \tag{3}$$

The condition of (3) is sufficient but not necessary for the unique decomposition of $\mathcal{X}$ [4], [6], [7], [28].

Several algorithms exist for decomposing tensor $\mathcal{X}$ into components $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. [21]. The method that we use in this paper is COMplex parallel FACtor analysis (COMFAC), which is a fast least square PARAFAC algorithm applied on a compressed version of the data [21], [29]. COMFAC consists of first compressing the array, second initializing and fitting the PARAFAC model on that compressed array, and finally decomposing and refining the solution in the raw data space.

III. Problem Formulation

Let us consider a N_i-input N_o-output LTI system. Let $\mathbf{s}(n) = [s_1(n) \ldots s_{N_i}(n)]^T_{n \in \mathcal{Z}}$ be a vector of N_i inputs; $\mathbf{h}(l)$ the finite-impulse-response (FIR) MIMO system impulse response matrix whose (i, j) element is denoted by $\{h_{ij}(l)\}_{1 \leq i \leq N_o, 1 \leq j \leq N_i}$; $\mathbf{x}(n) = [x_1(n) \ldots x_{N_o}(n)]^T_{n \in \mathcal{Z}}$ the vector of observations; and $\mathbf{w}(n) = [w_1(n) \ldots w_{N_o}(n)]^T_{n \in \mathcal{Z}}$ observation noise. Here, n denotes discrete time. All signals can be real or complex.

The MIMO system output equals

$$\mathbf{x}(n) = \sum_{l=0}^{L-1} \mathbf{h}(l)\mathbf{s}(n-l) + \mathbf{w}(n) \tag{4}$$

where L is the length of the longest $h_{ij}(l)$.

Let $\mathbf{H}(k)$ $(N_o \times N_i)$ be the N-point discrete Fourier transform (DFT) of $\mathbf{h}(n)$, i.e.,

$$\mathbf{H}(k) = \sum_{n=0}^{L-1} \mathbf{h}(n) e^{-j\frac{2\pi}{N}kn}, \;\; k = 0, \ldots, N-1 \tag{5}$$

where $N > L$. Our goal is to estimate $\mathbf{H}(k)$ based on the system output.

The problem contains inherent ambiguities. At best, we can find $\hat{\mathbf{H}}(k)$ such that [16]

$$\hat{\mathbf{H}}(k) = \mathbf{H}(k)\mathbf{P}\mathbf{\Lambda} e^{j\frac{2\pi}{N}k\mathbf{M}} \tag{6}$$

where $\mathbf{P}$ is a column permutation matrix, $\mathbf{\Lambda}$ a constant diagonal matrix and $\mathbf{M}$ diagonal matrix with integer elements. Equivalently, the impulse response matrix, at best, can be found within a column permutation matrix, a constant diagonal matrix, and a cyclic shift. We will refer to these ambiguities as *trivial ambiguities*. Using DFT properties, it is easy to see that if such a system estimate was used to recover the inputs, it would yield each input within a scalar ambiguity and a circular shift. In addition, the inputs would be recovered in some unknown order [16].

We next provide a list of all the assumptions considered in this paper. The subset of assumptions that will be needed in each case will be stated along with the proposed methods.

A1) Each $s_i(.)$ is a zero mean, nonsymmetrically distributed, independent identically distributed (i.i.d.), stationary process with nonzero skewness, i.e., $\gamma^3_{s_i} = \mathrm{Cum}[s_i(n), s_i^*(n), s_p(n)]$. The s_i's are mutually independent.

A2) The matrix $\mathbf{H}(k)$ is invertible for all $k = 0, \ldots, N-1$.

A3) The k-rank of $\mathbf{H}(k)$ satisfies $3k_H \geq 2N_i + 2$.

A4) $w_i(.), i = 1, \ldots, N_o$ are zero-mean Gaussian stationary random processes with variance σ_w^2, mutually independent and independent of the inputs.

A5) Each $s_i(.)$ is a zero mean, i.i.d., stationary process with nonzero kurtosis. The s_i's are mutually independent.

A6) The k-rank of $\mathbf{H}(k)$ satisfies $4k_H \geq 2N_i + 3$.

Discussion on Assumptions: Assumption A1) guarantees that there are nonzero samples in the third-order cumulant sequence of the system output. It will be used in the proposed methods that involve third-order statistics. Assumption A2) requires that the channel matrix $\mathbf{H}(k)$ is full column rank for all k's. As it will be seen later, this is the strongest channel assumption made by the methods to be proposed, but still, it is less stringent than the assumptions made in [16]. We note that time-domain methods do not require such assumption; however, they require channel length information and are sensitive to length mismatch. A matrix whose columns are drawn independently from an absolutely continuous distribution is both full rank and full k-rank with probability one [40]. Thus, if the elements of matrix $\mathbf{h}(n)$ are independent for all n, $\mathbf{h}(n)$, and thus also $\mathbf{H}(k)$, are full rank and also full k-rank. In the latter case, condition A3) is equivalent to requiring that $3\min(N_i, N_o) \geq 2N_i + 2$. For $N_i \leq N_o$ A3) is satisfied for $N_i \geq 2$. Assumption A4) is needed in order for additive noise to be suppressed in the higher order cumulant domain. Assumption A5) requires that the fourth-order cumulants of the inputs are not identically zero. Assumption A5), unlike assumption A1), is satisfied by most communication signals. Under A2), and assuming that $N_i \leq N_o$, A6) is satisfied for $N_i \geq 2$.

IV. Channel Estimation

Under assumptions A1) and A4), the third-order cross cumulants of the system outputs $x_l(k)$, $x_i^*(k)$, $x_j(k)$ [35] equal

$$c^3_{lij}(\tau,\rho) \triangleq \mathrm{Cum}\left[x_l(k), x_i^*(k+\tau), x_j(k+\rho)\right],\ |\tau|,|\rho| < L = \sum_{p=1}^{N_i} \gamma^3_{s_p} \sum_{m=0}^{L-1} h_{lp}(m) h^*_{ip}(m+\tau) h_{jp}(m+\rho) \quad (7)$$

where $l, i, j \in [1, \ldots, N_o]$.

The $N \times N$ discrete-frequency cross bispectrum of $x_l(k)$, $x_i^*(k)$, and $x_j(k)$ is the 2-D DFT of $c^3_{lij}(\tau,\rho)$, and equals

$$C^3_{lij}(k_1,k_2) = \sum_{p=1}^{N_i} \gamma^3_{s_p} H_{lp}(-k_1-k_2) H^*_{ip}(-k_1) H_{jp}(k_2), \quad k_1,k_2 = 0,\ldots,N-1. \quad (8)$$

For fixed k_1 and k_2, $C^3_{lij}(k_1,k_2)$ can be viewed as the (l,i,j)th element of tensor $\mathcal{C}^3(k_1,k_2)$ $(N_o \times N_o \times N_o)$. The lth slice of that tensor equals

$$\mathbf{C}^3_l(k_1,k_2) = \mathbf{H}^*(-k_1)\mathbf{\Gamma}^3\mathbf{D}_l\left[\mathbf{H}(-k_1-k_2)\right]\mathbf{H}^T(k_2) \quad (9)$$

where $\mathbf{\Gamma}^3 = \mathrm{Diag}\{\gamma^3_{s_1}, \ldots, \gamma^3_{s_{N_i}}\}$.

In the following, we will consider the tensors $\mathcal{C}^3(-m+r\delta,\delta)$, $r = 0,1,2,..,N-1$, for some constant m, δ, and show how they can be used to recover $\mathbf{H}(m-r\delta-2\delta)$ for each r. The choice of δ, m will affect the estimation result. The criterion for selection of these parameters will be discussed in a later section.

Let us define

$$\mathbf{A}_r \triangleq \mathbf{H}(m-r\delta-\delta) \quad (10)$$

$$\mathbf{B}_r \triangleq \mathbf{H}^*(m-r\delta)\mathbf{\Gamma}^3 \quad (11)$$

$$\mathbf{C}_r \triangleq \mathbf{H}(\delta). \quad (12)$$

Under assumption A3) and via Theorem 1, the tensor $\mathcal{C}^3(-m+r\delta,\delta)$ can be decomposed into

$$\hat{\mathbf{A}}_r = \mathbf{A}_r\mathbf{P}_{1r}\mathbf{\Lambda}_{1r}, \quad \hat{\mathbf{B}}_r = \mathbf{B}_r\mathbf{P}_{2r}\mathbf{\Lambda}_{2r}, \quad \hat{\mathbf{C}}_r = \mathbf{C}_r\mathbf{P}_{3r}\mathbf{\Lambda}_{3r} \quad (13)$$

where $\mathbf{P}_{ir}$ is a permutation matrix, and $\mathbf{\Lambda}_{ir}$ is a complex diagonal matrix. Both $\mathbf{P}_{ir}$ and $\mathbf{\Lambda}_{ir}$ depend on δ and m; however, since in the sequel we will only vary r, that dependence is not shown mainly for notational convenience.

If N and δ are not coprime integers, then $\mathbf{A}_r$, $\mathbf{B}_r$, $\mathbf{C}_r$, $\hat{\mathbf{A}}_r$, $\hat{\mathbf{B}}_r$, and $\hat{\mathbf{C}}_r$ will be periodic with period N/δ. In that case, by varying r, $\mathbf{A}_r$ or $\mathbf{B}_r$ provide N/δ independent samples of the system frequency response matrix. Since for the recovery of an impulse response of length L we only need L frequency response samples, as long as $(N/\delta) > L$, we can still recover the channel matrix. However, the more frequency-domain samples we obtain, the lower the estimation error in the system impulse response will be.

On the other hand, if N, δ are coprime integers, $\mathbf{A}_r$ and $\mathbf{B}_r$ are periodic with period N. As r takes values from 0 to $N-1$, $\mathbf{A}_r$ becomes equal to the samples of the channel matrix starting from sample $m-\delta$, and moving to the left in steps of δ. Due to the periodic extension of the DFT, and since N, δ are coprime, eventually all samples of the channel matrix will be visited. $\hat{\mathbf{A}}_r$ and $\hat{\mathbf{B}}_r$ will also be periodic; thus, $\mathbf{\Lambda}_{1r}$, $\mathbf{\Lambda}_{2r}$ will be periodic with period N.

Proposition 1: Consider the PARAFAC decomposition of the tensor $\mathcal{C}^3(-m+r\delta,\delta)$ into $\hat{\mathbf{A}}_r$, $\hat{\mathbf{B}}_r$, $\hat{\mathbf{C}}_r$, as defined in (13). Under assumption A3), it holds

$$\mathbf{P}_{1r} = \mathbf{P}_{2r} = \mathbf{P}_{3r} = \mathbf{P}_r, \quad \mathbf{\Lambda}_{2r}\mathbf{\Lambda}_{1r}\mathbf{\Lambda}_{3r} = \mathbf{I}. \quad (14)$$

The proof of Proposition 1 can be found in [25] and [40]. For the reader's convenience, it is also given in Appendix I.

According to Proposition 1, the decomposition of $\mathcal{C}^3(-m+r\delta,\delta)$ results in the same permutation ambiguity, $\mathbf{P}_r$, in all three terms, $\hat{\mathbf{A}}_r$, $\hat{\mathbf{B}}_r$, $\hat{\mathbf{C}}_r$. Noting that $\mathbf{C}_r$ is independent of r (see

(12)), it turns out that the permutation ambiguity in the decomposition of $\mathcal{C}^3(-m+r\delta,\delta)$ for different r's can be made independent of r. Consider the operation that normalizes the elements of matrix $\hat{\mathbf{C}}_r$ so that its first row consists of ones. The normalized matrix equals $\tilde{\hat{\mathbf{C}}}_r = \hat{\mathbf{C}}_r(\mathbf{D}_1[\hat{\mathbf{C}}_r])^{-1}$. The permutation of $\tilde{\hat{\mathbf{C}}}_r$ with respect to $\tilde{\hat{\mathbf{C}}}_0$ equals $\tilde{\hat{\mathbf{C}}}_0^{-1}\tilde{\hat{\mathbf{C}}}_r$.

Let us column-order $\hat{\mathbf{A}}_r$, $\hat{\mathbf{B}}_r$, and $\hat{\mathbf{C}}_r$ according to the order of columns of $\hat{\mathbf{C}}_0$ for all r. The column-ordered components are

$$\hat{\mathbf{A}}'_r \triangleq \hat{\mathbf{A}}_r\tilde{\hat{\mathbf{C}}}_r^{-1}\tilde{\hat{\mathbf{C}}}_0 = \mathbf{A}_r\mathbf{P}\mathbf{\Lambda}_{1r} = \mathbf{H}(m-r\delta-\delta)\mathbf{P}\mathbf{\Lambda}_{1r} \quad (15)$$

$$\hat{\mathbf{B}}'_r \triangleq \hat{\mathbf{B}}_r\tilde{\hat{\mathbf{C}}}_r^{-1}\tilde{\hat{\mathbf{C}}}_0 = \mathbf{B}_r\mathbf{P}\mathbf{\Lambda}_{2r} \quad (16)$$

$$\hat{\mathbf{C}}'_r \triangleq \hat{\mathbf{C}}_r\tilde{\hat{\mathbf{C}}}_r^{-1}\tilde{\hat{\mathbf{C}}}_0 = \mathbf{C}_r\mathbf{P}\mathbf{\Lambda}_{3r} \quad (17)$$

where $\mathbf{P}$ is a constant permutation matrix.

A. Channel Estimation Based on Decomposition of Multiple Tensors: Multiple PARAFAC Decomposition (MPD) Approach

Equation (15) indicates that, under assumption A3), the decomposition of tensors $\mathcal{C}^3(-m+r\delta,\delta)$, $r=0,\ldots,N-1$, followed by column-ordering of the matrix components yields $\mathbf{H}(m-r\delta-\delta)$, $r=0,\ldots,N-1$ within a constant permutation ambiguity and a diagonal scalar ambiguity $\mathbf{\Lambda}_{1r}$. As a result of the latter ambiguity, all elements of the ith column of $\mathbf{h}(n)$ will be recovered within the same filtering ambiguity. We will refer to this estimation approach as MPD-FA (muliple PARAFAC decomposition with filtering ambiguity). The latter ambiguity would be reflected as filtering ambiguity on each input and could be resolved using a single-input single-output (SISO) blind channel estimation approach.

It is interesting to note that MPD-FA also applies to MIMO systems that have more inputs than outputs. In such cases $\mathbf{H}(k)$ is a fat matrix. Assuming that the channel matrix elements are random and independent, then the k-rank of $\mathbf{H}(k)$ would be $k_H = \min\{N_o, N_i\} = N_o$ and requirement A3) would become

$$3k_H = 3N_o \geq 2N_i + 2. \quad (18)$$

In other words, there are values of N_o, N_i for which PARAFAC still yields unique decomposition, e.g., $N_o=4$, $N_i=5$ or $N_o=6$, $N_i=8$.

If, in addition to A3), assumption A2) is also satisfied, then the diagonal frequency-dependent scaling ambiguity can be reduced to a fixed diagonal scaling ambiguity. This can be achieved via the iteration defined in the following proposition. We should note that A2) cannot be satisfied by a MIMO system with more inputs than outputs.

Proposition 2: For some fixed m, δ in $[0, N-1]$, N, δ coprime, and under assumptions A1)–A4), consider the PARAFAC decomposition of tensors $\mathcal{C}^3(-m+r\delta,\delta)$, $r=0,\ldots,N-1$. Let $\hat{\mathbf{A}}'_r$, $\hat{\mathbf{B}}'_r$, $\hat{\mathbf{C}}'_r$ be the corresponding column reordered components. For

$$\mathbf{Q}(r) = \hat{\mathbf{A}}'_r\hat{\mathbf{C}}_0^{-1}\hat{\mathbf{C}}'_r\left(\mathbf{Q}^{-1}(r-1)\right)^*\hat{\mathbf{B}}'_r, \quad r=1,\ldots,N-1 \quad (19)$$

$$\mathbf{Q}(0) = \hat{\mathbf{A}}'_0 \quad (20)$$

it holds

$$\mathbf{Q}(r) = \mathbf{H}(m-r\delta-\delta)\mathbf{P}\mathbf{K}_{((r))_2}e^{j(\mathbf{\Phi}_1+r\mathbf{\Phi}_2)}, \ r=0,\ldots,N-1 \quad (21)$$

where $\mathbf{K}_0$, $\mathbf{K}_1$ are diagonal matrices with positive elements, and $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2$ are diagonal matrices.

The proof of Proposition 2 is given in Appendix II.

Equation (21) provides $\mathbf{H}(m-r\delta-\delta)$ within a fixed permutation matrix, a diagonal matrix that assumes a different fixed value depending on whether r is odd or even, and a phase diagonal ambiguity that depends on r. Since the DFT domain contains enough redundancy, using the even or odd samples of the channel matrix would suffice for recovering the system impulse response, as long as $N > 2L$.

Considering (21) for r even (or odd), and comparing to (15), we can see that there is a frequency-dependent diagonal ambiguity in both (i.e., $\mathbf{\Lambda}_{1r}$ and $e^{j(\mathbf{\Phi}_1+r\mathbf{\Phi}_2)}$, respectively) but in (21) that ambiguity has unit modulus. Although both ambiguities can be resolved using a SISO approach, as will be shown next, there is a simpler way to compensate for the unit modulus one.

We next show that the phase term $\mathbf{\Phi}_2$ can actually be computed within an integer multiple of 2π and/or an integer multiple of $2\pi/N$.

Consider $\mathbf{Q}(N+i)$ for some $i\in[0,N-1]$ as given in (21). For N, δ coprime, it holds

$$\begin{aligned}\mathbf{Q}(N+i) &= \mathbf{H}(m-(N+i)\delta-\delta)\\ &\quad\times\mathbf{P}\mathbf{K}_{((N+i))_2}e^{j(\mathbf{\Phi}_1+(N+i)\mathbf{\Phi}_2)}\\ &= \mathbf{H}(m-i\delta-\delta)\mathbf{P}\mathbf{K}_{((N+i))_2}e^{j(\mathbf{\Phi}_1+(N+i)\mathbf{\Phi}_2)} \quad (22)\\ \mathbf{Q}(i) &= \mathbf{H}(m-i\delta-\delta)\mathbf{P}\mathbf{K}_{((i))_2}e^{j(\mathbf{\Phi}_1+i\mathbf{\Phi}_2)}. \quad (23)\end{aligned}$$

Combining (22) and (23), we get

$$\mathbf{Q}^{-1}(N+i)\mathbf{Q}(i) = \mathbf{K}^{-1}_{((N+i))_2}\mathbf{K}_{((i))_2}e^{-jN\mathbf{\Phi}_2}. \quad (24)$$

Thus, $\mathbf{Q}^{-1}(N+i)\mathbf{Q}(i)$ is a diagonal matrix; if N is even, it has unit modulus; otherwise, its modulus depends on whether i is odd or even.

Let us consider N to be even, and δ to be coprime to N. Under A1)–A4), $\mathbf{H}(m-r\delta-\delta)$ can be obtained within trivial ambiguities as

$$\begin{aligned}&\hat{\mathbf{H}}(m-r\delta-\delta)\\ &\quad\triangleq \mathbf{Q}(r)\left[\mathbf{Q}^{-1}(N+i)\mathbf{Q}(i)\right]^{r/N}, \ r=0,\ldots,N-1 \quad (25)\\ &\quad= \mathbf{H}(m-r\delta-\delta)\mathbf{P}\mathbf{K}'_{((r))_2}e^{j\left(\mathbf{\Phi}_1+\frac{2\pi}{N}kr\right)} \quad (26)\end{aligned}$$

where i is some integer in $[0,N-1]$; k : integer; and $\mathbf{K}'_{((r))_2} \triangleq \mathbf{K}_{((r))_2}[\mathbf{K}^{-1}_{((N+i))_2}\mathbf{K}_{((i))_2}]^{r/N}$, i.e., it is a diagonal matrix with positive elements taking two different values depending on whether r is even or odd.

At the starting point of the iteration of Proposition 2, $\mathbf{A}'_0$ is needed. This is a function of $\mathbf{H}(m-\delta)$. Therefore, we can choose m so that we start the iteration with a well-conditioned

Take integers N, L, δ, i, m such that: N even; $N > 2L$; $i, \delta \in [0, ...N-1]$; N, δ co-prime; $m = R\delta$ (R:integer)

$\mathcal{C}^3(-m, \delta) \overset{PARAFAC}{\longrightarrow} \hat{\mathbf{A}}_0, \hat{\mathbf{B}}_0, \hat{\mathbf{C}}_0$

$\mathbf{Q}(0) = \hat{\mathbf{A}}_0$

for $r = 1, ..., N + i - 1$

$\mathcal{C}^3(-m + r\delta, \delta) \overset{PARAFAC}{\longrightarrow} \hat{\mathbf{A}}_r, \hat{\mathbf{B}}_r, \hat{\mathbf{C}}_r \overset{column\ ordering}{\longrightarrow} \hat{\mathbf{A}}'_r, \hat{\mathbf{B}}'_r, \hat{\mathbf{C}}'_r$

$\mathbf{Q}(r) = \hat{\mathbf{A}}'_r \hat{\mathbf{C}}_0^{-1} \hat{\mathbf{C}}'_r (\mathbf{Q}^{-1}(r-1))^* \hat{\mathbf{B}}'_r$

end

Form $\hat{\mathbf{H}}(-r\delta) = \mathbf{Q}(r+R)[\mathbf{Q}^{-1}(N+i)\mathbf{Q}(i)]^{(r+R)/N}, \quad r = -R, ..., N-1-R$

Find $\hat{\mathbf{h}}(n)$ as the IDFT of the even-indexed samples of $\hat{\mathbf{H}}(-((r))_N\delta)$

Downsample $\hat{\mathbf{h}}(n)$ by $(-\delta)$

At this point we have the system impulse response within a circular shift, a fixed diagonal scaling ambiguity and a fixed permutation ambiguity.

Fig. 1. MPD approach.

Take integers N, L, δ, i, m such that: N even; $N > 2L$; $i, \delta \in [0, ...N-1]$; N, δ co-prime; $m = R\delta$ (R:integer)

$\mathcal{C}^3(-m, \delta) \overset{PARAFAC}{\longrightarrow} \hat{\mathbf{A}}_0, \hat{\mathbf{B}}_0, \hat{\mathbf{C}}_0$

$\mathbf{F}(0) = \hat{\mathbf{A}}_0$

for $r = 1, ..., N + i - 1$

for $l = 1, ..., N_o$

$\mathbf{F}_l(r) = (\mathbf{F}^*(r-1))^{-1} \mathbf{C}_l^3(-m + r\delta, \delta)(\hat{\mathbf{C}}_0^T)^{-1}$

end

$\mathbf{F}(r) = [diag(\mathbf{F}_1(r)), ..., diag(\mathbf{F}_{N_o}(r))]^T$

end

Form $\hat{\mathbf{H}}(-r\delta) = \mathbf{F}(r+R)[\mathbf{F}^{-1}(N+i)\mathbf{F}(i)]^{(r+R)/N}, \quad r = -R, ..., N-1-R$

Find $\hat{\mathbf{h}}(n)$ as the IDFT of the even-indexed samples of $\hat{\mathbf{H}}(-((r))_N\delta)$

Downsample $\hat{\mathbf{h}}(n)$ by $(-\delta)$

At this point we have the system impulse response within a circular shift, a fixed diagonal scaling ambiguity and a fixed permutation ambiguity.

Fig. 2. SPD approach.

matrix. Let us take $m - \delta = R\delta$, where R is some integer. Then (25) can be written as

$$\hat{\mathbf{H}}(-r\delta) = \mathbf{Q}(r+R)\left[\mathbf{Q}^{-1}(N+i)\mathbf{Q}(i)\right]^{(r+R)/N}, \quad r = -R, \ldots, N-1-R \tag{27}$$

$$= \mathbf{H}(-\delta r)\mathbf{P}\mathbf{K}'_{((r+R))_2} e^{j\left(\mathbf{\Phi}'_1 + \frac{2\pi}{N} kr\right)} \tag{28}$$

where $\mathbf{\Phi}'_1 = (\mathbf{\Phi}_1 + (2\pi/N)kR)$.

Let us reorder the samples of $\hat{\mathbf{H}}(-r\delta)$, $r = -R, \ldots, N-1-R$, i.e., form $\hat{\mathbf{H}}(-((r))_N\delta)$, $r = 0, \ldots, N-1$ Applying an $N/2$-point inverse DFT (IDFT) on the even samples of the latter sequence, we get

$$\hat{\mathbf{h}}(n) = \mathbf{h}_{-\delta}\left((n+k)\right)_{N/2} \mathbf{P}\mathbf{K}'_{((R))_2} e^{j\mathbf{\Phi}'_1} \tag{29}$$

which is an upsampled by $-\delta$ version of $\mathbf{h}(n)$ circularly shifted by k. By downsampling $\hat{\mathbf{h}}(n)$ by $-\delta$, we can get a circularly shifted version of $\mathbf{h}(n)$.

We will refer to the above channel estimation methods as the multiple PARAFAC decomposition (MPD) approach. A summary of MPD is given in Fig. 1.

B. Channel Estimation Based on a Single PARAFAC Decomposition: SPD Approach

Proposition 3: Assume that A1–A4) hold and that $N_i \leq N_o$ with $N_i \geq 2$. For some $\delta \in [0, \ldots N-1]$, with N, δ coprime, consider the PARAFAC decomposition of tensor $\mathcal{C}^3(-m, \delta)$ resulting in components $\hat{\mathbf{A}}_0$, $\hat{\mathbf{B}}_0$, and $\hat{\mathbf{C}}_0$.

For $r = 1, 2, \ldots N-1$, define

$$\mathbf{F}_l(r) \triangleq (\mathbf{F}^*(r-1))^{-1} \mathbf{C}_l^3(-m + r\delta, \delta)\left(\hat{\mathbf{C}}_0^T\right)^{-1}, \quad l = 1, \ldots, N_o \tag{30}$$

where

$$\mathbf{F}(r) \triangleq \left[\mathrm{diag}\left(\mathbf{F}_1(r)\right), \ldots, \mathrm{diag}\left(\mathbf{F}_{N_o}(r)\right)\right]^T \tag{31}$$

$$\mathbf{F}(0) = \mathbf{A}_0 \tag{32}$$

It holds

$$\mathbf{F}(r) = \mathbf{H}(m - r\delta - \delta)\mathbf{P}\mathbf{K}_{((r))_2} e^{j(\mathbf{\Phi}_1 + r\mathbf{\Phi}_2)} \tag{33}$$

where $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2$ are diagonal matrices and $\mathbf{K}_1$, $\mathbf{K}_0$ are diagonal matrices with positive elements.

The phase ambiguity $\mathbf{\Phi}_2$ can be solved along the lines of (24), where $\mathbf{Q}(r)$ is replaced with $\mathbf{F}(r)$.

Again, taking $m - \delta = R\delta$, where R is some integer, we get

$$\hat{\mathbf{H}}(-r\delta) = \mathbf{F}(r+R)\left[\mathbf{F}^{-1}(N+i)\mathbf{F}(i)\right]^{(r+R)/N}, \quad r = -R, \ldots, N-1-R \tag{34}$$

$$= \mathbf{H}(-\delta r)\mathbf{P}\mathbf{K}'_{((r+R))_2} e^{j\left(\mathbf{\Phi}'_1 + \frac{2\pi}{N} kr\right)} \tag{35}$$

where $\mathbf{\Phi}'_1 = (\mathbf{\Phi}_1 + (2\pi/N)kR)$.

Finally, the channel impulse response matrix can be obtained as in (29).

The above proposition provides a method for estimating the system response. We will refer to that methods as the single PARAFAC decomposition (SPD) method, a summary of which is given in Fig. 2. We should note that the channel assumptions for this approach are just a subset of those in [16].

The SPD method is computationally simpler than the MPD approach.

V. Extensions to Fourth-Order Statistics

The ideas presented above can be extended to HOS. We next discuss the extension to fourth-order statistics.

Based on the assumption A4) and A5), the fourth-order cross cumulant of the received signals equals [35]

$$\begin{aligned} &c^4_{ijlm}(\tau_1,\tau_2,\tau_3) \\ &\quad\triangleq \mathrm{Cum}\left[x_i(k), x_j^*(k+\tau_1), x_l(k+\tau_2), x_m^*(k+\tau_3)\right] \\ &\quad= \sum_{p=1}^{r} \gamma^4_{s_p} \sum_{q=0}^{L-1} h_{ip}(q) h^*_{jp}(q+\tau_1) h_{l_1 p}(q+\tau_2) h^*_{l_2 p}(q+\tau_3) \end{aligned} \quad (36)$$

where $\gamma^4_{s_p} = \mathrm{Cum}[s_p(m), s_p^*(m), s_p(m), s_p^*(m)]$ is the fourth-order cumulant of $s_p(.)$. The discretized fourth-order cross spectrum, defined as the three-dimensional (3-D) DFT of $c^4_{ijlm}(\tau_1,\tau_2,\tau_3)$ equals [35]

$$\begin{aligned} C^4_{ijlm}(k_1,k_2,k_3) = &\sum_{p=1}^{r} \gamma^4_{s_p} H_{ip}(-k_1-k_2-k_3) \\ &\times H^*_{jp}(-k_1) H_{lp}(k_2) H^*_{mp}(-k_3). \end{aligned} \quad (37)$$

There are two approaches one could follow. The first is based on a trilinear decomposition of a three-way tensor formed based on fourth-order spectra given by (37), where k_1, k_2, k_3, and m are fixed. Let $\boldsymbol{\mathcal{C}}^4_m(k_1,k_2,k_3)$ denote a 3-D tensor whose (i,j,l)th element is equal to $C^4_{ijlm}(k_1,k_2,k_3)$. Its lth slice equals

$$\begin{aligned} \mathbf{C}^4_{lm}(k_1,k_2,k_3) = &\mathbf{H}(-k_1-k_2-k_3) \\ &\times \mathbf{\Gamma}^4 \mathbf{D}_l\left[\mathbf{H}(k_2)\mathbf{D}_m\left[\mathbf{H}^*(-k_3)\right]\right]\mathbf{H}^H(-k_1). \end{aligned} \quad (38)$$

where $\mathbf{\Gamma}^4 = \mathrm{Diag}\{\gamma^4_{s1},\ldots,\gamma^4_{s_{N_i}}\}$. Equation (38) is similar to (9), so the results presented in the previous section are directly applicable.

However, the above approach does not exploit all information available in fourth order statistics. Alternatively, we can use four-dimensional PARAFAC decomposition.

Consider a four-way tensor $\mathcal{X}^4$ with dimensions $I\times J\times L\times M$ and the F-component decomposition

$$x_{i,j,l,m} = \sum_{f=1}^{F} a_{i,f} b_{j,f} c_{l,f} g_{m,f} \quad (39)$$

for $i=1,\ldots,I$, $j=1,\ldots,J$, $l=1,\ldots,L$, $m=1,\ldots,M$. Let $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{G}$ be matrices with elements $a_{i,f}$, $b_{j,f}$, $c_{l,f}$, and $g_{m,f}$, respectively. Given $\mathcal{X}^4$, the matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{G}$ are unique up to permutation and complex scaling of columns provided that [41]

$$k_A + k_B + k_C + k_G \geq 2F+3. \quad (40)$$

In a compact form, $\mathcal{X}$ can be expressed in terms of its slices $\mathbf{X}_{il}(J\times M)$ as

$$\mathbf{X}_{il} = \mathbf{B}\mathbf{D}_i\left[\mathbf{A}\mathbf{D}_l[\mathbf{C}]\right]\mathbf{G}^T. \quad (41)$$

For our case, let us consider the four-way tensor $\mathcal{C}^4(k_1,k_2,k_3)$ whose (i,j,l,m)th element equals $C^4_{ijlm}(k_1,k_2,k_3)$, and define

$$\mathbf{A}_0 \triangleq \mathbf{H}(k_2) \quad (42)$$

$$\mathbf{B}_0 \triangleq \mathbf{H}(-k_1-k_2-k_3)\mathbf{\Gamma}^4 \quad (43)$$

$$\mathbf{C}_0 \triangleq \mathbf{H}^*(-k_3) \quad (44)$$

$$\mathbf{G}_0 \triangleq \mathbf{H}^*(-k_1). \quad (45)$$

The PARAFAC decomposition on $\mathcal{C}^4(k_1,k_2,k_3)$ yields

$$\begin{aligned} \hat{\mathbf{A}}_0 &= \mathbf{A}_0\mathbf{P}_{10}\mathbf{\Lambda}_{10} \\ \hat{\mathbf{B}}_0 &= \mathbf{B}_0\mathbf{P}_{20}\mathbf{\Lambda}_{20} \\ \hat{\mathbf{C}}_0 &= \mathbf{C}_0\mathbf{P}_{30}\mathbf{\Lambda}_{30} \\ \hat{\mathbf{G}}_0 &= \mathbf{G}_0\mathbf{P}_{40}\mathbf{\Lambda}_{40}. \end{aligned} \quad (46)$$

As in the third-order cumulant case, it can be shown that

$$\mathbf{P}_{10} = \mathbf{P}_{20} = \mathbf{P}_{30} = \mathbf{P}_{40} = \mathbf{P}, \quad \mathbf{\Lambda}_{10}\mathbf{\Lambda}_{20}\mathbf{\Lambda}_{30}\mathbf{\Lambda}_{40} = \mathbf{I}. \quad (47)$$

Proposition 4: Assume A2), A4), A5), $N_i \leq N_o$, and $N_o \geq 2$. For some fixed k_1, k_2, k_3 in $[0, N-1]$, with (k_1+k_2), N coprime, consider the tensor $\mathcal{C}^4(k_1,k_2,k_3)$. Let $\hat{\mathbf{A}}_0$, $\hat{\mathbf{B}}_0$, $\hat{\mathbf{C}}_0$, and $\hat{\mathbf{G}}_0$ be the components of its PARAFAC decomposition.

For $i,l\in[0,\ldots,N_0]$, define

$$\begin{aligned} \mathbf{F}^4_{il}(r) \triangleq &\left(\mathbf{F}^{4*}(r-1)\right)^{-1} \mathbf{C}^4_{il}\left(k_1,k_2,k_3-r(k_1+k_2)\right) \\ &\times \left(\hat{\mathbf{G}}_0^T\right)^{-1}, \quad r=1,2,\ldots,N-1 \end{aligned} \quad (48)$$

$$\mathbf{F}^4(0) = \hat{\mathbf{C}}_0^*. \quad (49)$$

Construct

$$\mathbf{F}^4_l(r) \triangleq \left[\mathrm{diag}\left(\mathbf{F}^4_{1l}(r)\right),\ldots,\mathrm{diag}\left(\mathbf{F}^4_{N_o l}(r)\right)\right]^T \quad (50)$$

$$\mathbf{F}^4(r) \triangleq \left[\hat{\mathbf{A}}_0^{-1}\mathrm{diag}\left(\mathbf{F}^4_1(r)\right),\ldots,\hat{\mathbf{A}}_0^{-1}\mathrm{diag}\left(\mathbf{F}^4_{N_o}(r)\right)\right]^T. \quad (51)$$

It holds

$$\mathbf{F}^{4*}(r) = \mathbf{H}\left(-k_3+r(k_1+k_2)\right)\mathbf{P}\mathbf{S}_{((r))_2} e^{j\left(-\mathbf{\Phi}^4_1+(r+1)\left(\mathbf{\Phi}^4_2\right)\right)} \quad (52)$$

where $\mathbf{\Phi}^4_1$, $\mathbf{\Phi}^4_2$ are constant diagonal matrices, and $\mathbf{S}_1$ and $\mathbf{S}_0$ are diagonal matrices with positive elements.

The proof is given in Appendix IV.

By setting $m=-k_3$ and $\delta=-(k_1+k_2)$, (52) can be rewritten as

$$\mathbf{F}^{4*}(r) = \mathbf{H}(m-r\delta)\mathbf{P}\mathbf{S}_{((r))_2} e^{j\left(-\mathbf{\Phi}^4_1+(r+1)\left(\mathbf{\Phi}^4_2\right)\right)}. \quad (53)$$

Similar to (33), (53) provides the even- or odd-indexed samples of the system frequency response within a phase and constant permutation and ambiguities. We should take m to be an

integer multiple of δ to facilitate resolution along the line of (33).

Proposition 4 provides a method to obtain the MIMO system response using fourth-order statistics. We will refer to this method as the *fourth-order SPD method*.

VI. Implementation Issues

As already discussed, N, δ should be taken to be coprime. To facilitate phase recovery, N should be taken to be even. Also, m should be taken as an integer multiple of δ.

When implementing the MPD method, to facilitate the column reordering step, we should pick δ so that it maximizes the minimum Euclidean distance between the columns of $\hat{\mathbf{C}}_0(\mathbf{D}_1[\hat{\mathbf{C}}_0])^{-1}$ (normalized $\hat{\mathbf{C}}_0$).

In both MPD and third-order-cumulants-based SPD, since we are using the inverse of $\hat{\mathbf{C}}_0$, we should pick δ so that the smallest eigenvalue of $\hat{\mathbf{C}}_0$ is as large as possible. Also, $\hat{\mathbf{A}}_0$ should be a well-conditioned matrix. Although there is no formal way to ensure the above, simulations suggest that, in the case where all cross channels occupy approximately the same frequency range, the following ad hoc method can be used. We can estimate the trace of the power spectrum matrix corresponding to the system output. Then, we can experiment with different values of m, δ selected from the high-energy region of the trace, that is, perform PARAFAC decomposition of $\mathcal{C}^3(-m, \delta)$ and check whether the resulting $\hat{\mathbf{C}}_\mathbf{0}$ and $\hat{\mathbf{A}}_\mathbf{0}$ satisfy the above conditions.

For the fourth-order-cumulants-based estimation, we set $m = -k_3$ and $\delta = -(k_1 + k_2)$. Equation (48) requires the inverse of $\hat{\mathbf{G}}_0^T = \mathbf{H}^T(-k_1)$ in each iteration, and the inverse of $\hat{\mathbf{C}}_0^T = (\mathbf{H}^*(-k_3))^T$ at the first step. In order to enable phase recovery, m should be an integer multiple of δ. Thus, we can always set $k_2 = 0$ and choose $\delta = -k_1$ and $m = -k_3 = R\delta$, where $-k1$ and m lie in the high-energy area of the power spectrum trace.

VII. Potential Applications

MIMO problems that can be solved using third-order statistics arise in polarimetric calibration of radar images, where inputs corresponding to several transmit–receive pairs of polarizations are observed through FIR distortion filters [18], [38]. Synthetic aperture imaging (STA) is a technique that has shown promise in medical imaging applications [19]. In STA, an image is created by making a number of consecutive defocused transmissions from different locations of the aperture, and an image is formed from every single element transmission. Again here, image recovery can be formulated as a MIMO problem [22]. Separation of two or more speakers based on the output of multiple microphones can also be posed as a MIMO problem. In this case, the channels represent acoustic impulse responses. Site response analysis based on free surface recordings can be viewed as a MIMO problem. The channels represent path reflectivities and the inputs are seismic signals at some depth [49].

When applying the proposed third-order-statistics-based methods to the aforementioned cases, a point of concern is the validity of assumption A2), i.e., the condition number of matrix $\mathbf{H}(k)$ for each k. In cases where the channel coefficients represent attenuation of various paths, e.g., speaker separation, site response analysis, it is reasonable to assume that A2) will hold.

Due to complexity considerations, one would employ fourth-order statistics for MIMO system estimation only in cases where third-order statistics are not applicable. Multiuser multipath communications is one such case [31], [36]. Since most communications signals are symmetrically distributed, their third-order cumulants are identically zero. Again, the elements of channel matrix represent attenuation of various paths and can be assumed independent [40] thus satisfying assumption A2).

VIII. Simulations

In this section, we demonstrate the performance of the proposed approaches.

In all cases, the additive noise processes were white, zero-mean, complex Gaussian with identical variances and they were independent of the source signals.

The sample cross-cumulant estimates were windowed by a Hamming window of size $L_e \times L_e$ for third-order cumulants, or $L_e \times L_e \times L_e$ for fourth-order cumulants, where L_e is an upper bound for the channel length $(L_e > L)$. The data length used to obtain the cross-cumulant estimates is denoted by T. The channel impulse response was obtained as the L_e-samples long segment (modulo $N/2$) with the maximum energy.

In most practical applications, the channels are bandpass signals. For this reason we conducted our simulations using channels generated as

$$h_{ij}(n) = r_1 \mathrm{sinc}\,(0.25(n-2)) + r_2 \mathrm{sinc}\,(0.25(n-1)) \quad (54)$$

where the r_i's are zero-mean Gaussian random variables. By varying the r_i's, we can generate multiple bandpass channels and the performance of the proposed methods can be looked at as the average of the performance of all different channels.

We should note that bandpass channels represent challenging cases for the proposed methods, since for certain frequencies, $\mathbf{H}(k)$ can have a large condition number.

For each channel estimation, we performed M_c Monte Carlo simulations. The normalized mean-square error (NMSE) was used as a performance index, i.e.,

$$\mathrm{NMSE}_{ij} \triangleq = \frac{\frac{1}{M_c}\sum_{l=1}^{M_c}\sum_{k=0}^{L_e-1}\left(\hat{h}_{ij}(k) - h_{ij}(k)\right)^2}{\sum_{k=0}^{L_e-1}\left(h_{ij}(k)\right)^2} \quad (55)$$

where $\hat{h}_{ij}(k)$ denotes the cross-channel estimate. The latter was extracted as the L_e-sample long segment (its lags taken modulo $N/2$) that corresponded to the maximum correlation with the true $h_{ij}(k)$.

The overall NMSE (ONMSE) was then obtained by averaging over all subchannels, i.e.,

$$\mathrm{ONMSE} \triangleq \frac{1}{N_i N_o}\sum_{i=1}^{N_o}\sum_{j=1}^{N_i}\mathrm{NMSE}_{ij}. \quad (56)$$

The PARAFAC decomposition was performed using the MATLAB code downloaded from http://www.ece.umn.edu/users/nikos/public_html/3SPICE/code.html.

A. Estimation Using Third-Order Statistics

Here, the inputs were taken to be i.i.d. single-sided and exponentially distributed. The cross-third-order spectrum was estimated via the indirect class method [35]. The MPD and SPD methods were applied as described in Figs. 1 and 2. To avoid getting a reversed version of $\mathbf{h}$ we used negative δ.

When applying the SPD method, instead of (30), the following equation was implemented:

$$\mathbf{F}_l(r) \triangleq \left(\left(\hat{\mathbf{C}}_0^T\right)\left(\mathbf{C}_l^3(-m+r\delta,\delta)\right)^{-1}\mathbf{F}^*(r-1)\right)^{-1}, \quad l=1,\ldots,N_o. \quad (57)$$

Although theoretically the same as (30), (57) resulted in better performance since, for the case where $N_o = N_i$ the bispectrum estimate $\mathbf{C}_l^3(-m+r\delta,\delta)$ contains less errors than the estimate of $\mathbf{F}^*(r-1)$. If $N_o > N_i$, $\mathbf{C}_l^3(-m+r\delta,\delta)$ is no longer invertible and one would still use (30).

Decoupling of Inputs Followed by SISO Equalization: We mentioned in Proposition (1) that the channel estimate $\hat{\mathbf{A}}'_r$ of (10) can be used to decouple the input signals, leaving a filtering ambiguity in each input. Then, we can apply a SISO equalizer to solve for that ambiguity. This was implemented as follows. Based on the system estimate we applied a zero-forcing equalizer to the system outputs. We then used the SISO simplification of the MIMO approach of [46] in order to cancel the filtering ambiguity in each recovered input. We chose the method of [46] mainly because it is one of the few time domain methods that are not sensitive to the channel length overestimation.

The channel $\hat{\mathbf{h}}(n-m)$ was subsequently estimated by cross-correlating the recovered input vector $\hat{\mathbf{s}}(n-m)$ with the output vector $\mathbf{x}(n)$ for various values of m.

1) Example 1—Selection of Parameters: In the following we discuss parameter selection and show the effect of various parameters on estimation performance.

We considered a 2×2 MIMO channel of length $L = 6$ produced based on (54). The channel taps were

$$\begin{aligned}
h_{11}(1\ldots L-1) &= [0.0827,\ -0.0830,\ -0.2204,\ -0.1641,\\
&\qquad 0.2290,\ 1.0000]^T\\
h_{12}(1\ldots L-1) &= [0.3571,\ 0.3281,\ 0.0596,\\
&\qquad -0.3757,\ -0.7997,\ -1.0000]^T\\
h_{21}(1\ldots L-1) &= [-0.2299,\ -0.3280,\ -0.2302,\ 0.0888,\\
&\qquad 0.5517,\ 1.0000]^T\\
h_{22}(1\ldots L-1) &= [-0.1625,\ -0.1669,\ -0.0384,\ 0.2343,\\
&\qquad 0.6090,\ 1.0000]^T. \quad (58)
\end{aligned}$$

We used data length $T = 8000$, signal-to-noise ratio SNR $= 20$ dB, $N = 128$, $L_e = 10$. We performed PARAFAC decomposition of $\mathcal{C}^3(-m,\delta)$ for $m = 0$ and all the allowable values of δ's, i.e., $\delta = -N+1, -N+3, \ldots, -3, -1$ and computed the minimum Euclidean distance between the columns of $\hat{\mathbf{C}}_0(\mathbf{D}_1[\hat{\mathbf{C}}_0])^{-1}$. The average of that distance corresponding to five independent input realizations is shown in Fig. 3. The true minimum distance, which was calculated based on $\mathbf{H}(\delta)(\mathbf{D}_1[\mathbf{H}(\delta)])^{-1}$, is also shown on the same figure. Fig. 3(b) shows the power spectrum trace of the system output, and also the true power spectrum trace, i.e., $trace\{\mathbf{H}(k)\mathbf{H}(k)^H\}$. Comparing Fig. 3(a) and (b), we see that the estimate of the distance corresponding to a δ taken in the low power spectrum trace region is not always accurate. Therefore, it is preferable to always choose δ from the high power spectrum trace region. Based on Fig. (3), one can see that the best value for δ is -11.

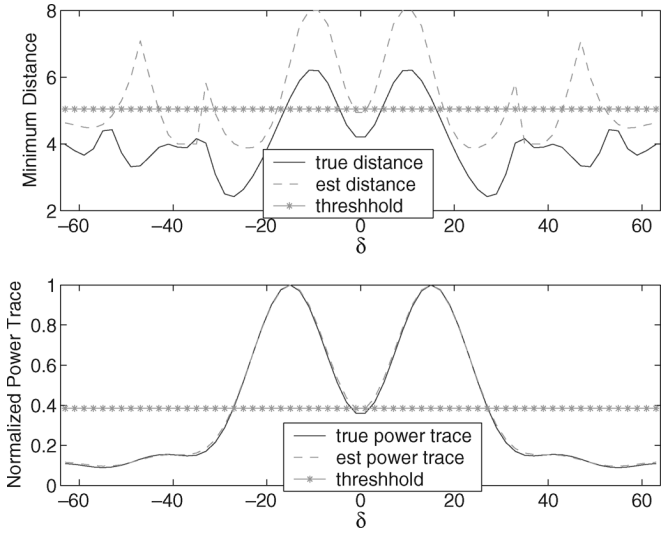

Fig. 3. Minimum distance between the columns of the normalized $\hat{\mathbf{C}}_0$, and the power spectrum trace. Both are used to select the parameter δ (Example 1).

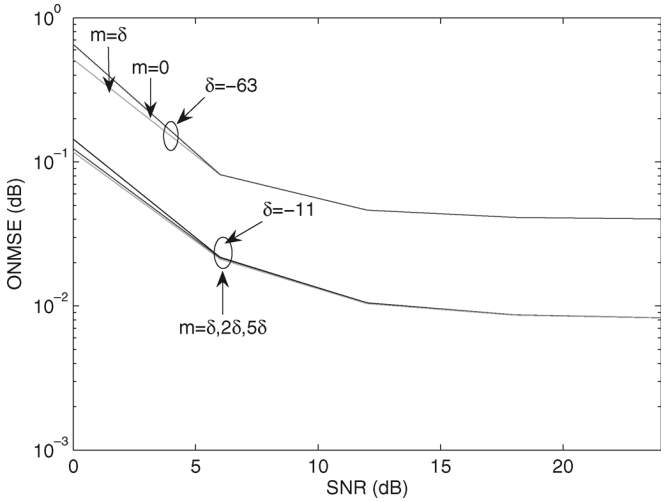

Fig. 4. ONMSE performance of the SPD approach for different values of m, δ (Example 1).

To illustrate the effect of choosing different δ's and m's, we show in Fig. 4 the ONMSE for the SPD method corresponding to different combinations of m, δ, where m was always taken to be an integer multiple of δ. It can be seen that the selection

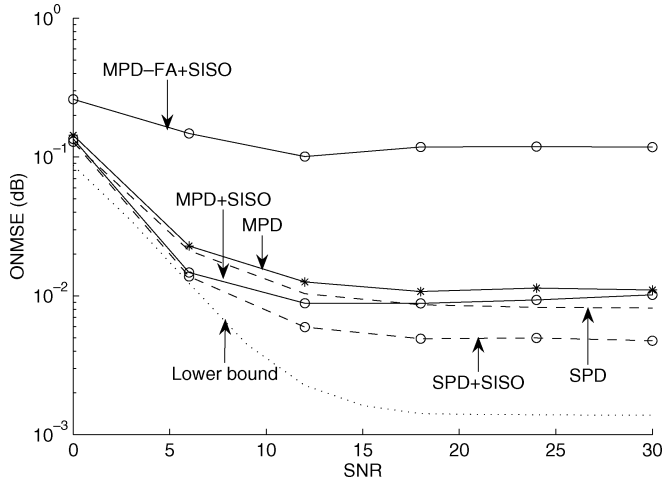

Fig. 5. ONMSE performance comparison of the proposed methods (Example 1).

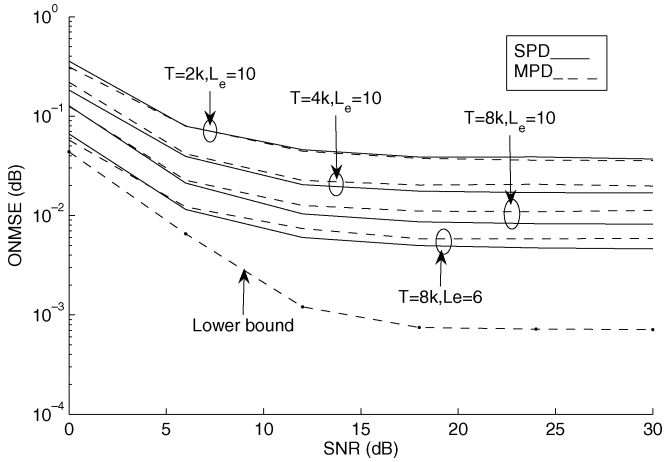

Fig. 6. ONMSE performance of the proposed methods with different Le and N (Example 1).

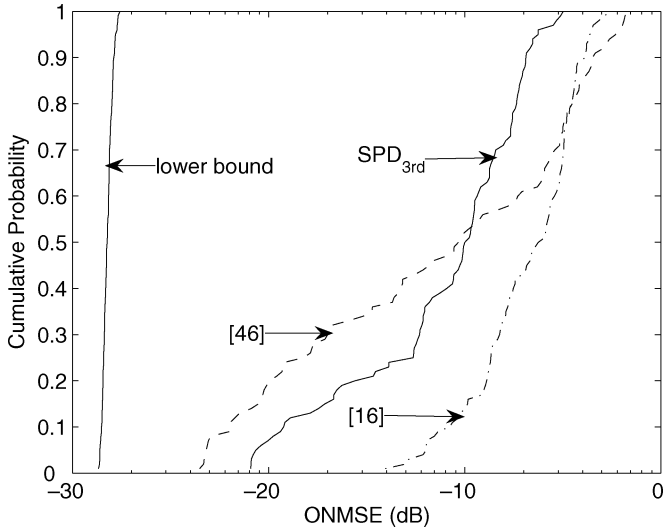

Fig. 7. Cumulative distribution of ONMSEs for SPD and the methods of [16] and [46] (Example 2).

TABLE I
ONMSE COMPARISON

ONMSE	SPD	SPD + SISO	[16]	[46]	[16] + SISO
mean	**0.1104**	**0.1034**	0.2417	0.1938	0.1564
std	**0.0770**	**0.0483**	0.1220	0.2045	0.1169

of δ is more critical than that of m. The behavior of the MPD approach with m, δ was very similar to that of Fig. 4.

In Fig. 5, we show the ONMSE performance based on 50 Monte Carlo runs for the MPD and SPD methods. We used $L_e = 10$, $m = \delta = -11$, $T = 8000$, and $N = 128$ for all the cases. As a reference point, we also show the error corresponding to the channel estimate for the case of known input. The channel estimate, referred to as the *ideal channel estimate*, was found by cross-correlating the system output vector with the known input vector. The corresponding ONMSE was taken here as the lower bound. One can see that the SPD method yields lower error as compared with the MPD one and is of course less computationally intensive.

The estimation and cancellation of $\mathbf{\Phi}_2$ (see (24)) is sometimes a sensitive step that leaves some diagonal filtering ambiguity in the channel matrix. This can be mitigated by using a SISO method. In Fig. 5, we also show the results for the MPD and SPD method followed by the SISO equalizer of [46], where one can see the improvement in performance. The equalizer length was set to 15 taps in all cases.

In Fig. 6, we show the ONMSE as a function of SNR for both MPD and SPD methods for different values of T and L_e. It can be concluded from the figure that the estimation improves by increasing the data length T. As for the extended channel length, we can see that although the error is smaller when $L_e = L = 6$, the difference is not significant, indicating that the proposed methods do not depend critically on channel length information.

2) Example 2—Performance Over a Large Number of Channels and Comparison With Existing Methods: To render the comparisons independent of the channel, we tested performance based on 100 2×2 bandpass channels of length $L = 6$, simulated according to (54), with the r_i's taken to be independent Gaussian random variables with zero mean and unit variance. Each time domain subchannel was normalized to have a maximum absolute channel tap equal to 1. For each channel we performed 50 Monte Carlo runs. For all the methods, we took $T = 8,000$, $L_e = 10$. For the proposed method, we used $m = \delta = -11$, $N = 128$.

We compared the performance of the SPD approach against that of the frequency-domain approach of [16] and also the time-domain method of [46]. According to [16], a closed-form solution for the system frequency response is obtained based on joint diagonalization of matrices constructed based on slices of higher order polyspectra of the system output. Prewhitening is used to make the system matrix unitary. The method of [46] is a deflation-type approach, where the input sequences are extracted and removed one by one. At the end, the system is estimated based on the system output and the estimated input. We choose this method mainly because it is one of the few time domain methods that are not sensitive to the channel length Le.

In Fig. 7, we show the cumulative probability function of the ONMSE for (I) the proposed SPD method; (II) the method of

TABLE II
COMPLEXITY COMPARISON

	SPD	SPD + SISO	[16]	[46]	[16]+SISO
Average running time (sec)	**2.9511**	**7.3053**	7.7766	7.3910	11.9622

[16]; (III) the method of [46]; and (IV) the ideal channel estimation. The graph shows that for the SPD method the probability that during the 100 runs the ONMSE will remain below -10 db is much higher than for the other two comparison methods. On the other hand, the method of [46] exhibits a lower error floor.

For the SPD method, the errors are caused because at some frequencies, k, due to the bandlimited nature of the channel, matrix $\mathbf{H}(\mathbf{k})$ has large condition number.

The average ONMSE during the 100 runs and its standard deviation are shown in Table I.

Regarding complexity, it is not simple to determine the complexity of COMFAC based PARAFAC decomposition. For an $I \times J \times K$ tensor and an F component decomposition the complexity of each iteration is $\mathcal{O}(IJKF)$. However, the number of iterations depends on the data to be decomposed [8], [40]. No analysis results are available except for very simple cases. To give an idea of the complexity involved, we show in Table II the average running time of each method for 100 Monte Carlo runs for one channel realization as described in (54). We can see that the running time of the SPD method is lower than that of the joint diagonalization method of [16] and the time domain method of [46]. In general, the complexity of MPD is one order of magnitude higher that that of the SPD.

We should note that for the proposed approach and also for the method of [16], 2 s of the total running time were taken up by the bispectrum estimation step.

B. Estimation Using Fourth-Order Statistics

The inputs here were taken to be BPSK with unit power, and the independent between the users.

The code for fourth-order cumulants estimation was downloaded from http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=3013 [43].

1) Example 3—Selection of Parameters: We considered a 2 $\times$ 2 MIMO bandlimited channel of length 4 generated based on (54). The channel taps were

$$\begin{aligned} h_{11}(1\ldots L-1) &= [0.4360,\ 0.7583,\ 0.9713,\ 1.0000]^T \\ h_{12}(1\ldots L-1) &= [0.5080,\ 0.8186,\ 1.0000,\ 0.9882]^T \\ h_{21}(1\ldots L-1) &= [-0.2861,\ -0.6240,\ -0.8934,\ -1.0000]^T \\ h_{22}(1\ldots L-1) &= [0.3121,\ 0.6472,\ 0.9069,\ 1.0000]^T. \end{aligned} \quad (59)$$

Fig. 8 illustrates the effect of parameters k_1, k_2 and k_3. Note that in the figure, $\delta = -(k_1 + k_2)$. We set $k_2 = 0$ and took both $-k_1 = -k_3 = \delta$ to lie in the high energy area of the power spectrum of the output. Based on Fig. 8 the choice $k1 = -11$, $k2 = 0$ and $k3 = -11$ results in the best performance.

Fig. 9 shows ONMSE performance of the SPD method for different values of T and L_e as a function of SNR. As expected,

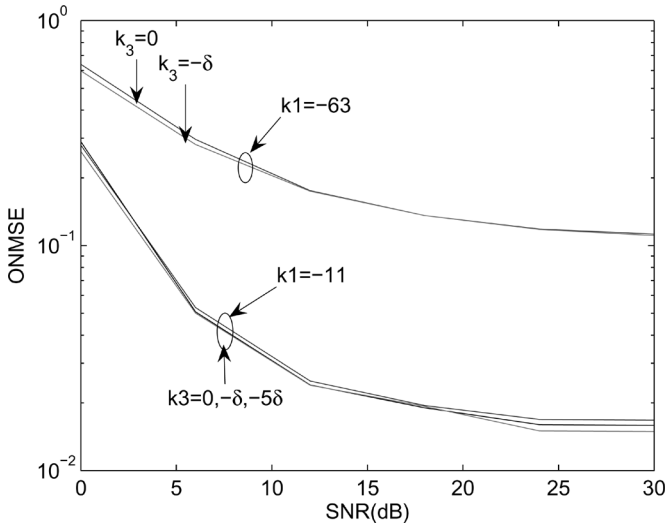

Fig. 8. ONMSE performance for different values of k_1, k_3 of the fourth-order SDP method (Example 3).

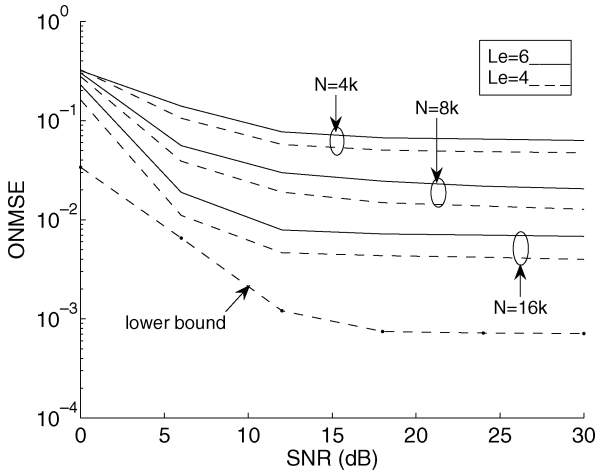

Fig. 9. ONMSE performance for different values of T and Le of the fourth-order SDP method (Example 3).

the estimation improves as the data length increases. As for the extended channel length, we can see that, although the error is smaller when $L_e = L = 4$, the difference is not significant, indicating that the proposed method does not depend critically on channel length information.

2) Example 4—Performance Based on Many Channels and Comparison With Existing Methods: To render the comparisons independent of the channel we tested performance based on 50 2 $\times$ 2 bandpass channels of length $L = 4$, simulated according to (54), with the r_i's taken to be independent Gaussian random variables with zero mean and unit variance. Each time domain subchannel was normalized to have a maximum absolute channel tap equal to 1. For each channel, we performed 30

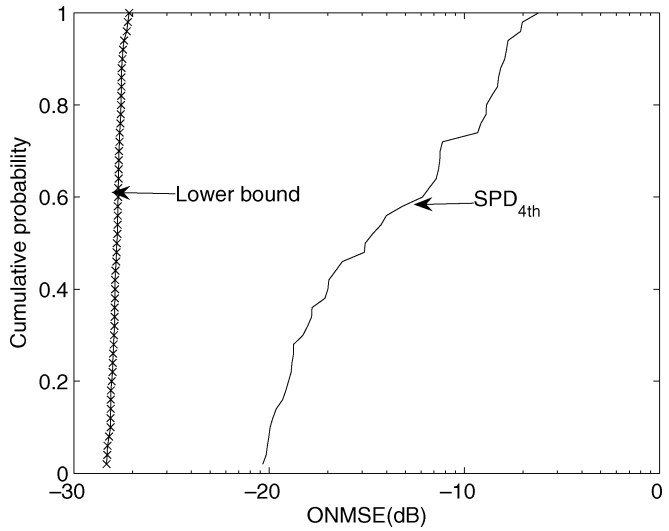

Fig. 10. Cumulative distribution of ONMSEs for the fourth-order SPD method (Example 4).

Monte Carlo runs. We used $L_e = 6, T = 8,000$, and the output cross cumulants were estimated using the same parameters (cumulant window and segment sizes).

Fig. 10 shows the cumulative ONMSE of the fourth-order SPD and also the lower bound.

The performance of the fourth-order SPD approach was also compared to that of [31]. In [31], the MIMO system is estimated from the common nullspace of a set of fourth-order cumulant matrices of the system output. As a subspace approach, the method of [31] requires exact channel length knowledge. We repeated Example 1 of [31] for quadrature phase-shift keying (QPSK) inputs. The performance was comparable to that in [31], even when the length was overestimated by 1.

IX. Conclusion

We presented a new frequency-domain framework for the identification of a MIMO system driven by white, mutually independent unobservable inputs.

The MPD approach requires multiple PARAFAC decompositions, while SPD requires a single decomposition. In the first step, all methods produce an estimate of the system frequency response matrix within a constant permutation and a diagonal frequency dependent scaling ambiguity. The latter ambiguity is reduced to a diagonal linear phase ambiguity after an iterative scheme.

The MPD involves much higher complexity than the SPD. Results showed that the ONMSE performance of the two methods is very similar, with the SPD method resulting is lower ONMSE as the data length increases. The basis of the multiple PARAFAC decompositions method can be used to estimate a MIMO system with more inputs than outputs within a filtering ambiguity. However, the iteration of the MPD method does not apply in this case to resolve the ambiguity.

The SPD approach compares favorably to existing methods, and can further improve by via use of a SISO method applied on each recovered input.

Appendix I
Proof of Proposition 1

Since both $(\mathbf{A}_r, \mathbf{B}_r, \mathbf{C}_r)$ and $(\hat{\mathbf{A}}_r, \hat{\mathbf{B}}_r, \hat{\mathbf{C}}_r)$ satisfy (9) for $k_1 = -m + r\delta$, $k_2 = \delta$, it holds

$$\mathbf{C}_l^3(-m + r\delta, \delta) = \mathbf{B}_r^* \mathbf{D}_l[\mathbf{A}_r]\mathbf{C}_r^T = \hat{\mathbf{B}}_r^* \mathbf{D}_l[\hat{\mathbf{A}}_r]\hat{\mathbf{C}}_r^T. \quad (60)$$

Substituting (10)–(12) into the above equation, and after some cancellations, we get

$$\begin{aligned}
&\mathbf{D}_l\left[\mathbf{H}(m - r\delta - \delta)\right] \\
&\quad = \mathbf{P}_{2r}\mathbf{\Lambda}_{2r}\mathbf{D}_l\left[\mathbf{H}(m - r\delta - \delta)\mathbf{P}_{1r}\mathbf{\Lambda}_{1r}\right]\mathbf{\Lambda}_{3r}\mathbf{P}_{3r}^T && (61)\\
&\quad = \mathbf{P}_{2r}\mathbf{\Lambda}_{2r}\mathbf{P}_{1r}^T\mathbf{D}_l\left[\mathbf{H}(m - r\delta - \delta)\right]\mathbf{P}_{1r}\mathbf{\Lambda}_{1r}\mathbf{\Lambda}_{3r}\mathbf{P}_{3r}^T && (62)\\
&\quad = \mathbf{P}_{2r}\mathbf{P}_{1r}^T\mathbf{D}_l\left[\mathbf{H}(m - r\delta - \delta)\right]\mathbf{P}_{1r}\mathbf{\Lambda}_{1r}\mathbf{\Lambda}_{2r}\mathbf{\Lambda}_{3r}\mathbf{P}_{3r}^T. && (63)
\end{aligned}$$

where it was taken into account that $\mathbf{D}_l[\mathbf{H}(.)\mathbf{P}_{1r}\mathbf{\Lambda}_{1r}] = \mathbf{P}_{1r}^T\mathbf{D}_l[\mathbf{H}(.)]\mathbf{P}_{1r}\mathbf{\Lambda}_{1r}$.

Note that $\mathbf{P}_{1r}^T\mathbf{D}_l[\mathbf{H}(.)]\mathbf{P}_{1r}$ is a diagonal matrix, whose diagonal elements are those of $\mathbf{D}_l[\mathbf{H}(.)]$ in an order permuted according to $\mathbf{P}_{1r}$. For the right-hand side of (63) to be diagonal, it must hold that $\mathbf{P}_{2r} = \mathbf{P}_{3r}$. Furthermore, for it to be equal to the left-hand side of (63) for all l's, it must hold $\mathbf{\Lambda}_{2r}\mathbf{\Lambda}_{1r}\mathbf{\Lambda}_{3r} = \mathbf{I}$ and $\mathbf{P}_{2r} = \mathbf{P}_{3r} = \mathbf{P}_{1r}$.

Appendix II
Proof of Proposition 2

It holds

$$\begin{aligned}
\mathbf{Q}(1) &\triangleq \hat{\mathbf{A}}_1'\hat{\mathbf{C}}_0^{-1}\hat{\mathbf{C}}_1'\left(\hat{\mathbf{A}}_0^{-1}\right)^*\hat{\mathbf{B}}_1' \\
&= \mathbf{H}(m - 2\delta)\mathbf{P}\mathbf{\Lambda}_{11}\mathbf{\Lambda}_{30}^{-1}\mathbf{\Lambda}_{31}\left(\mathbf{\Lambda}_{10}^{-1}\right)^*\mathbf{P}^T\mathbf{\Gamma}^3\mathbf{P}\mathbf{\Lambda}_{21} \\
&= \mathbf{H}(m - 2\delta)\mathbf{P}\mathbf{\Lambda}_{11}\mathbf{\Lambda}_{30}^{-1}\mathbf{\Lambda}_{31}\left(\mathbf{\Lambda}_{10}^{-1}\right)^*\mathbf{\Gamma}_p^3\mathbf{\Lambda}_{21} \\
&= \mathbf{H}(m - 2\delta)\mathbf{P}|\mathbf{\Lambda}_{20}|\left|\mathbf{\Gamma}_p^3\right|e^{j\left(\mathbf{\Phi}_{10} + \mathbf{\Phi}_{\Gamma_p^3} - \mathbf{\Phi}_{30}\right)}. \quad (64)
\end{aligned}$$

Based on the PARAFAC decomposition of $\mathcal{C}^3(-m + r\delta, \delta)$, $r = 1, 2, \ldots, N-1$, and by recalling

$$\mathbf{Q}(0) = \hat{\mathbf{A}}_0' = \mathbf{H}(m - \delta)\mathbf{P}|\mathbf{\Lambda}_{10}|e^{j(\mathbf{\Phi}_{10})} \quad (65)$$

it holds

$$\begin{aligned}
\mathbf{Q}(r) &= \hat{\mathbf{A}}_r'\hat{\mathbf{C}}_0^{-1}\hat{\mathbf{C}}_r'\left(\mathbf{Q}^{-1}(r-1)\right)^*\hat{\mathbf{B}}_r', \quad r = 0, \ldots, N-1 \\
&= \mathbf{H}(m - r\delta - \delta)\mathbf{P}\mathbf{K}_r e^{j\left(\mathbf{\Phi}_{10} + r\left(\mathbf{\Phi}_{\Gamma_p^3} - \mathbf{\Phi}_{30}\right)\right)} && (66)\\
&= \mathbf{H}(m - r\delta - \delta)\mathbf{P}\mathbf{K}_r e^{j(\mathbf{\Phi}_1 + r\mathbf{\Phi}_2)} && (67)
\end{aligned}$$

where

$$\mathbf{K}_r = \begin{cases} |\mathbf{\Lambda}_{20}|.\left|\mathbf{\Gamma}_p^3\right|, & \text{for } r \text{ odd} \\ |\mathbf{\Lambda}_{10}|, & \text{for } r \text{ even} \end{cases}. \quad (68)$$

APPENDIX III
PROOF OF PROPOSITION 3

$$\begin{aligned}\mathbf{F}_l(1) &\triangleq \left(\hat{\mathbf{A}}_0^*\right)^{-1}\mathbf{C}_l^3(-m+\delta,\delta)\left(\hat{\mathbf{C}}_0^T\right)^{-1}\\ &= \left(\mathbf{\Lambda}_{10}^*\right)^{-1}\mathbf{P}^T\mathbf{\Gamma}^3\mathbf{D}_l\left[\mathbf{H}(m-2\delta)\right]\mathbf{P}\mathbf{\Lambda}_{30}^{-1}\\ &= \left(\mathbf{\Lambda}_{10}^*\right)^{-1}\mathbf{D}_l\left[\mathbf{H}(m-2\delta)\mathbf{P}\right]\mathbf{\Gamma}_p^3\mathbf{\Lambda}_{30}^{-1}\\ &= \mathbf{D}_l\left[\mathbf{H}(m-2\delta)\mathbf{P}\right]|\mathbf{\Lambda}_{20}|\mathbf{\Gamma}_p^3 e^{j(\mathbf{\Phi}_{10}-\mathbf{\Phi}_{30})}. \end{aligned} \quad (69)$$

It can be seen that $\mathbf{F}_l(1)$ is a diagonal matrix.

Based on $\mathbf{C}_l^3(-m+\delta,\delta)$, for $l=1,\ldots,N_o$, and placing the diagonal elements of $\mathbf{F}_l(1)$ at the lth row of $\mathbf{F}(1)$, we get

$$\mathbf{F}(1) = \mathbf{H}(m-2\delta)\mathbf{P}|\mathbf{\Lambda}_{20}|\mathbf{\Gamma}_p^3 e^{j(2\mathbf{\Phi}_{10}+\mathbf{\Phi}_{20})}. \quad (70)$$

Similarly, based on $\mathbf{C}_l^3(-m+r\delta,\delta)$, we can compute $\mathbf{F}_l(r)$ along the lines of (69) and then construct $\mathbf{F}(r)$ by placing as its lth row the diagonal elements of $\mathbf{F}_l(r)$. It holds

$$\mathbf{F}(r) = \mathbf{H}(m-r\delta-\delta)\mathbf{P}\mathbf{K}_r e^{j\left((r+1)\mathbf{\Phi}_{10}+r\left(\mathbf{\Phi}_{\Gamma_p^3}+\mathbf{\Phi}_{20}\right)\right)} \quad (71)$$

which can be written as

$$\mathbf{F}(r) = \mathbf{H}(m-r\delta-\delta)\mathbf{P}\mathbf{K}_r e^{j(\mathbf{\Phi}_1+r\mathbf{\Phi}_2)} \quad (72)$$

where

$$\mathbf{K}_r = \begin{cases} |\mathbf{\Lambda}_{20}|.\left|\mathbf{\Gamma}_p^3\right|, & \text{for } r \text{ odd} \\ |\mathbf{\Lambda}_{10}|, & \text{for } r \text{ even} \end{cases}. \quad (73)$$

APPENDIX IV
PROOF OF PROPOSITION 4

Due to assumption A2) and also $N_i < N_o$ and $N_o \geq 2$, Assumption A6) is also valid. This allows the PARAFAC decomposition of tensor $\mathcal{C}^4(k_1,k_2,k_3)$.

$$\begin{aligned}\mathbf{F}_{il}^4(1) &\triangleq \left(\hat{\mathbf{C}}_0^*\right)^{-1}\mathbf{C}_{il}^4(k_1,k_2,k_3-k_1-k_2)\left(\hat{\mathbf{G}}_0^T\right)^{-1}\\ &= \left(\mathbf{\Lambda}_{30}^*\right)^{-1}\mathbf{P}^T\mathbf{\Gamma}^4\\ &\quad\times\mathbf{D}_i\left[\mathbf{H}(k_2)\mathbf{D}_l\left[\mathbf{H}^*(-k_3+k_1+k_2)\right]\right]\mathbf{P}\mathbf{\Lambda}_{40}^{-1}\\ &= \mathbf{D}_i\left[\mathbf{H}(k_2)\mathbf{D}_l\left[\mathbf{H}^*(-k_3+k_1+k_2)\right]\mathbf{P}\right]\\ &\quad\times\left(\mathbf{\Lambda}_{30}^*\right)^{-1}\mathbf{\Lambda}_{40}^{-1}\mathbf{\Gamma}_p^4. \end{aligned} \quad (74)$$

By varying i of $\mathbf{C}_{il}^4(k_1,k_2,k_3-k_1-k_2)$ from 1 to N_o, we get $\mathbf{F}_l^4(1)$ as

$$\mathbf{F}_l^4(1) = \mathbf{H}(k_2)\mathbf{D}_l\left[\mathbf{H}^*(-k_3+k_1+k_2)\right]\mathbf{P}\left(\mathbf{\Lambda}_{30}^*\right)^{-1}\mathbf{\Lambda}_{40}^{-1}\mathbf{\Gamma}_p^4. \quad (75)$$

After multiplying the above equation with $\hat{\mathbf{A}}_0^{-1}$ and varying l of $\mathbf{C}_{il}^4(k_1,k_2,k_3-k_1-k_2)$ from 1 to N_o for all i, we get

$$\mathbf{F}^4(1) = \mathbf{H}^*(-k_3+k_1+k_2)\mathbf{P}\mathbf{\Lambda}_{20}^{-1}\left(\mathbf{\Lambda}_{30}^*\right)^{-1}\mathbf{\Lambda}_{40}^{-1}\mathbf{\Gamma}_p^4. \quad (76)$$

If we take the complex conjugate of $\mathbf{F}^4(1)$ and using equation (47), we get

$$\mathbf{F}^{4*}(1) = \mathbf{H}(-k_3+k_1+k_2)\mathbf{P}|\mathbf{\Lambda}_{10}|\left|\mathbf{\Gamma}_p^4\right|e^{j\left(-\mathbf{\Phi}_1^4+\mathbf{\Phi}_2^4\right)}. \quad (77)$$

Similarly, based on $\mathbf{C}_{il}^4(k_1,k_2,k_3-r(k_1+k_2))$, we can compute $\mathbf{F}_{il}^4(r)$ for $r>1$ as

$$\begin{aligned}\mathbf{F}_{il}^4(r) &\triangleq \left(\mathbf{F}^{4*}(r-1)\right)^{-1}\\ &\quad\times\mathbf{C}_{il}^4\left(k_1,k_2,k_3-r(k_1+k_2)\right)\left(\hat{\mathbf{G}}_0^T\right)^{-1}\\ &\quad\mathbf{C}_{il}^4\mathbf{C}_{k_1},\ldots. \end{aligned} \quad (78)$$

It can be shown that

$$\mathbf{F}^{4*}(r) = \mathbf{H}\left(-k_3+r(k_1+k_2)\right)\mathbf{P}\mathbf{S}_r^4 e^{j\left(-\mathbf{\Phi}_1^4+r\mathbf{\Phi}_2^4\right)} \quad (79)$$

where

$$\mathbf{S}_r^4 = \begin{cases} |\mathbf{\Lambda}_{10}|.\left|\mathbf{\Gamma}_p^4\right|, & \text{for } r \text{ odd} \\ |\mathbf{\Lambda}_{30}|, & \text{for } r \text{ even} \end{cases}. \quad (80)$$

ACKNOWLEDGMENT

The authors would like to thank N. Sidiropoulos for helpful discussions.

REFERENCES

[1] K. Abed-Meraim, P. Loubaton, and E. Moulines, "Prediction error for second-order blind identification: Algorithm and statistical performance," *IEEE Trans. Signal Process.*, vol. 47, no. 3, pp. 694–705, Mar. 1997.

[2] ——, "A subspace algorithm for certain blind identification problems," *IEEE Trans. Inf. Theory*, vol. 43, pp. 499–511, Mar. 1997.

[3] A. Belouchrani, K. A. Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[4] J. M. F. ten Berge and N. D. Sidiropoulos, "On uniqueness in CANDECOMP/PARAFAC," *Psychometrika*, vol. 67, 2002.

[5] I. Bradaric, A. P. Petropulu, and K. I. Diamantaras, "Blind MIMO FIR channel identification based on second-order spectra correlations," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1668–1674, Jun. 2003.

[6] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Independent component analysis and (simultaneous) third-order tensor diagonalization," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2262–2271, Oct. 2001.

[7] B. Chen, A. Petropulu, and L. De Lathauwer, "Blind identification of convolutive MIMO systems with 3 sources and 2 sensors," *EURASIP J. Appl. Signal Process. (Special Issue on Space–Time Coding and Its Applications)*, vol. 2002, no. 5, pp. 487–496, May 2002.

[8] R. Bro, N. D. Sidiropoulos, and G. B. Giannakis, "A fast least squares algorithm for separating trilinear mixtures," in *Proc. Int. Workshop Independent Component Analysis Blind Signal Separation (ICA)*, Aussois, France, Jan. 11–15, 1999, pp. 289–294.

[9] V. Capdevielle, C. Serviere, and J. Lacoume, "Blind separation of wideband sources in the frequency domain," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Detroit, MI, May 9–12, 1995, vol. 3, pp. 2080–2083.

[10] J. D. Carrol and J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.

[11] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *Proc. Inst. Elect. Eng.—F*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[12] M. Castella, J. C. Pesquet, and A. P. Petropulu, "Family of frequency and time-domain contrasts for blind separation of convolutive mixtures of temporally dependent signals," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 107–120, Jan. 2005.

[13] R. B. Cattell, "'Parallel Proportional Profiles' and other principles for determining the choice of factors by rotation," *Psychometrika*, vol. 9, pp. 267–283, 1944.

[14] P. Comon, "Independent component analysis, a new concept," *Signal Process.*, vol. 36, pp. 287–314, Apr. 1994.

[15] ——, "Contrasts for multichannel blind deconvolution," *IEEE Signal Process. Lett.*, vol. 3, pp. 209–211, Jul. 1996.

[16] B. Chen and A. P. Petropulu, "Frequency domain blind MIMO system identification based on second and higher order statistics," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1677–1688, Aug. 2001.
[17] K. I. Diamantaras, A. P. Petropulu, and B. Chen, "Blind two-input-two-output FIR channel identification based on frequency domain second-order statistics," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 534–542, Feb. 2000.
[18] E. Ertin and L. C. Potter, "Polarimetric calibration for wide band synthetic aperture radar imaging," *Proc. Inst. Elect. Eng.*, vol. 145, no. 5, pp. 275–280, Oct. 1998.
[19] C. H. Frazier and W. D. O'Brien, "Synthetic aperture techniques with virtual source element," *IEEE Trans. Ultrason., Ferroelectr. Freq. Control*, vol. 45, no. 1, pp. 196–207, Jan. 1998.
[20] G. B. Giannakis and J. M. Mendel, "Identification of nonminimum phase system using higher order statistics," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 360–377, Mar. 1989.
[21] G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, *Signal Processing Advances in Wireless & Mobile Communications*. Upper Saddle River, NJ: Prentice-Hall, 2001, vol. 2.
[22] F. Gran and J. A. Jensen, "Multi element synthetic aparture transmission using a frequency division approach," in *Proc. IEEE Ultrasonics Symp.*, 2003, pp. 1942–1946.
[23] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an "exploratory" multi-modal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970, UMI Serials in Microform, No. 10, 085.
[24] ——, "Determination and proof of minimum uniqueness conditions for PARAFAC1," UCLA Working Papers in Phonetics pp. 22111–22117, (UMI Serials in Microform, No. 10 085), 1972.
[25] R. A. Harshman and M. E. Lundy, "The PARAFAC model for three-way factor analysis and multidimensioanal scaling," in *Res.. Methods Multimode Data Anal.*, H. G. Law, C. W. Snyder, Jr, J. Hattie, and R. P. McDonald, Eds. New York: Praeger, 1984, pp. 122–215.
[26] S. Haykin, J. H. Justice, N. L. Owsley, J. L. Yen, and A. C. Kak, *Array Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
[27] Y. Inouye and K. Hirano, "Cumulant-based blind identification of linear multi-input-multi-output systems driven by colored inputs," *IEEE Trans. Signal Process.*, vol. 45, no. 6, pp. 1543–1552, Jun. 1997.
[28] T. Jiang and N. D. Sidiropoulos, "Kruskal's permutation Lemma and the identification of CANDECOMP/PARAFAC and bilinear models with constant modulus constraints," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2625–2636, Sep. 2004.
[29] H. A. L. Kiers, "A three-step algorithm for CANDECOMP/PARAFAC analysis of large data sets with multicollinearity," *J. Chemometrics*, vol. 12, pp. 155–171, 1988.
[30] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Its Appl.*, vol. 18, pp. 95–138, 1977.
[31] J. Liang and Z. Ding, "Blind MIMO system identification based on cumulant subspace decomposition," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1457–1468, Jun. 2003.
[32] X. Liu and N. D. Sidiropoulos, "Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 2074–2086, Sep. 2001.
[33] E. Moreau and J. C. Pesquet, "Generalized contrasts for multichannel blind deconvolution of linear systems," *IEEE Trans. Signal Process. Lett.*, vol. 4, pp. 182–183, Jun. 1997.
[34] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516–525, Feb. 1995.
[35] C. L. Nikias and A. P. Petropulu, *Higher Order Spectra Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[36] C. B. Papadias, "Unsupervised receiver processing techniques for linear space-time equalization of wideband multiple input/multiple output channels Papadias," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 472–482, Feb. 2004.
[37] T. Quatieri and A. Oppenheim, "Iterative techniques for minimum phase signal reconstruction from phase or magnitude," *IEEE Trans. Signal Process.*, vol. 29, no. 6, pp. 1187–1193, Dec. 1981.
[38] R. Rajagopal and L. C. Potter, "Multivariate MIMO FIR inverses," *IEEE Trans. Image Process.*, vol. 12, no. 4, pp. 458–465, Apr. 2003.
[39] L. Rota, P. Comon, and S. Icart, "Blind equalization of MIMO channels," presented at the IEEE Workshop Signal Processing Advances Wireless Communications (SPAWCC), Rome, Italy, Jun. 15–18, 2003.
[40] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. Siganl Process.*, vol. 48, no. 3, pp. 810–823, Mar. 2000.
[41] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilienar decomposition of N-way arrays," *J. Chemometrics*, vol. 14, pp. 229–239, 2000.
[42] A. Swami, G. B. Giannakis, and G. Zhou, "Bibliography on higher-order statistics," *Signal Process.*, vol. 60, no. 1, pp. 65–126, 1997.
[43] A. Swami, HOSA—Higher Order Spectral Analysis Toolbox, The MathWorks, Inc,. Nattick, MA [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=3013
[44] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, pp. 1951–1968, Oct. 1998.
[45] M. Torlak and G. Xu, "Blind multiuser channel estimation in asynchronous CDMA systems," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 137–147, Jan. 1997.
[46] J. K. Tugnait, "Identification and deconvolution of multichannel linear non-Gaussian processes using higher order statistics and inverse filter criteia," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 658–672, Mar. 1997.
[47] ——, "Blind spatio-temporal equalization and impulse response estimation for MIMO channels using a Godard cost function," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 268–271, Jan. 1997.
[48] J. K. Tugnait and B. Huang, "Multistep linear predictors-based blind identificaion and equalization by the subspace method: Identifibiality results," *IEEE Trans. Signal Process.*, vol. 48, no. 1, pp. 26–38, Jan. 2000.
[49] A. Zerva, A. P. Petropulu, and P.-Y. Bard, "Blind deconvolution methodology for site response evaluation exclusively from ground surface seismic recordings," *J. Soil Dynamics Earthquake Eng.*, Jan. 1999.

Turev Acar (S'00–M'02) was born in Denizli, Turkey. After attending Ankara Science High School, he received the B.Sc. degree from Bilkent University, Turkey, in 2002 and the M.S. degree in electrical and computer engineering from Drexel University, Philadelphia, PA, in 2004.

Currently, he is with Maxim Integrated Products, Sunnyvale, CA.

Yuanning Yu (S'04) was born in Anhui, China. He received the B.Sc. degree from the Department of Electronic Engineering at Tsinghua University, Beijing, China, in 2002. He is currently working towards the Ph.D. degree at Drexel University, Philadelphia, PA.

Currently, he is a Research Assistant in the Department of Electrical and Computer Engineering, Drexel University. His research interests are in the area of system identification and blind source separation.

Athina P. Petropulu (S'87–M'87–SM'01) received the Diploma degree in electrical engineering from the National Technical University of Athens, Greece, in 1986 and the M.Sc. and Ph.D. degrees, both in electrical and computer engineering, from Northeastern University, Boston, MA, in 1988 and 1991, respectively.

In 1992, she joined the Department of Electrical and Computer Engineering at Drexel University, where she is now a Professor. During the academic year 1999–2000, she was an Associate Professor at Université Paris Sud, École Supérieure d'Electrícité, Paris, France. She is the coauthor (with C. L. Nikias) of the textbook entitled *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework* (Englewood Cliffs, NJ: Prentice-Hall, 1993). Her research interests span the area of statistical signal processing, wireless communications and networking, and ultrasound imaging.

Dr. Petropulu is the recipient of the 1995 Presidential Faculty Fellow Award in Electrical Engineering given by NSF and the White House. She has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS, and is a member of the editorial board of the *IEEE Signal Processing Magazine* and the *EURASIP Journal on Wireless Communications and Networking*. She is IEEE Signal Processing Society (SPS) Vice President-Conferences, member of the IEEE Signal Processing Board of Governors and the Executive Committee. She is the corecipient of the 2005 *Signal Processing Magazine* Best Paper Award. She was the General Chair of the 2005 International Conference on Acoustics Speech and Signal Processing (ICASSP), Philadelphia PA.