Blind PARAFAC Receivers for DS-CDMA Systems

Nicholas D. Sidiropoulos, *Senior Member, IEEE*, Georgios B. Giannakis, *Fellow, IEEE*, and Rasmus Bro

***Abstract*—This paper links the direct-sequence code-division multiple access (DS-CDMA) multiuser separation-equalization-detection problem to the *parallel factor (PARAFAC)* model, which is an analysis tool rooted in psychometrics and chemometrics. Exploiting this link, it derives a deterministic blind PARAFAC DS-CDMA receiver with performance close to nonblind minimum mean-squared error (MMSE). The proposed PARAFAC receiver capitalizes on code, spatial, and temporal diversity-combining, thereby supporting small sample sizes, more users than sensors, and/or less spreading than users. Interestingly, PARAFAC does not require knowledge of spreading codes, the specifics of multipath (interchip interference), DOA-calibration information, finite alphabet/constant modulus, or statistical independence/whiteness to recover the information-bearing signals. Instead, PARAFAC relies on a fundamental result regarding the uniqueness of low-rank three-way array decomposition due to Kruskal (and generalized herein to the complex-valued case) that guarantees *identifiability* of all relevant signals and propagation parameters. These and other issues are also demonstrated in pertinent simulation experiments.**

I. Introduction

BLIND separation of signals impinging on an antenna array is of paramount importance in commercial and military applications, including source localization, sensor calibration, blind signal copy, mitigation of co-channel interference, and eavesdropping, just to name a few. Existing self-recovering approaches separate the inaccessible sources from their mixtures within a scale and permutation ambiguity (which is inherently nonidentifiable from output data only). Depending on modeling assumptions and tools adopted, the resulting algorithms can be classified into

i) those that assume narrowband sources and rely on known manifolds to estimate directions of arrival using subspace techniques (which are also capable of handling wideband sources per frequency bin) [22];
ii) those that assume a special property of the sources, such as finite-alphabet (FA), constant-modulus (CM), or cyclostationarity, and attempt to restore it at the receiver [1], [35], [42];
iii) those that assume spatial independence of non-Gaussian sources and exploit cumulant statistics for separation [8], [12], [31], [36];
iv) those that assume known signatures that the inaccessible sources are coded with, and capitalize on, their orthogonality to remove multiuser interference and unknown multipath effects [17], [39].

Direct-sequence code-division multiple access (DS-CDMA) is the major motivation behind iv) because it offers a promising alternative to traditional time- and frequency-division communications.

Blind multiuser detection of DS-CDMA signals is dealt with in this paper both for downlink as well as for uplink multipoint transmissions through frequency-selective fading channels that arise due to multipath propagation in wireless environments. Blind receivers target the ultimate CDMA goal for minimal cooperation among users and are well motivated because they mitigate unknown multipath effects while obviating bandwidth-consuming training sequences (which are source signals known to both transmitters and receivers). Following [38], subspace methods for blind multichannel estimation and equalization of CDMA signals were proposed in [2], [27], [37], [39], and [48], all assuming that the underlying transfer function matrix is irreducible. In a recent contribution, [14] proposes antenna diversity to alleviate the irreducibility assumption and trade-off performance for receiver complexity. Existing approaches require knowledge of the spreading codes (at least of a "user of interest") that may not be available in nonco-operative scenarios and, most importantly, do not fully exploit all available forms of diversity (in time, space, and code domains) in identifying the underlying channels and reliably recovering the information bearing signals of interest. We explore herein a novel approach for blind DS-CDMA multiuser detection that relies on *parallel factor (PARAFAC) analysis*.

PARAFAC analysis is a tool with roots in psychometrics and chemometrics. PARAFAC is a subset of *multi-way analysis*, which can be viewed as linear algebra for *multi-way arrays* (a matrix is a two-way array). Our work links the DS-CDMA multiuser separation-equalization-detection problem to PARAFAC analysis and derives a deterministic blind PARAFAC receiver whose performance is close to nonblind minimum mean-squared error (MMSE). The proposed blind receiver supports small sample sizes, more users than sensors, and/or less spreading than users, while obviating the need to resort to space/time statistical independence/whiteness, FA/CM constraints, or DOA and array calibration information. Most notably, it does not require knowledge of the spreading codes (not even for a "user of interest"), or the specifics of the (interchip interference) multipath channels. Instead, PARAFAC relies on a fundamental result of Kruskal [24] regarding the uniqueness of low-rank three-way array decomposition to simultaneously recover all information bearing signals and

Manuscript received December 29, 1998; revised August 17, 1999. This work was supported by NSF/CAREER CCR-9733540, NSF CCR-9805350, the Nordic Industry Foundation Project P93-149, and the FØTEK Foundation, through Prof. L. Munck. The associate editor coordinating the review of this paper and approving it for publication was Prof. Michail K. Tsatsanis.

N. D. Sidiropoulos was with the Department of Electrical Engineering, University of Virginia, Charlottesville, VA 22903 USA. He is now with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: nikos@ece.umn.edu).

G. B. Giannakis is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 (e-mail: georgios@ece.umn.edu).

R. Bro is with the Chemometrics Group, Food Technology, Department of Dairy and Food Science, The Royal Veternary and Agricultural University, Frederiksberg, Denmark (e-mail: rb@kvl.dk).

Publisher Item Identifier S 1053-587X(00)01591-9.

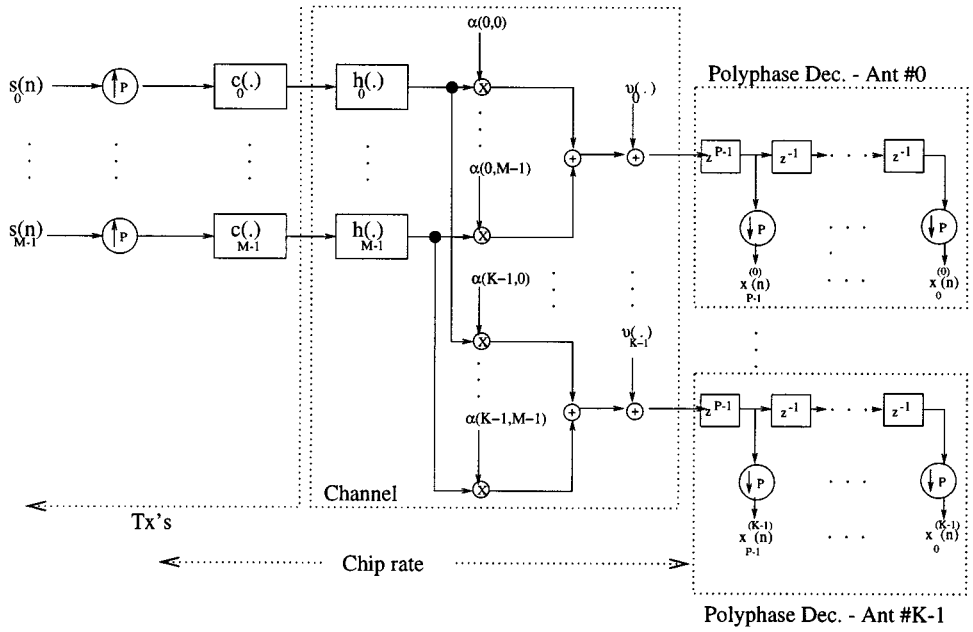

Fig. 1. Multiuser/multirate discrete-time equivalent baseband CDMA model (chip rate).

associated system parameters from output data only. This result is also generalized herein to the complex-valued case in order to handle complex modulations.

mMulti-way analysis is largely unknown to the signal processing community, although elements of it have implicitly or explicitly appeared in the context of higher-than-second-order statistics, under source-independence assumptions, for blind multichannel system identification [15] and, thus, source separation, even when the available sensors are less than the sources [31] (see also [13] and [26] for an independent component analysis (ICA) perspective and related Schur/Jacobi-based algorithmic approaches). However, we herein (and in related conference papers [6], [7], and [32]) take a deterministic approach in linking multi-way analysis tools with some key signal processing and communications problems, without requiring statistical independence.

A. Organization

This paper is structured as follows. Section II lays out our assumptions and develops the discrete-time baseband-equivalent data model. Section III summarizes important background in parallel factor analysis. Section IV discusses identifiability issues, whereas Section V deals with algorithmic issues. Section VI discusses modes of usage and establishes links to other applications of PARAFAC ideas in signal processing and communications. Section VII presents simulation results, whereas Section VIII summarizes our conclusions. The proof of a key theorem can be found in the Appendix.

II. Data Model

A. Modeling Preliminaries

The block diagram in Fig. 1 represents a CDMA system described in terms of its discrete-time baseband-equivalent model, where signals, codes, and channels are represented by samples of their complex envelopes taken at the chip rate (see also [38] and [39]). Upsamplers and downsamplers serve the purpose of multiplexing and demultiplexing (spreading and despreading by a factor P). Each of the M users spreads the corresponding information sequence $s_m(n)$ with the upsampler and encodes it using a code $c_m(p)$ of length P before transmission through the *unknown* FIR channel $h_m(l)$, which, in addition to multipath, includes the transmit spectral-shaping pulse, the receive filter, and the mth user's asynchronism in the form of delay factors. If Nyquist pulse shapers are adopted, their effect disappears in our discrete-time model [28, p. 542]. The receiver employs an array of K antennas. The baseband output of each antenna is sampled at the chip rate and decomposed into its P polyphase components.

We now summarize our working assumptions—some of which are already implicit in the model in Fig. 1. The spreading codes $c_m(p)$ are assumed to be symbol-periodic, and the spreading gain (P) is known or has been estimated (e.g., using cyclostationarity tests). The number of active users (M) is known, and symbol-level (but not necessarily chip-level) synchronization is available. The linear channels $\{h_m(l)\}_{m=0}^{M-1}$ remain time invariant over N symbols (as we will see, N can be as small as 2), and an upper bound $L \geq \max_{m=0,\ldots,M-1} \text{order}(h_m(l))$ on the order of all channels is known. In addition, we assume the following.

- a1) The multipath/delay channel between user m and antenna k only depends on m, modulo a flat fading/antenna response pattern factor $\alpha(k,m)$ that is time-invariant over N symbols. This is valid when multipath reflectors are in the far field (relative to the receive antenna array); see also [41] and references therein.
- a2) Either one of the following blocking assumptions common to block processing methods (see e.g., [2], [14], [27], [37]–[39]) is applicable:
  - **a2.1)** The spreading gain $P = P_f + L$, and $P_f \geq 2$ intersymbol interference (ISI)-free chips are available ("discard prefix").
  - **a2.2)** The codes $c_m(p)$ include $L_g \geq L$ trailing zeros ("guard chips").

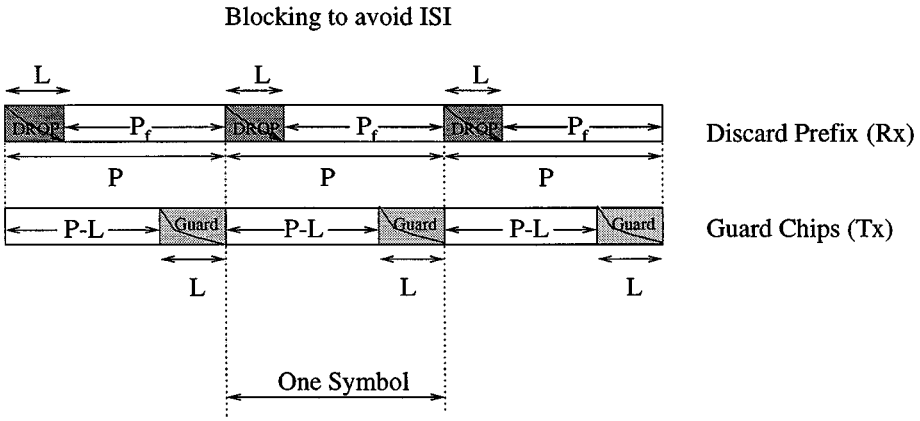

Fig. 2. Blocking.

Either of a2.1)–a2.2) enables data blocking that avoids ISI; interchip interference (ICI) remains, but, as we will see, it can be relatively easily dealt with. If a2.1) is in effect, then an ISI-free block data model can be derived using a "discard prefix" strategy that is implemented at the receiver; thus, it is also applicable in a noncooperative context. The "guard chips" assumption a2.2) requires a cooperative transmitter. The "discard prefix"/"guard chips" idea is illustrated in Fig. 2. Since the channels also incorporate relative delays, both a2.1) and a2.2) effectively assume delay spreads in the order of a few chips. This is the case in quasisynchronous CDMA systems [21]. As an example, consider $M = 32$ users in an urban microcell of radius 1 Km, and suppose that each user transmits at 10 Kbps with a spreading gain of $P = M = 32$ chips. The chip duration is $T_c \approx 3\ \mu$ s. Suppose that the base station broadcasts a pilot signal and that the mobiles synchronize their transmissions with the pilot (e.g., transmit immediately upon receipt of the pilot). Assuming that the maximum round-trip path length is ≈3 Km, the round-trip delay spread is ≈10 μs, or three to four chips. The same figure applies to a rural setting with a radius of 12 Km and $P = M = 4$ users in line of sight. The model also applies to "synchronous" systems in the presence of oscillator drifts and relative motion between the mobiles and the base station (see also [21]). Henceforth, we use P to denote the number of ISI-free chips available, with the understanding that it stands for P_f when "discard prefix" is in effect or the spreading gain itself when "guard chips" is in effect. In the interest of brevity, we also index the ISI-free chips starting from 0, regardless of whether "discard prefix" or "guard chips" is in effect.

B. Prelude—No ICI

It is instructive to begin with a noiseless/memoryless (synchronous) model and build from it. Toward this end, consider Fig. 1, discarding the noise v_k, $k = 0, \ldots, K-1$, and multipath channels h_m, $m = 0, \ldots, M-1$. Recall that

- $\alpha(k,m)$ fading/gain between user m and antenna element k;
- $c_m(p)$ pth chip of the spreading code of user m;
- $s_m(n)$ nth symbol transmitted by user m.

Letting $x_{k,n,p}$ denote the baseband output of the kth antenna, for symbol n and chip p, it holds that

$$x_{k,n,p} = \sum_{m=0}^{M-1} \alpha(k,m) c_m(p) s_m(n) \qquad \begin{cases} k = 0, \ldots, K-1 \\ n = 0, \ldots, N-1 \\ p = 0, \ldots, P-1. \end{cases} \tag{1}$$

It will prove convenient to recast this model in matrix form. Let us define $K \times N$ data matrices

$$\mathbf{X}_p := \begin{bmatrix} x_{0,0,p} & x_{0,1,p} & \cdots & x_{0,N-1,p} \\ x_{1,0,p} & x_{1,1,p} & \cdots & x_{1,N-1,p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{K-1,0,p} & x_{K-1,1,p} & \cdots & x_{K-1,N-1,p} \end{bmatrix}$$

for $p = 0, 1, \ldots, P-1$. After some algebraic manipulation, it can be shown that $\mathbf{X}_p$ admits the factorization

$$\mathbf{X}_p = \mathbf{A}\mathbf{D}_p(\mathbf{H})\mathbf{S} \tag{2}$$

where

$$\mathbf{A} := \begin{bmatrix} \alpha(0,0) & \cdots & \alpha(0,M-1) \\ \alpha(1,0) & \cdots & \alpha(1,M-1) \\ \vdots & \vdots & \vdots \\ \alpha(K-1,0) & \cdots & \alpha(K-1,M-1) \end{bmatrix}$$

$$\mathbf{H} := \begin{bmatrix} c_0(0) & \cdots & c_{M-1}(0) \\ c_0(1) & \cdots & c_{M-1}(1) \\ \vdots & \vdots & \vdots \\ c_0(P-1) & \cdots & c_{M-1}(P-1) \end{bmatrix}$$

$$\mathbf{S} := \begin{bmatrix} s_0(0) & \cdots & s_0(N-1) \\ s_1(0) & \cdots & s_1(N-1) \\ \vdots & \vdots & \vdots \\ s_{M-1}(0) & \cdots & s_{M-1}(N-1) \end{bmatrix}$$

and $\mathbf{D}_p(\cdot)$ is understood as an operator that extracts the $(p+1)$ row of its matrix argument and constructs a diagonal matrix out of it. By convention, indices start from zero, whereas the first row of a matrix is row one; hence, we get $+1$. Matrix $\mathbf{A} \in$

$\mathbb{C}^{K\times M}$ denotes the compound flat fading/array response pattern, $\mathbf{H}\in\mathbb{C}^{P\times M}$ is the spreading code matrix, and $\mathbf{S}\in\mathbb{C}^{M\times N}$ is the information bearing signal matrix.

In the presence of noise, the observation model becomes

$$\tilde{\mathbf{X}}_p=\mathbf{X}_p+\mathbf{V}_p=\mathbf{A}\mathbf{D}_p(\mathbf{H})\mathbf{S}+\mathbf{V}_p \quad p=0,1,\ldots,P-1$$

where the $K\times N$ matrix $\mathbf{V}_p$ is the measurement noise corresponding to the pth polyphase matrix component (tilde will be used throughout to denote noisy measurements).

C. Incorporating ICI

If ICI is present, the only thing that changes [assuming either a2.1) "discard prefix" or a2.2) "guard chips")] is that the individual user codes now become "effective codes" or "effective signatures": the result of convolving a user's spreading code with the impulse response of the respective multipath channel and, if a2.1) is in effect, discarding the first L chips. Thus, only $\mathbf{H}$ changes (including number of rows = number of ISI-free chips), and it is now more appropriately called the aggregate multipath channel/spreading matrix—hence the choice of symbol $\mathbf{H}$.

III. Intermezzo: Parallel Factor Analysis

Recall (1), reproduced here for convenience:

$$x_{k,n,p}=\sum_{m=0}^{M-1}\alpha(k,m)c_m(p)s_m(n).$$

Note that $x_{k,n,p}$ in (1) is a sum of triple products; it is variably known as the trilinear model, trilinear decomposition, triple product decomposition, canonical decomposition, or PARAFAC analysis. It has been first introduced as a data analysis tool in psychometrics,[1] where it has been used for "individual differences multidimensional scaling," but also in phonetics, exploratory data analysis, statistics, arithmetic complexity, and other fields and disciplines [3]. Nowadays, most of the research in the area is conducted in the context of chemometrics,[2] where it is used for spectrophotometric, chromatographic, and flow injection analyses. The term PARAFAC is widely adopted in chemometrics, and we adopt it here as well, although we also use the other terms when convenient, particularly in stating theoretical results.

Harshman [18] developed the PARAFAC model as a natural extension of the ideas put forward by Cattell [11]. At the same time, Caroll and Chang [9] introduced the canonical decomposition (CANDECOMP) model, which is essentially identical to PARAFAC but was developed as a natural extension of multidimensional scaling tools. In signal processing and communications terms, PARAFAC can be thought of as a generalization of ESPRIT ideas—actually as a general principle underlying the most general formulation of ESPRIT to date. PARAFAC can also be seen as generalizing joint approximate diagonalization ideas [41]. We will return to these points in Section VI.

[1]To the best of our knowledge, the first seed of PARAFAC ideas appeared as early as 1944, in a paper by R. B. Cattell in *Psychometrika* [11].

[2]Data analysis and its applications in chemistry.

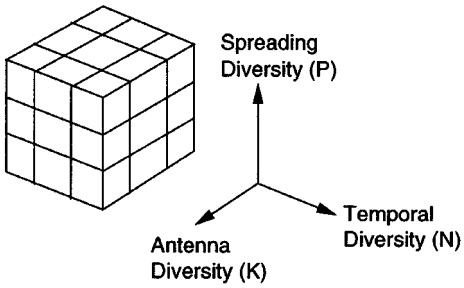

Fig. 3. Datacube in diversity space.

Define a $K\times N\times P$ *three-way array* $\underline{\mathbf{X}}$ with typical element $x_{k,n,p}$. A rank-1 three-way array $\underline{\mathbf{E}}$ with typical element $\epsilon_{k,n,p}$ can be written as

$$\epsilon_{k,n,p}=a(k)c(p)s(n).$$

Equation (1) is an M-component triple product decomposition of $\underline{\mathbf{X}}$. The *rank* of $\underline{\mathbf{X}}$ is defined to be the minimum number of rank-1 three-way components needed to decompose $\underline{\mathbf{X}}$ [24]. The above definition is consistent with matrix (two-way array) rank: the minimum number of rank-1 matrices needed to decompose a given matrix. PARAFAC is thus naturally related to linear algebra for multi-way arrays [24], which is also known as multi-way analysis.

In the present context, $\underline{\mathbf{X}}$ contains the observation data arranged in a three-way *diversity space*—(antenna#,symbol#, chip#)—reflecting the three different kinds of diversity available:

- spatial;
- temporal;
- spreading diversity.

We refer to $\underline{\mathbf{X}}$ as the *diversity datacube*[3]. This concept is illustrated in Fig. 3.

Note that (1) affords a scalar view of the diversity datacube. Another view, namely $\mathbf{X}_p=\mathbf{A}\mathbf{D}_p(\mathbf{H})\mathbf{S}, p=0,1,\ldots,P-1$, is afforded by the polyphase matrix decomposition of the data in (2). This alternative view can be interpreted as "slicing" the 3-D array $\underline{\mathbf{X}}$ in a series of "slabs" (2-D arrays) perpendicular to one of the diversity dimensions: in this case, the spreading dimension. Indeed, it is easy to see that according to our definitions and for a fixed p, $\mathbf{X}_p$ is nothing but $[x_{:,:,p}]$: the $K\times N$ 2-D slice of $\underline{\mathbf{X}}$ corresponding to the given p.

The perfect symmetry of the trilinear model in (1) allows two more matrix system rearrangements, which can be interpreted as "slicing" the three-way array $\underline{\mathbf{X}}$ along different dimensions. In particular

$$\mathbf{Y}_k=\mathbf{S}^T\mathbf{D}_k(\mathbf{A})\mathbf{H}^T,\quad k=0,1,\ldots,K-1 \tag{3}$$

where the $N\times P$ matrix $\mathbf{Y}_k:=[x_{k,:,:}]$. Similarly

$$\mathbf{Z}_n=\mathbf{H}\mathbf{D}_n(\mathbf{S}^T)\mathbf{A}^T,\quad n=0,1,\ldots,N-1 \tag{4}$$

where the $P\times K$ matrix $\mathbf{Z}_n:=[x_{:,n,:}]$. These alternative "views" of the data are useful in understanding the proofs of

[3]Although it does not literally have to be a *cube*, i.e., in general, $K\neq N\neq P$.

core theoretical results, as well as algorithmic PARAFAC issues. Note that we may choose the rows of any one of the three matrices $\mathbf{A}, \mathbf{S}^T$, or $\mathbf{H}$ to construct the diagonal matrices that appear in the middle of the decomposition. An important feature of the different data rearrangements is revealed when two of the three parameters K, P, N are $\geq M$ (e.g., $P \geq M$ and $N \geq M$). We may then *always* select a matrix system representation for which the left matrix is tall and the right matrix is fat.

Our link of the DS-CDMA setup to PARAFAC affords a powerful identifiability result, plus the opportunity to tap on and extend the available expertise for fitting the PARAFAC model, to derive a deterministic (least squares) joint blind estimation algorithm. These are explored next.

IV. Identifiability

The following concept is key for this section.

Definition 1: Consider a matrix $\mathbf{B} \in \mathbb{C}^{I \times J}$. If $\text{rank}(\mathbf{B}) = r$, then $\mathbf{B}$ contains a collection of r linearly independent columns. Moreover, if *every* $\ell \leq J$ columns of $\mathbf{B}$ are linearly independent, but this does not hold for every $\ell+1$ columns, then $\mathbf{B}$ has k-**rank** $k_{\mathbf{B}} = \ell$. Note that $k_{\mathbf{B}} \leq \text{rank}(\mathbf{B}), \forall \mathbf{B}$.

The concept of k-rank is implicit in the seminal work of Kruskal [24], but the term was later coined by Harshman and Lundy [19] (k-rank stands for *Kruskal-rank*). The distinction between rank and k-rank is important. Consider the following two matrices:

$$\begin{bmatrix} \alpha & 1 & 0 \\ \beta & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} \alpha & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

For $\alpha, \beta \neq 0$, the left matrix has full column rank (two) *and* full k-rank (two), whereas the right matrix has full column rank (two) but *not* full k-rank (one). In other words, k-rank $= 2$ requires that *every* two columns are linearly independent, whereas rank $= 2$ simply requires that there exists at least *one* pair of linearly independent columns.

A matrix whose columns are drawn independently from an absolutely continuous distribution has full rank with probability one. Interestingly, it also has full k-rank because any combination of columns can be thought of as another random matrix with columns drawn independently from an absolutely continuous distribution.

A distinguishing feature of the trilinear model is its uniqueness. Under mild conditions and unlike the unconstrained bilinear model, the trilinear model does not suffer from rotational freedom (i.e., $\mathbf{A}, \mathbf{H}, \mathbf{S}$ are identifiable without unitary matrix ambiguities). Several results regarding uniqueness of this model have been known from early on (and, in fact, sparked the interest in trilinear modeling in the data analysis community), e.g., [18], [24], and [25]. Among them, the following (due to Kruskal [24]) is the deepest.

Theorem 1: [24] Consider the set of P matrices $\mathbf{X}_p = \mathbf{A}\mathbf{D}_p(\mathbf{H})\mathbf{S}$, $p = 0, 1, \ldots, P-1$, where $\mathbf{A} \in \mathbb{R}^{K \times M}$, $\mathbf{H} \in \mathbb{R}^{P \times M}$, $\mathbf{S} \in \mathbb{R}^{M \times N}$, and M denotes the common dimension. If

$$k_{\mathbf{A}} + k_{\mathbf{H}} + k_{\mathbf{S}^T} \geq 2(M+1) \tag{5}$$

then $\mathbf{A}$, $\mathbf{H}$, and $\mathbf{S}^T$ are unique up to (inherently unresolvable from output data only) permutation and scaling of columns, meaning that any other triple $\bar{\mathbf{A}}, \bar{\mathbf{H}}, \bar{\mathbf{S}}^T$ that gives rise to the data $\mathbf{X}_p$, $p = 0, 1, \ldots, P-1$ is related to $\mathbf{A}, \mathbf{H}, \mathbf{S}^T$ via

$$\bar{\mathbf{A}} = \mathbf{A}\mathbf{\Pi}\mathbf{\Delta}_1, \quad \bar{\mathbf{H}} = \mathbf{H}\mathbf{\Pi}\mathbf{\Delta}_2, \quad \bar{\mathbf{S}}^T = \mathbf{S}^T\mathbf{\Pi}\mathbf{\Delta}_3 \tag{6}$$

where $\mathbf{\Pi}$ is a permutation matrix, and $\mathbf{\Delta}_{1,2,3}$ are diagonal scaling matrices satisfying

$$\mathbf{\Delta}_1\mathbf{\Delta}_2\mathbf{\Delta}_3 = \mathbf{I}. \tag{7}$$

The result holds for $M = 1$, irrespective of condition (5), as long as $\underline{\mathbf{X}}$ does not contain an identically zero 2-D slice along any dimension.

Note that for finite M, the set of permutation matrices is finite (as opposed to the set of unitary matrices, which is infinite). Furthermore, in our DS-CDMA context (as well as in other applications, such as source separation), symbol matrix recovery within a permutation matrix ambiguity amounts to shuffling the users. This is not a major concern, and it can be resolved by resorting to *a priori* or embedded information, e.g., spreading codes, and user ID bits, respectively. The scale ambiguity is relevant for digital communications applications, but it can be resolved using AGC and differential encoding/decoding [28, p. 187] or phase estimation.

Kruskal, in [24], states and proves his uniqueness of trilinear decomposition explicitly under the working assumption of real-valued arrays. Communication systems employing complex modulation (in-phase/quadrature processing) lead to discrete-time baseband equivalent models involving complex arrays. The generalization of Kruskal's results to the complex case is nontrivial because of the following.

- If we try to cast the complex model in (1) or (2) in terms of real and imaginary parts, we have the following.
  - Starting from (1), we are led to two real scalar trilinear models for the real and imaginary parts of the data, each involving *four times* the factors of the complex scalar trilinear model, albeit constrained in a particular fashion.
  - Starting from the parallel slabs model (2), we are led to a block PARAFAC-like model for the unfolded real-imaginary data slabs in which the middle matrix is tri-diagonal, instead of diagonal. If we constrain one of the factors to be real, then the middle matrix can be made diagonal, but then, the k-rank of the matrix constructed out of these diagonals is always less than or equal to one, violating (5), which requires that the k-rank of each of the three matrices is at least two[4].
- We should be careful in extending factorization results from the real-valued case to the complex-valued case (e.g., uniqueness of prime factorization).

[4]The k-rank of any one of the three matrices involved can never be greater than the number of factors (users) M. If the k-rank of one of the three matrices is 1, then the other two can never make up for it; the sum of the three k-ranks will be bounded above by $2M + 1 < 2(M + 1)$ in this case. Hence, it is necessary for all three matrices to have k-rank at least 2 for (5) to be satisfied.

- Kruskal [25] gives an example of a $2 \times 2 \times 2$ array of real numbers whose rank is 3 when considered over the field of reals, yet when the *same* array of *real numbers* is *simply thought of* as being a complex array, it has rank 2. Care should therefore be exercised in treating the complex case.

In his seminal paper [24], Kruskal actually proves a series of partial uniqueness results, culminating in a set of conditions that are jointly satisfied by the k-rank condition (5). His line of argument is exquisite but long and laborious (≈20 pages). Herein, we are interested in generalizing the k-rank condition to the complex case and applying it to DS-CDMA systems. Based on this hindsight and using ideas from Kruskal, we develop a concise proof for the complex case, which also helps illuminate the meaning and role of k-rank. The result is summarized in the following Theorem, which is proven in the Appendix.

Theorem 2: Given $\mathbf{X}_p = \mathbf{A}\mathbf{D}_p(\mathbf{H})\mathbf{S}$, $p = 0, 1, \ldots, P-1$, $\mathbf{A} \in \mathbb{C}^{K \times M}$, $\mathbf{H} \in \mathbb{C}^{P \times M}$, $\mathbf{S} \in \mathbb{C}^{M \times N}$, if $k_{\mathbf{A}} + k_{\mathbf{H}} + k_{\mathbf{S}^T} \geq 2(M+1)$, then $\mathbf{A}$, $\mathbf{H}$, and $\mathbf{S}^T$ are unique up to permutation and (complex) scaling of columns.

Before presenting the algorithmic ramifications of Theorem 2, we elaborate on its implications to our blind DS-CDMA problem at hand.

A. Interpretation from a Communications Viewpoint

Even though the PARAFAC uniqueness result is purely deterministic, it also admits a statistical characterization. Recall that a matrix whose columns are drawn independently from an absolutely continuous distribution has full k-rank (equal to its rank) with probability one, even when the elements across a given column are dependent random variables. In our present context, we have the following.

- User-wise independent fading (or, user-wise independent DOA's and unambiguous array manifold) imply that $\mathbf{A}$ is full k-rank with probability one.
- "Persistence of excitation" of user symbols implies that $\mathbf{S}^T$ is full k-rank. This is certainly valid for communications transmissions, provided N is big enough. For small N, the presence of—inevitable—nuisance effects like residual carriers will also render $\mathbf{S}^T$ full k-rank with probability one.
- User-wise independent multipath/delay channel taps assure that $\mathbf{H}$ is also full k-rank.

Under these justifications for the full k-rank conditions, (5) yields

$$\min(K, M) + \min(N, M) + \min(P, M) \geq 2M + 2 \quad (8)$$

which has the following important corollaries.

1) If $N \geq M$, $P \geq M$ (typical in DS-CDMA), then $K \geq 2$ antennas are sufficient for M users; hence, our system is capable of supporting more users than sensors.
2) If $K \geq M$, $P \geq M$, then $N \geq 2$ symbols are sufficient; hence, fading and multipath can potentially vary as fast as half the symbol rate.
3) If $K \geq M$, $N \geq M$, then $P \geq 2$ (ISI-free) chips are sufficient (recall that the actual spreading gain must be $\geq L+2$); therefore, we do not necessarily need orthogonal or quasiorthogonal spreading sequences, which is particularly important in oversaturated systems (number of users greater than the spreading gain).
4) However, the most interesting case is precisely when some or all of $\mathbf{A}$, $\mathbf{H}$, $\mathbf{S}^T$ are not of "proper" dimensions (i.e., fat instead of tall), yet the k-rank condition (5)–(8) is satisfied, and the model is identifiable. An interesting instance occurs when $K = N = P = 4$, and $M = 5$; none of $\mathbf{A}$, $\mathbf{H}$, $\mathbf{S}^T$ is tall, yet $\min(K, M) + \min(N, M) + \min(P, M) = 12 = 2M+2$, and the model is generically (meaning with probability 1 under our statistical assumptions) identifiable. Of course, more significant diversity gains are possible for bigger M, e.g., $M = 8$.

Remark 1: Conditions (5)–(8) characterize the *diversity tradeoff*, i.e., the intricate balance of different kinds of diversity (spatial, temporal, spreading) that leads to identifiability. Fig. 4 depicts the boundary of the identifiability region [where $\min(K, M) + \min(N, M) + \min(P, M) = 2M + 2$] for $M = 8$ users.

V. Trilinear Alternating Least Squares Regression

The principle of alternating least squares (ALS) can be used to fit the trilinear model on the basis of noisy observations. In the noiseless case, and under considerably stronger conditions, $\mathbf{A}$, $\mathbf{H}$, and $\mathbf{S}$ can indeed be found using eigenanalysis. The following result—generalizing ESPRIT, but rediscovered many times ([30] is one of the earliest)—is a simpler *constructive* proof of uniqueness, albeit under restrictive (dimensionality, full-rank) conditions. It is useful in initializing ALS iterations whenever its conditions are met. A short proof follows.

Theorem 3 (Constructive Proof of Uniqueness—Full-Rank Case): Given $\mathbf{X}_p = \mathbf{A}\mathbf{D}_p(\mathbf{H})\mathbf{S}$, $p = 0, 1, \ldots, P-1$, $\mathbf{A} \in \mathbb{C}^{K \times M}$, $\mathbf{H} \in \mathbb{C}^{P \times M}$, $\mathbf{S} \in \mathbb{C}^{M \times N}$, suppose that two of the three matrices $\mathbf{A}$, $\mathbf{H}$, $\mathbf{S}^T$, say without loss of generality $\mathbf{A}$ and $\mathbf{S}^T$, are tall and full rank, and the corresponding third ($\mathbf{H}$ in this case) has a pair of columns corresponding to, say $p = 0$ and $p = 1$, such that the associated diagonals $\mathbf{D}_0(\mathbf{H})$ and $\mathbf{D}_1(\mathbf{H})$ have no zero elements, and $\mathbf{D}_1(\mathbf{H})\mathbf{D}_0^{-1}(\mathbf{H})$ has distinct elements. Then, $\mathbf{A}$, $\mathbf{H}$, and $\mathbf{S}^T$ are unique up to permutation and (complex) scaling of columns as in (6) and (7) and can be recovered from the data in a single shot.

Proof: Due to the inherent scale ambiguity, without loss of generality we can absorb the first diagonal into the $\mathbf{S}$ matrix and consider:

$$\mathbf{X}_0 = \mathbf{A}\mathbf{S}, \quad \mathbf{X}_1 = \mathbf{A}\mathbf{D}\mathbf{S} \quad (9)$$

where $\mathbf{D} := \mathbf{D}_1(\mathbf{H})\mathbf{D}_0^{-1}(\mathbf{H})$ is diagonal nonsingular $M \times M$, $\mathbf{A}$ is $K \times M$ tall and full rank, and $\mathbf{S}$ is $M \times N$ fat and full rank. Construct $N \times N$ sample auto- and cross-correlation matrices:

$$\mathbf{R}_0 = \mathbf{X}_0^H \mathbf{X}_0 = \mathbf{S}^H \mathbf{A}^H \mathbf{A} \mathbf{S}$$
$$\mathbf{R}_1 = \mathbf{X}_0^H \mathbf{X}_1 = \mathbf{S}^H \mathbf{A}^H \mathbf{A} \mathbf{D} \mathbf{S}$$

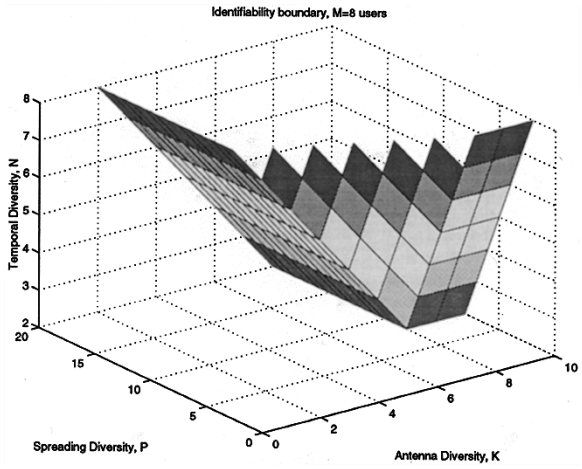

Fig. 4. Boundary of identifiability region for $M = 8$ users.

where $(\cdot)^H$ stands for Hermitian transpose. Define the $N \times M$ matrix $\mathcal{A} := \mathbf{S}^H \mathbf{A}^H \mathbf{A}$, and note that it is tall and full rank under our working assumptions. We have

$$\mathbf{R}_0 = \mathcal{A}\mathbf{S}, \quad \mathbf{R}_1 = \mathcal{A}\mathbf{D}\mathbf{S}. \tag{10}$$

Let $\mathbf{\Gamma} := \mathcal{A}^\dagger$, which is the pseudo-inverse of $\mathcal{A}$. It then follows from (10) that

$$\mathbf{\Gamma}\mathbf{R}_0 = \mathbf{D}^{-1}\mathbf{\Gamma}\mathbf{R}_1 \tag{11}$$

and determining the $M \times N$ matrix $\mathbf{\Gamma}$ from $\mathbf{X}_0$, $\mathbf{X}_1$ amounts to solving a generalized eigenvalue problem. Letting γ_m^H denote the mth row of $\mathbf{\Gamma}$ and λ_m be the mth element along the diagonal of $\mathbf{D}^{-1}$, we find from (11) that

$$\gamma_m^H(\mathbf{R}_0 - \lambda_m \mathbf{R}_1) = \mathbf{0}, \quad m = 1, \ldots, M.$$

Because $\mathbf{R}_0$, $\mathbf{R}_1$ are both of rank M, the λ_m's and γ_m^H's are the M generalized eigenvalues and left generalized eigenvectors (restricted in the column space of $\mathbf{R}_0, \mathbf{R}_1$) of the matrix pencil $(\mathbf{R}_0 - \lambda \mathbf{R}_1)$ (see, e.g., [20] and [50]). Once $\mathbf{\Gamma}$ is recovered (modulo permutation and scale), $\mathbf{S}$ can be recovered modulo permutation and scale from (10) as $\mathbf{S} = \mathbf{\Gamma}\mathbf{R}_0$. Then, $\mathbf{A}$ can be obtained from the original data in (9) as $\mathbf{A} = \mathbf{X}_0\mathbf{S}^\dagger$ (note that $\mathbf{S}$ is fat and full rank, and the pseudo-inverse simply carries over the permutation and scale ambiguity to the columns of $\mathbf{A}$). Finally, the diagonals can be recovered modulo permutation and scale using both pseudo-inverses as $\mathbf{D} = \mathbf{A}^\dagger \mathbf{X}_1 \mathbf{S}^\dagger$, and the proof is complete. □

Although the algorithm implied by Theorem 3 is single shot, its relatively restrictive assumptions on the dimensionality and rank of $\mathbf{A}$, $\mathbf{H}$, $\mathbf{S}$ prevent us from taking full advantage of the practically important (diversity combining) implications of the relaxed k-rank condition (5)–(8) of Theorem 2. In contrast, although iterative, the ALS algorithm we describe next capitalizes fully on the identifiability of Theorem 2 in addition to possessing least-squares optimality.

The basic idea behind ALS is simple: Each time, update a subset of parameters using least squares conditioned on previously obtained estimates of the remaining parameters; proceed to update another subset of parameters; repeat until convergence. We are given the noisy data $\underline{\tilde{\mathbf{X}}}$ and wish to estimate $\mathbf{A}, \mathbf{H}$, and $\mathbf{S}$. Consider the polyphase decomposition (2) and concatenate the $\mathbf{X}_p$'s to obtain

$$\begin{bmatrix} \mathbf{X}_0 \\ \vdots \\ \mathbf{X}_{P-1} \end{bmatrix}_{KP \times N} = \begin{bmatrix} \mathbf{A}\mathbf{D}_0(\mathbf{H}) \\ \vdots \\ \mathbf{A}\mathbf{D}_{P-1}(\mathbf{H}) \end{bmatrix}_{KP \times M} \mathbf{S}. \tag{12}$$

Least squares fitting (and ML parameter estimation, when the noise is modeled as temporally and spatially white Gaussian and all other parameters are treated as deterministic unknowns) amounts to

$$\min_{\mathbf{A},\mathbf{H},\mathbf{S}} \left\| \begin{bmatrix} \tilde{\mathbf{X}}_0 \\ \vdots \\ \tilde{\mathbf{X}}_{P-1} \end{bmatrix} - \begin{bmatrix} \mathbf{A}\mathbf{D}_0(\mathbf{H}) \\ \vdots \\ \mathbf{A}\mathbf{D}_{P-1}(\mathbf{H}) \end{bmatrix} \mathbf{S} \right\|_F^2 \tag{13}$$

where $\tilde{\mathbf{X}}_p$, $p = 0, 1, \ldots, P-1$ are the noisy slabs, and $\|\cdot\|_F$ stands for the Frobenius norm. It follows that the conditional least squares update for $\mathbf{S}$ is

$$\hat{\mathbf{S}} = \begin{bmatrix} \hat{\mathbf{A}}\mathbf{D}_0(\hat{\mathbf{H}}) \\ \vdots \\ \hat{\mathbf{A}}\mathbf{D}_{P-1}(\hat{\mathbf{H}}) \end{bmatrix}^\dagger \begin{bmatrix} \tilde{\mathbf{X}}_0 \\ \vdots \\ \tilde{\mathbf{X}}_{P-1} \end{bmatrix}$$

where $(\cdot)^\dagger$ stands for pseudo-inverse, and $\hat{\mathbf{A}}$, $\hat{\mathbf{H}}$ denote previously obtained estimates of $\mathbf{A}$ and $\mathbf{H}$. Similarly, from the second way of slicing the 3-D data $\mathbf{Y}_k = \mathbf{S}^T\mathbf{D}_k(\mathbf{A})\mathbf{H}^T$, $k = 0, 1, \ldots, K-1$, which is rewritten as

$$\begin{bmatrix} \mathbf{Y}_0 \\ \vdots \\ \mathbf{Y}_{K-1} \end{bmatrix} = \begin{bmatrix} \mathbf{S}^T\mathbf{D}_0(\mathbf{A}) \\ \vdots \\ \mathbf{S}^T\mathbf{D}_{K-1}(\mathbf{A}) \end{bmatrix} \mathbf{H}^T$$

it follows that LS fitting is equivalent to

$$\left\| \begin{bmatrix} \tilde{\mathbf{Y}}_0 \\ \vdots \\ \tilde{\mathbf{Y}}_{K-1} \end{bmatrix} - \begin{bmatrix} \mathbf{S}^T\mathbf{D}_0(\mathbf{A}) \\ \vdots \\ \mathbf{S}^T\mathbf{D}_{K-1}(\mathbf{A}) \end{bmatrix} \mathbf{H}^T \right\|_F^2$$

and the conditional LS update for $\mathbf{H}$ is

$$\hat{\mathbf{H}}^T = \begin{bmatrix} \hat{\mathbf{S}}^T\mathbf{D}_0(\hat{\mathbf{A}}) \\ \vdots \\ \hat{\mathbf{S}}^T\mathbf{D}_{K-1}(\hat{\mathbf{A}}) \end{bmatrix}^\dagger \begin{bmatrix} \tilde{\mathbf{Y}}_0 \\ \vdots \\ \tilde{\mathbf{Y}}_{K-1} \end{bmatrix}.$$

Finally, from the third way of slicing the data $\mathbf{Z}_n = \mathbf{H}\mathbf{D}_n(\mathbf{S}^T)\mathbf{A}^T$, $n = 0, 1, \ldots, N-1$, we find the conditional LS update for $\mathbf{A}$ as

$$\hat{\mathbf{A}}^T = \begin{bmatrix} \hat{\mathbf{H}}\mathbf{D}_0(\hat{\mathbf{S}}^T) \\ \vdots \\ \hat{\mathbf{H}}\mathbf{D}_{N-1}(\hat{\mathbf{S}}^T) \end{bmatrix}^\dagger \begin{bmatrix} \tilde{\mathbf{Z}}_0 \\ \vdots \\ \tilde{\mathbf{Z}}_{N-1} \end{bmatrix}.$$

Note that the conditional update of any given matrix may either improve or maintain but cannot worsen the current fit. Global monotone convergence to (at least) a local minimum follows directly from this observation.

ALS (initialized using eigenanalysis when Theorem 3 is applicable and randomly otherwise) has been proven to be the

method of choice in applications of PARAFAC in chemometrics, for a variety of reasons outlined in [3]. The trilinear ALS method is appealing primarily because it is guaranteed to converge monotonically but also because of its relative simplicity (no parameters to tune, and each step solves a standard LS problem) and good performance [3]. Experience shows that convergence to the global minimum is almost always achieved whenever the k-rank condition is satisfied [3]. The explanation (and fundamental difference *vis-a-vis* unstructured bilinear ALS solutions of $\mathbf{X} = \mathbf{AS}$ problems) lies in the inherent *uniqueness* of the trilinear model. The ALS approach also allows easy incorporation of a weighted loss function, missing values, and constraints on some or all of the factors [5].

VI. Further Considerations

In cooperative multiuser communication systems, the spreading codes are known. If no multipath is present, or if the multipath channels can be effectively estimated (e.g., using training), then $\mathbf{H}$ (or a good estimate thereof) can be assumed to be known. In this case, we can initialize ALS with the given $\mathbf{H}$. This speeds up convergence significantly and improves performance (smaller MSE). Similar arguments can be made if estimates are available for any one or two of $\mathbf{A}, \mathbf{H}, \mathbf{S}$. In practice, PARAFAC will be most useful for "cold-starting" or "boot-strapping" the equalization/detection process when little is known about the system. Once good estimates of the system parameters are acquired via PARAFAC, a simpler tracking loop, like, e.g., adaptive MMSE-DFE, can be used. Once every so often, PARAFAC can be invoked again to reopen the eye.

Reflecting on the form of the LS criterion in (13), we note that the multiple-parameter/multiple-invariance subspace fitting (SSF) formulation of ESPRIT [33], [34], [46], [47] (also see related work in [16] and [49]) is actually a special case of the mathematical statement of the problem considered in this paper when posed in eigenspace. In particular, given P subspace estimates $\hat{\mathbf{E}}_0, \ldots, \hat{\mathbf{E}}_{P-1}$, the aforementioned SSF problem is to find a square invertible matrix $\mathbf{T}$, unitary diagonal matrices $\mathbf{\Phi}_0, \ldots, \mathbf{\Phi}_{P-1}$, and a (sub)array response matrix $\mathbf{A}$, such that the following cost is minimized:

$$\left\| \begin{bmatrix} \hat{\mathbf{E}}_0 \\ \vdots \\ \hat{\mathbf{E}}_{P-1} \end{bmatrix} \mathbf{W}^{\frac{1}{2}} - \begin{bmatrix} \mathbf{A}\mathbf{\Phi}_0 \\ \vdots \\ \mathbf{A}\mathbf{\Phi}_{P-1} \end{bmatrix} \mathbf{T} \right\|_F^2 .$$

It follows that the identifiability results and trilinear ALS techniques presented herein carry over to such a SSF problem, and its many applications that include joint azimuth-elevation estimation, joint angle-delay estimation, and multidimensional harmonic retrieval; see [6] and [43]–[45].

Going back to the least squares formulation of PARAFAC in (13), we note that if we constrain the model by insisting that $\mathbf{S} = \mathbf{A}^H$, then a *joint approximate diagonalization* problem appears, and thus, the latter is a special case of PARAFAC. Our identifiability results of Section IV and fitting algorithms of Section V apply *a fortiori* to the joint approximate diagonalization problem as well. The latter is *the* basic problem in a variety of important papers dealing with source separation in the context of signal processing and communications, e.g., the analytical constant modulus algorithm (ACMA), and the JADE algorithm; see [41] for a review. These contributions have brought forward the significance of joint diagonalization as a core signal separation principle. As we have seen, PARAFAC is an even broader principle that encompasses these problems and answers some of the open questions in joint diagonalization, in particular, those related to identifiability with factors having dimensions unfit for eigenanalysis and/or factors that do lose rank.

VII. Simulation Results

A. Algorithm

In all of our simulations, we used the COMFAC algorithm [7], which is essentially a fast implementation of trilinear ALS. COMFAC speeds up the LS fitting procedure by working with a compressed version of the data, thereby avoiding brute-force implementation of ALS in the raw data space. The main steps of COMFAC are

i) compression;
ii) initialization and fitting of PARAFAC in compressed space;
iii) decompression and refinement in the raw data space.

These are outlined next.[5]

- As a first step, COMFAC compresses the $K \times N \times P$ three-way array $\underline{\mathbf{X}}$ into a (usually much) smaller $K' \times N' \times P'$ three-way array $\underline{\mathbf{G}}$. This is achieved by fitting a *Tucker3* model [40] to the raw data $\underline{\mathbf{X}}$ [4], yielding three bases that span the systematic variation in each of the three modes [23], [40]. Regressing the raw data onto these bases yields the compressed $K' \times N' \times P'$ array $\underline{\mathbf{G}}$. Using the CANDELINC theorem [4], [10], it can be shown that fitting the PARAFAC model to $\underline{\mathbf{X}}$, given the aforementioned bases, is equivalent to fitting PARAFAC to the (much smaller) $\underline{\mathbf{G}}$.
- Next, an M-component PARAFAC model is fitted to $\underline{\mathbf{G}}$. For initialization, we may use two-slab eigenanalysis (ESPRIT) or, better yet, a method known as *direct trilinear decomposition*, which generates two optimal pseudo-slabs from the available data and extracts initial estimates using eigenanalysis, provided the dimensions are appropriate; otherwise, a random initialization may be used. Given the relatively small size of $\underline{\mathbf{G}}$, ALS as well as other (e.g., Gauss–Newton using separable least squares, Levenberg-Marquardt, etc.) optimization techniques can be used in compressed space; a "universal" choice is yet not clear, and COMFAC actually employs a combination Gauss-Newton/ALS for this step, with ALS usually carrying most of the optimization burden.
- After fitting the model in the compressed space, the solution is decompressed by "post-multiplying" with the Tucker3 bases. This is followed by a few ALS steps in uncompressed space, which serve to guard for compression artifacts. Usually, the decompressed model is close

[5]Detailed description of these steps is omitted due to lack of space but can be accessed by downloading [7]

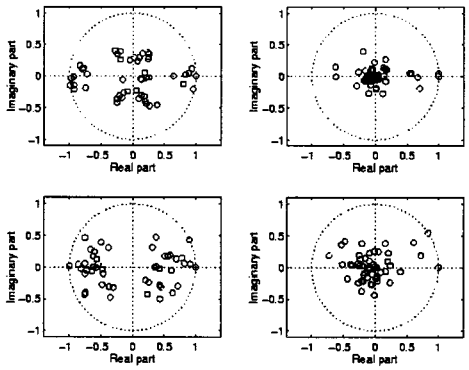

Fig. 5. Scatter diagrams for $M = 4$ users. Matched filter receiver.

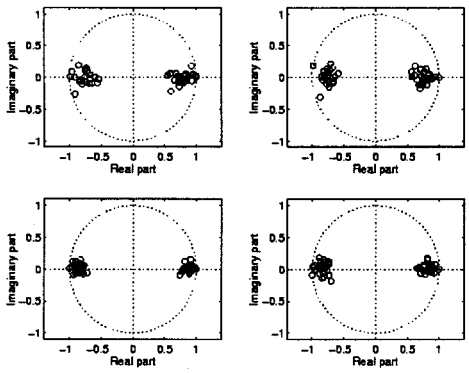

Fig. 6. Scatter diagrams for $M = 4$ users. Blind PARAFAC receiver.

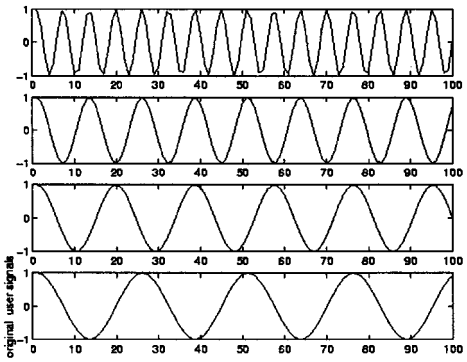

Fig. 7. Harmonics.

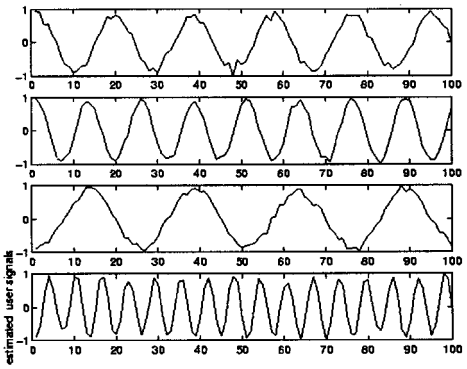

Fig. 8. PARAFAC-recovered harmonics.

to the LS solution, hence, <5 ALS steps are sufficient for this refinement stage.

As per Theorems 1 and 2, scale ambiguity is inherent to the separation problem without side information, and it manifests itself as a complex constant that multiplies each individual row of $\mathbf{S}$. For constant-modulus transmissions, this ambiguity can be removed via automatic gain control (AGC) and differential encoding (DE)/decoding [28, p. 187]. We assume differentially-encoded user signals throughout the simulations. For the purpose of performance evaluation only, the permutation ambiguity is resolved using a greedy least squares $(\mathbf{S}, \hat{\mathbf{S}})$-row matching algorithm.

B. Sample Runs

If $\underline{\mathbf{X}}$ is the noise-free data and $\underline{\tilde{\mathbf{X}}} = \underline{\mathbf{X}} + \underline{\mathbf{V}}$ is the noisy data, we define the sample SNR at the input of the multiuser receiver as

$$\text{SNR} = 10 \log_{10} \frac{||\underline{\mathbf{X}}||_F^2}{||\underline{\mathbf{V}}||_F^2} \text{ dB}$$

where the $||\underline{\mathbf{X}}||_F^2$ is the sum of squares of all elements of the 3-D array $\underline{\mathbf{X}}$, and *average SNR* (henceforth refered to as SNR) in the obvious fashion.[6]

Figs. 5–8 present the results of two illustrative simulation experiments. In our first experiment, we consider $M = 4$ differentially encoded BPSK (DE-BPSK) user signals, spread with Hadamard(4) codes and augmented by two trailing zeros, for a total of $P = 6$ chips/symbol (six ISI-free chips/symbol available). Each user undergoes a Rayleigh multipath/delay frequency selective channel of length $L + 1 = 3$ chips and is received at an SNR of 16 dB by $K = 2$ antennas. In addition to each user's own multipath/delay channel, each user-antenna pair experiences independent Rayleigh flat fading. Fig. 5 depicts scatter diagrams (at the input of the decision device) for the conventional Matched Filter receiver. Fig. 6 depicts the same for the proposed blind PARAFAC receiver. With $N = 50$ snapshots, the proposed deterministic blind receiver overwhelmingly outperforms the MF receiver. Note that the MF receiver exploits knowledge of the spreading codes and averages the independently faded signal copies received by the two antennas. We underscore that the proposed receiver does not utilize spreading code information.

To illustrate the waveform-preserving character of our PARAFAC receiver and using the same transmit/channel parameters, in our second experiment, we consider harmonic sources. At the receiver, we employ $K = 4$ antennas and collect $N = 100$ snapshots. SNR at the input of the multiuser receiver is 10 dB. Fig. 7 depicts the true source signals, whereas Fig. 8 depicts the PARAFAC-recovered signals. Observe the shuffling among waveforms due to the permutation ambiguity predicted by Theorem 2.

C. Monte Carlo Results—Comparison With Nonblind ZF/MMSE

In this section, we present Monte Carlo simulations that are designed to assess the bit/symbol error rate (BER/SER) performance of the proposed blind PARAFAC receiver. We compare

[6]This (and not E_b/N_o) is the pertinent measure of SNR because E_b/N_o does not take into account the effects of multiple antennas/fading/multipath.

against the nonblind zero-forcing (ZF also known as decorrelating) and nonblind minimum mean-squared error (MMSE) receivers. The latter offers a performance bound against which blind algorithms are often measured [14], [39]; therefore, it is a natural benchmark. As a line of reference, we also include results for the conventional matched filter (MF) receiver, although it clearly exhibits an error floor due to the presence of multipath and the ensuing loss of orthogonality that causes significant multiuser interference (MUI).

In contrast to our blind PARAFAC receiver, the nonblind ZF/MMSE receivers assume perfect knowledge of multipath/delay/fading and spreading codes. In addition, the nonblind MMSE receiver assumes knowledge of the SNR. Let us rewrite the (unfolded) polyphase decomposition of the data in (12) in the compact form as

$$\mathcal{X}_{KP\times N} = \mathcal{A}_{KP\times M}\mathbf{S}.$$

With noise present, $\mathcal{X}$ becomes $\tilde{\mathcal{X}}$, whereas $\mathcal{A}$ is assumed to be perfectly known in a nonblind setting. It follows that the ZF solution for $\mathbf{S}$ is

$$\hat{\mathbf{S}}_{\text{ZF}} = \mathcal{A}^{\dagger}\tilde{\mathcal{X}} = \begin{bmatrix} \mathbf{A}\mathbf{D}_0(\mathbf{H}) \\ \vdots \\ \mathbf{A}\mathbf{D}_{P-1}(\mathbf{H}) \end{bmatrix}^{\dagger} \begin{bmatrix} \tilde{\mathbf{X}}_0 \\ \vdots \\ \tilde{\mathbf{X}}_{P-1} \end{bmatrix}$$

where $\tilde{\mathbf{X}}_p$ denotes the noisy slabs. Similarly, the nonblind MMSE solution is a regularized inverse given by

$$\hat{\mathbf{S}}_{\text{MMSE}} = \left(\mathcal{A}^H\mathcal{A} + \frac{1}{\text{SNR}}\mathbf{I}\right)^{-1}\mathcal{A}^H\tilde{\mathcal{X}}.$$

The MF receiver exploits knowledge of the spreading codes and takes advantage of antenna diversity by combining independently faded array outputs. On the average, and for a typical simulation setup like the ones below, PARAFAC takes a few seconds to run in interpreted MATLAB on a SUN ULTRA-1.

Recall that

- M number of users;
- N number of snapshots;
- P number of ISI-free chips per symbol;
- K number of antennas.

Throughout, $\mathcal{MC}$ =number of Monte Carlo trials = 1000. Note that the effective BER averaging is of order $\mathcal{MC} \times N$ per user with N varying from experiment to experiment. For all users and all experiments, $h_m(l)$ is of order $L = 2 \rightarrow L+1 = 3-$chip multipath/delay channels. $L_g = 2$ trailing zeros ("guard chips") are used for blocking, and $P = M + L_g$ ISI-free chips are available. For each Monte Carlo run, the multipath/delay channel taps are redrawn from an i.i.d. Rayleigh generator and are thus the fading coefficients. User signals are redrawn from an i.i.d. distribution and then differentially encoded for each run. Note that the channel impulse responses are not normalized to unit norm; this, along with the presence of fading, means that the effective signal power varies considerably from run to run and from user to user, which is a challenging setup when power control is absent. No power control is assumed. The results are reported as average BER/SER for all users versus average SNR.

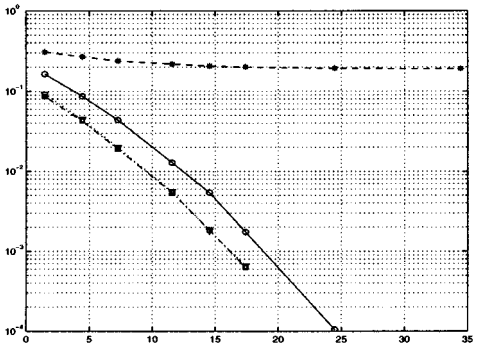

Fig. 9. BER versus SNR. $M = 4, P = 6, N = 50, K = 2$ DE-BPSK.

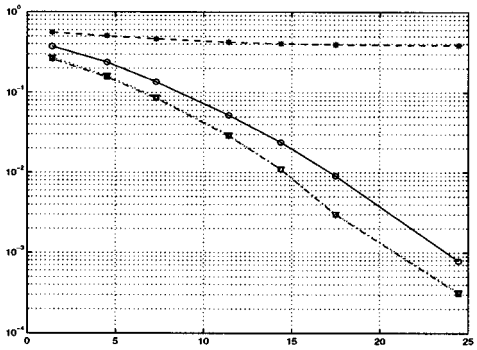

Fig. 10. SER versus SNR. $M = 4, P = 6, N = 50, K = 2$ DE-QPSK.

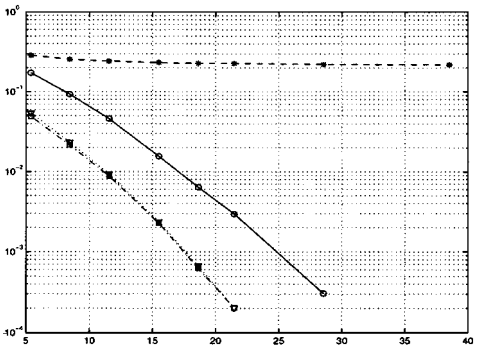

Fig. 11. BER versus SNR:. $M = 8, P = 10, N = 50, K = 2$ DE-BPSK.

Averaging is performed over the statistics of fading, multipath, user signals, and (AWG) noise.

Fig. 9 depicts results for $M = 4, P = 6, N = 50, K = 2$, and DE-BPSK transmissions. The blind PARAFAC receiver is depicted using solid line with circles, MF using dashed line with stars, nonblind MMSE using dash-dot with squares, and nonblind ZF using double-dot with triangles (the ZF and MMSE receivers appear almost identical in the log-log plot, especially for high SNR, as expected). Fig. 10 presents the counterpart of Fig. 9 but for DE-QPSK. Notice that in both cases, the blind PARAFAC receiver is very close to the nonblind MMSE receiver.

Figs 11 and 12 present results for $M = 8, P = 10, N = 10, K = 2$ DE-BPSK and $M = 16, P = 18, N = 10, K = 2$ DE-BPSK, respectively. Note that the gap between blind PARAFAC and (nonblind) MMSE increases as M increases. This is to be expected, of course, and it can be easily compensated by increasing K from the current (minimum possible)

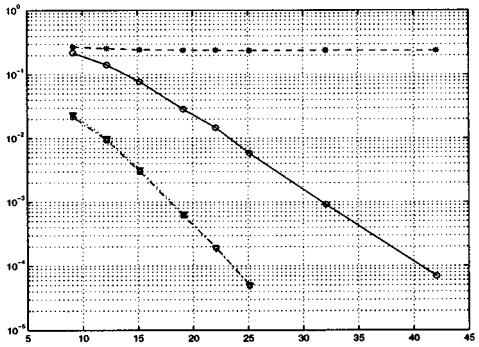

Fig. 12. BER versus SNR. $M = 16$, $P = 18$, $N = 50$, $K = 2$ DE-BPSK. Increasing gap with increasing M.

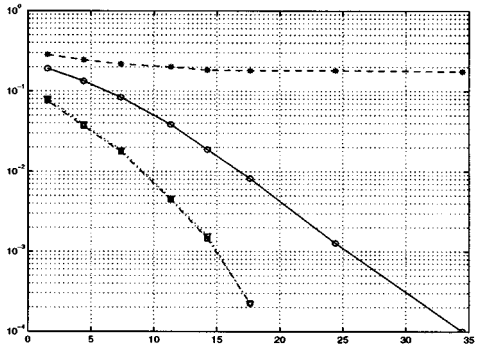

Fig. 13. Works with small sample sizes. BER versus SNR. $M = 4$, $P = 6$, $N = 10$, $K = 2$ DE-BPSK.

value of 2 to, e.g., 4 or, alternatively, by increasing some other form of diversity (P or N).

Fig. 13 shows small sample results: $M = 4, P = 6, N = 10$, $K = 2$ DE-BPSK. It is clear that PARAFAC performs well even for very small sample sizes.

Remark 2—Regarding the Case $M > P$ (More Users than Spreading): Observe that due to symmetry, conclusions for this case can be drawn from the above simulation results, noting that the role of antenna diversity $(2 < M)$ can be switched with spreading diversity $(P \geq M)$. In addition, note that due to the presence of frequency-selective multipath, no two users will share the same signature code ("effective" spreading code), even if they start out with the same spreading code on transmit. If $M > P$, then what happens is that *all* signature codes become correlated, albeit at (randomly) varying degrees, since the multipath channels (and hence the signature codes) are redrawn for each Monte Carlo trial. PARAFAC works with these "effective" spreading codes, and therefore , it does not make sense to single out any subset of users; average BER for all users is still pertinent.

VIII. Reprise

This paper has developed a link between PARAFAC analysis and blind multiuser separation-equalization-detection for DS-CDMA systems. Relying on the uniqueness of low-rank three-way array decomposition and trilinear alternating least squares regression, a deterministic blind PARAFAC receiver has been proposed that offers some unique advantages: notably blind performance close to nonblind MMSE; identifiability (consistency as SNR $\rightarrow \infty$); and deterministic joint LS/conditional ML estimates of signals, array response, and effective codes.

Unlike existing approaches, PARAFAC requires no statistical independence or whiteness in space or time, and it does not rely on known codes, DOA-calibration information, or FA/CM properties to achieve separation, although it can capitalize on any of these, if available. PARAFAC works well for small sample sizes by virtue of its deterministic nature, and it is robust with respect to near-far effects (LS/ML). Interestingly, PARAFAC can recover the users even if less spreading than users is available, which is an important advantage in oversaturated systems, or whenever bandwidth comes at a premium. This is achieved by trading spreading diversity for spatio-temporal diversity. Rapidly varying frequency-flat fading, as well as unknown frequency-selective multipath, can be handled, provided that the multipath reflectors are in the far-field, the system is quasisynchronous so the delay spread is in the order of a few chips, and sufficient diversity is available for identifiability in the sense of the k-rank condition (5)–(8).

The ideas presented herein also apply to non-CDMA systems where fractional sampling diversity is available (not necessarily induced by spreading). As shown recently in [29], TDMA, orthogonal frequency division multiplexing/multiple access (OFDMA), and discrete multitone (DMT) can be thought of as special cases of the CDMA model considered herein. For example, in OFDMA, the codes are complex exponentials. It follows that our results apply to these schemes as well. These are of interest in their own right, and will be expanded upon in a sequel to this paper. Multi-way arrays, multilinear models, and the trilinear model in particular have many additional applications in signal processing and communications that are currently under investigation.

Appendix

Proof of Theorem 2: We will use the following result, which is of interest in its own right, and is stated and proved separately (without reference to uniqueness of triple product decomposition) in Kruskal [24]. Kruskal proves it (in $\approx$ four pages) under the working assumption of real matrices, but in contrast with the uniqueness of triple product decomposition results, it only involves rank (instead of k-rank) and span arguments for a pair of matrices, and it readily generalizes to the complex case. Like Kruskal, we will use the result for the "last mile," i.e., to finish the line of argument in the main proof.

Lemma 1 (Permutation Lemma): Let $w(\mathbf{v})$ denote the number of nonzero elements of $\mathbf{v} \in \mathbb{C}^P$. Given two matrices $\mathbf{H}$ and $\bar{\mathbf{H}}$ with the same number of columns (M), suppose that $\mathbf{H}$ has no identically zero columns, and assume that the following implication holds:

$$w(\mathbf{v}^T\bar{\mathbf{H}}) \leq M - \text{rank}(\bar{\mathbf{H}}) + 1 \Rightarrow w(\mathbf{v}^T\mathbf{H}) \leq w(\mathbf{v}^T\bar{\mathbf{H}}).$$

We then have that $\bar{\mathbf{H}} = \mathbf{H}\mathbf{\Pi}\mathbf{\Lambda}$, where $\mathbf{\Pi}$ is a permutation matrix, and $\mathbf{\Lambda}$ is a nonsingular complex diagonal scaling matrix.

Main Proof: Let us first consider uniqueness of $\mathbf{H}$. The rest will follow by the complete symmetry of both the complex trilinear model and the k-rank condition. Suppose there also exist

$\bar{\mathbf{A}} \in \mathbb{C}^{K\times M}$, $\bar{\mathbf{H}} \in \mathbb{C}^{P\times M}$, $\bar{\mathbf{S}} \in \mathbb{C}^{M\times N}$, such that $\mathbf{X}_p = \mathbf{A}\mathbf{D}_p(\mathbf{H})\mathbf{S} = \bar{\mathbf{A}}\mathbf{D}_p(\bar{\mathbf{H}})\bar{\mathbf{S}}$, for $p = 0, 1, \ldots, P-1$. We wish to show that

$$\begin{aligned}&\textbf{IF: } w(\mathbf{v}^T\bar{\mathbf{H}}) \le M - \text{rank}(\bar{\mathbf{H}}) + 1,\\ &\textbf{THEN: } w(\mathbf{v}^T\mathbf{H}) \le w(\mathbf{v}^T\bar{\mathbf{H}})\end{aligned}$$

for all $\mathbf{v} \in \mathbb{C}^P$. Taking linear combinations $\sum_{p=0}^{P-1} v_p\mathbf{X}_p$, it follows that

$$\mathbf{A}\,\text{diag}(\mathbf{v}^T\mathbf{H})\mathbf{S} = \bar{\mathbf{A}}\,\text{diag}(\mathbf{v}^T\bar{\mathbf{H}})\bar{\mathbf{S}}, \quad \forall \mathbf{v} := [v_0, \ldots, v_{P-1}]^T \in \mathbb{C}^P. \quad (14)$$

Because the rank of a matrix product is always less than or equal to the rank of any factor, we have

$$w(\mathbf{v}^T\bar{\mathbf{H}}) = \text{rank}(\text{diag}(\mathbf{v}^T\bar{\mathbf{H}})) \ge \text{rank}(\bar{\mathbf{A}}\,\text{diag}(\mathbf{v}^T\bar{\mathbf{H}})\bar{\mathbf{S}}) = \text{rank}(\mathbf{A}\,\text{diag}(\mathbf{v}^T\mathbf{H})\mathbf{S}). \quad (15)$$

Let $\gamma := w(\mathbf{v}^T\mathbf{H})$ be the number of nonzero elements in $\mathbf{v}^T\mathbf{H}$. Drop those columns of $\mathbf{A}$ and associated rows of $\mathbf{S}$ corresponding to the zeros of $\mathbf{v}^T\mathbf{H}$; this results in truncated matrices $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{S}}$ having γ columns and rows, respectively. Let $\mathbf{t} \in \mathbb{C}^{\gamma\times 1}$ be the corresponding nonzero part of $\mathbf{v}^T\mathbf{H}$. Then, relying on Sylvester's inequality, we obtain

$$\begin{aligned}&\text{rank}(\mathbf{A}\,\text{diag}(\mathbf{v}^T\mathbf{H})\mathbf{S})\\ &\quad = \text{rank}(\tilde{\mathbf{A}}\,\text{diag}(\mathbf{t})\tilde{\mathbf{S}})\\ &\quad \ge \text{rank}(\tilde{\mathbf{A}}) + \text{rank}((\text{diag}(\mathbf{t})\tilde{\mathbf{S}})^T) - \gamma. \end{aligned} \quad (16)$$

Since all elements of $\mathbf{t}$ are nonzero, it follows that

$$\begin{aligned}&\text{rank}(\tilde{\mathbf{A}}) + \text{rank}((\text{diag}(\mathbf{t})\tilde{\mathbf{S}})^T) - \gamma\\ &\quad = \text{rank}(\tilde{\mathbf{A}}) + \text{rank}(\tilde{\mathbf{S}}) - \gamma. \end{aligned} \quad (17)$$

Now, $\tilde{\mathbf{A}}$ has γ columns of the original $\mathbf{A}$, and $\tilde{\mathbf{S}}$ is constructed out of γ rows of $\mathbf{S}$; by the definition of k-rank, it holds that

$$\text{rank}(\tilde{\mathbf{A}}) \ge \min(\gamma, k_{\mathbf{A}}), \quad \text{rank}(\tilde{\mathbf{S}}) \ge \min(\gamma, k_{\mathbf{S}^T}). \quad (18)$$

Thus, (15)–(18) allow us to infer that

$$w(\mathbf{v}^T\bar{\mathbf{H}}) \ge \min(\gamma, k_{\mathbf{A}}) + \min(\gamma, k_{\mathbf{S}^T}) - \gamma. \quad (19)$$

For different ranges of γ, (19) implies (20), shown at the bottom of the page. This is the first leg of the main proof. Now, observe that in order to establish the implication required by the *permutation lemma*, it suffices to show that the k-rank condition (5) excludes the latter two ($\dagger, \bullet$) possibilities. We will argue by contradiction.

We start by proving that (5) implies $\text{rank}(\mathbf{H}) \le \text{rank}(\bar{\mathbf{H}})$. From (14), it follows that

$$\mathbf{v}^T\bar{\mathbf{H}} = \mathbf{0}_{1\times M}^T \Rightarrow \mathbf{A}\,\text{diag}(\mathbf{v}^T\mathbf{H})\mathbf{S} = \mathbf{0}_{K\times N}.$$

Consider again $w(\mathbf{v}^T\mathbf{H})$, which is the number of nonzero elements in $\mathbf{v}^T\mathbf{H}$. We will show that $w(\mathbf{v}^T\mathbf{H}) = 0$. Suppose the opposite, namely, that $1 \le w(\mathbf{v}^T\mathbf{H}) \le M$. As part of the derivation leading to (20), we see that (21), shown at the bottom of the page, holds. Since $k_{\mathbf{S}^T} \le M$, from (5)

$$\begin{aligned}&k_{\mathbf{A}} + k_{\mathbf{H}} + k_{\mathbf{S}^T} \ge 2(M+1)\\ &\quad \Rightarrow \begin{cases} k_{\mathbf{A}} + k_{\mathbf{S}^T} \ge M + 2 > M + 1\\ \min(k_{\mathbf{A}}, k_{\mathbf{S}^T}) \ge 2. \end{cases}\end{aligned}$$

Equation (21) then shows that $\text{rank}(\mathbf{A}\,\text{diag}(\mathbf{v}^T\mathbf{H})\mathbf{S}) \ge 1$ whenever $1 \le w(\mathbf{v}^T\mathbf{H}) \le M$. However, $\mathbf{A}\,\text{diag}(\mathbf{v}^T\mathbf{H})\mathbf{S} = \mathbf{0}_{K\times N}$, and therefore, its rank should be zero. This shows that

$$\mathbf{v}^T\bar{\mathbf{H}} = \mathbf{0}_{1\times M} \Rightarrow \mathbf{v}^T\mathbf{H} = \mathbf{0}_{1\times M} \Rightarrow \text{rank}(\mathbf{H}) \le \text{rank}(\bar{\mathbf{H}}).$$

Recall the assumption of the *permutation lemma* and work from the second line and to the left:

$$\begin{aligned}M - k_{\mathbf{H}} + 1 &\ge M - \text{rank}(\mathbf{H}) + 1\\ &\ge M - \text{rank}(\bar{\mathbf{H}}) + 1 \ge w(\mathbf{v}^T\bar{\mathbf{H}}) \end{aligned} \quad (22)$$

where the top left-most inequality follows from definition of k-rank ($\le$rank). Here is now another point where the k-rank condition plays a key role:

$$\begin{aligned}k_{\mathbf{A}} + k_{\mathbf{H}} + k_{\mathbf{S}^T} \ge 2(M+1) &\Rightarrow k_{\mathbf{A}} + k_{\mathbf{S}^T} - M - 1\\ &\ge M - k_{\mathbf{H}} + 1. \end{aligned} \quad (23)$$

Combining (22) and (23), we arrive at the second key inequality

$$k_{\mathbf{A}} + k_{\mathbf{S}^T} - M - 1 \ge w(\mathbf{v}^T\bar{\mathbf{H}}). \quad (24)$$

Consider now inequalities (20) and (24) jointly. Suppose that the third leg ($\bullet$) of (20) is in effect, that is, $w(\mathbf{v}^T\mathbf{H}) \ge \max(k_{\mathbf{A}}, k_{\mathbf{S}^T})$; then, we obtain

$$k_{\mathbf{A}} + k_{\mathbf{S}^T} - M - 1 \ge w(\mathbf{v}^T\bar{\mathbf{H}}) \ge k_{\mathbf{A}} + k_{\mathbf{S}^T} - w(\mathbf{v}^T\mathbf{H})$$

$$w(\mathbf{v}^T\bar{\mathbf{H}}) \ge \begin{cases} w(\mathbf{v}^T\mathbf{H}), \text{ for } w(\mathbf{v}^T\mathbf{H}) \le \min(k_{\mathbf{A}}, k_{\mathbf{S}^T}) & (*)\\ \min(k_{\mathbf{A}}, k_{\mathbf{S}^T}), \text{ for } \min(k_{\mathbf{A}}, k_{\mathbf{S}^T}) \le w(\mathbf{v}^T\mathbf{H}) \le \max(k_{\mathbf{A}}, k_{\mathbf{S}^T}) & (\dagger)\\ k_{\mathbf{A}} + k_{\mathbf{S}^T} - w(\mathbf{v}^T\mathbf{H}), \text{ for } w(\mathbf{v}^T\mathbf{H}) \ge \max(k_{\mathbf{A}}, k_{\mathbf{S}^T}) & (\bullet). \end{cases} \quad (20)$$

$$\text{rank}(\mathbf{A}\text{diag}(\mathbf{v}^T\mathbf{H})\mathbf{S}) \ge \begin{cases} w(\mathbf{v}^T\mathbf{H}), \text{ for } w(\mathbf{v}^T\mathbf{H}) \le \min(k_{\mathbf{A}}, k_{\mathbf{S}^T})\\ \min(k_{\mathbf{A}}, k_{\mathbf{S}^T}), \text{ for } \min(k_{\mathbf{A}}, k_{\mathbf{S}^T}) \le w(\mathbf{v}^T\mathbf{H}) \le \max(k_{\mathbf{A}}, k_{\mathbf{S}^T})\\ k_{\mathbf{A}} + k_{\mathbf{S}^T} - w(\mathbf{v}^T\mathbf{H}), \text{ for } w(\mathbf{v}^T\mathbf{H}) \ge \max(k_{\mathbf{A}}, k_{\mathbf{S}^T}). \end{cases} \quad (21)$$

which is impossible because $w(\mathbf{v}^T\mathbf{H}) \leq M$. Similarly, suppose that the second leg (†) of (20) is in effect, that is, $\min(k_{\mathbf{A}}, k_{\mathbf{S}^T}) \leq w(\mathbf{v}^T\mathbf{H}) \leq \max(k_{\mathbf{A}}, k_{\mathbf{S}^T})$; then, we obtain

$$k_{\mathbf{A}} + k_{\mathbf{S}^T} - M - 1 \geq w(\mathbf{v}^T\bar{\mathbf{H}}) \geq \min(k_{\mathbf{A}}, k_{\mathbf{S}^T})$$

which is impossible since

$$\begin{aligned} k_{\mathbf{A}} + k_{\mathbf{S}^T} - M - 1 &= \min(k_{\mathbf{A}}, k_{\mathbf{S}^T}) \\ &\quad + \max(k_{\mathbf{A}}, k_{\mathbf{S}^T}) - M - 1 \leq \min(k_{\mathbf{A}}, k_{\mathbf{S}^T}) - 1 \end{aligned}$$

since $\max(k_{\mathbf{A}}, k_{\mathbf{S}^T}) \leq M$. The only remaining option is leg $(*)$, i.e., $w(\mathbf{v}^T\mathbf{H}) \leq \min(k_{\mathbf{A}}, k_{\mathbf{S}^T})$, and thus

$$w(\mathbf{v}^T\bar{\mathbf{H}}) \geq w(\mathbf{v}^T\mathbf{H})$$

which is the condition required by the *permutation lemma* (in addition, note the trivial fact that $k_{\mathbf{A}}+k_{\mathbf{H}}+k_{\mathbf{S}^T} \geq 2(M+1) \Rightarrow \min(k_{\mathbf{A}}, k_{\mathbf{H}}, k_{\mathbf{S}^T}) \geq 2$, which implies that none of $\mathbf{A}, \mathbf{H}, \mathbf{S}^T$ has an all-zero column). This shows that $\bar{\mathbf{H}} = \mathbf{H}\mathbf{\Pi}\mathbf{\Lambda}$, where $\mathbf{\Pi}$ is a permutation matrix, and $\mathbf{\Lambda}$ is a nonsingular complex diagonal scaling matrix. Since the trilinear model is completely symmetric ($\Rightarrow$ any one of the three matrices can be put in the middle of the decomposition), and the k-rank condition is also symmetric (*sum* of k-ranks), it follows that $\mathbf{A}$ and $\mathbf{S}$ are also unique modulo permutation and scale. This completes the proof. □

Remark 3: It can be shown that the permutation is common to all three matrices, and the product of the respective scales is identity. This is not critical in the DS-CDMA context, and we therefore skip it for space considerations.

ACKNOWLEDGMENT

The authors wish to acknowledge the influential work of Dr. J. Kruskal of the Mathematical Sciences Center at Bell Labs. N. Sidiropoulos further acknowledges personal communication (and a revised copy of [24] including clarifications) with Dr. Kruskal.

REFERENCES

[1] B. G. Agee, S. V. Schell, and W. A. Gardner, "Spectral self-coherence restoral: A new approach to blind adaptive signal extraction using antenna arrays," *Proc. IEEE*, vol. 78, pp. 753–767, Apr. 1990.

[2] S. E. Bensley and B. Aazhang, "Subspace-based channel estimation for code division multiple access communication systems," *IEEE Trans. Commun.*, vol. 44, pp. 1009–1020, Aug. 1996.

[3] R. Bro, "PARAFAC: Tutorial and applications," *Chemometr. Intell. Lab. Syst.*, vol. 38, pp. 149–171, 1997.

[4] R. Bro and C. A. Andersson, "Improving the speed of multi-way algorithms. Part II: Compression," *Chemometr. Intell. Lab. Syst.*, to be published.

[5] R. Bro and N. D. Sidiropoulos, "Least squares regression under unimodality and nonnegativity constraints," *J. Chemometr.*, vol. 12, pp. 223–247, 1998.

[6] R. Bro, N. D. Sidiropoulos, and G. B. Giannakis, "Optimal joint Azimuth-elevation and signal-array response estimation using parallel factor analysis," presented at the Proc. 32nd Asilomar Conf. Signals, Syst., Comput., Monterey, CA, Nov. 1998.

[7] ——, "A fast least squares algorithm for separating trilinear mixtures," presented at the Proc. Int. Workshop Independent Component Anal. Blind Signal Separation, Aussois, France, Jan. 11–15, 1999.

[8] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *Proc. Inst. Elect. Eng. F*, vol. 140, pp. 362–370, 1993.

[9] J. D. Carroll 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.

[10] J. D. Carroll, S. Pruzansky, and J. B. Kruskal, "CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters," *Psychometrika*, vol. 45, pp. 3–24, 1980.

[11] R. B. Cattell, "Parallel proportional profiles and other principles for determining the choice of factors by rotation," *Psychometrika*, no. 9, p. 267, 1944.

[12] P. Comon, "Independent component analysis: A new concept," *Signal Process.*, vol. 36, pp. 287–314, 1994.

[13] P. Comon, "Blind channel identification and extraction of more sources than sensors," in *Proc. SPIE*, San Diego, CA, July 19–24, 1998.

[14] D. Gesbert, J. Sorelius, and A. Paulraj, "Blind multi-user MMSE detection of CDMA signals," in *Proc. of Int. Conf. Acoust., Sppech, Signal Process.*, May 1998, pp. 3161–3164.

[15] G. B. Giannakis, Y. Inouye, and J. M. Mendel, "Cumulant based identification of multichannel moving-average models," *IEEE Trans. Automat. Contr.*, vol. 34, pp. 783–787, July 1989.

[16] M. Haardt, M. D. Zoltowski, C. P. Mathews, and J. A. Nossek, "2D unitary ESPRIT for efficient 2D parameter estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, Detroit, MI, May 8–12, 1995, pp. 2096–2099.

[17] M. L. Honig, U. Madhow, and S. Verdú, "Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, pp. 944–996, July 1995.

[18] R. A. Harshman, "Foundations of the PARAFAC procedure: Model and conditions for an 'explanatory' multi-mode factor analysis," *UCLA Working Papers Phonetics*, vol. 16, pp. 1–84, Dec. 1970.

[19] R. A. Harshman and M. E. Lundy, "The PARAFAC model for three-way factor analysis and multidimensional scaling," in *Research Methods for Multimode Data Analysis*, H. G. Law, C. W. Snyder Jr., J. Hattie, and R. P. McDonald, Eds. New York: Praeger, 1984, pp. 122–215.

[20] Y. Hua and T. K. Sarkar, "On SVD for estimating generalized eigenvalues of singular matrix pencil in noise," *IEEE Trans. Signal Processing*, vol. 39, pp. 892–900, Apr. 1991.

[21] R. A. Iltis and L. Mailaender, "Multiuser detection of quasisynchronous CDMA signals using linear decorrelators," *IEEE Trans. Commun.*, pp. 1561–1571, Nov. 1996.

[22] H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Signal Processing Mag.*, pp. 67–95, July 1996.

[23] P. M. Kroonenberg, *Three-Mode Principal Component Analysis. Theory and Applications*. Leiden, Germany: DSWO, 1983.

[24] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Applicat.*, vol. 18, pp. 95–138, 1977.

[25] J. B. Kruskal, "Rank decomposition, and uniqueness for 3-way and N-way arrays," in *Multi-way Data Analysis*, R. Coppi and S. Bolasco, Eds. Amsterdam, The Netherlands: North-Holland, 1988, pp. 7–18.

[26] L. de Lathauwer, "Signal processing based on multilinear algebra," Ph.D. dissertation, Katholieke Univ., Leuven, Belgium, Sept. 1997.

[27] H. Liu and G. Xu, "A subspace method for signature waveform estimation in synchronous CDMA systems," *IEEE Trans. Commun.*, pp. 1346–1354, 1996.

[28] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[29] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers Part I: Unification and optimal designs," *IEEE Trans. Signal Processing*, vol. 47, pp. 1988–2006, July 1999.

[30] P. H. Schonemann, "An algebraic solution for a class of subjective metrics models," *Psychometrika*, vol. 37, p. 441, 1972.

[31] S. Shamsunder and G. B. Giannakis, "Modeling of nongaussian array data using cumulants: DOA estimation of more sources with less sensors," *Signal Process.*, vol. 30, no. 3, pp. 279–297, Feb. 1993.

[32] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Deterministic waveform-preserving blind separation of DS-CDMA signals using an antenna array," in *Proc. IEEE Signal Processing Workshop Stat. Signal Array Process.*, Portland, OR, Sept. 14–16, 1998, pp. 304–307.

[33] A. Swindlehurst and T. Kailath, "Algorithms for Azimuth/elevation direction finding using regular array geometries," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, pp. 145–156, Jan. 1993.

[34] A. Swindlehurst, B. Ottersten, R. Roy, and T. Kailath, "Multiple invariance ESPRIT," *IEEE Trans. Signal Processing*, vol. 40, pp. 868–881, Apr. 1992.

[35] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array—Part I: Algorithms," *IEEE Trans. Signal Processing*, vol. 44, pp. 1184–1197, May 1996.

[36] L. Tong, Y. Inouye, and R.-W. Liu, "Waveform-preserving blind estimation of multiple independent sources," *IEEE Trans. Signal Processing*, vol. 41, pp. 2461–2470, 1993.
[37] M. K. Tsatsanis, "Inverse filtering criteria for CDMA systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 102–112, Jan. 1997.
[38] M. K. Tsatsanis and G. B. Giannakis, "Multirate filter banks for code-division multiple access systems," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, Detroit, MI, May 8–12, 1995, pp. 1484–1487.
[39] M. K. Tsatsanis and Z. Xu, "Performance analysis of minimum variance CDMA receivers," *IEEE Trans. Signal Processing*, vol. 46, pp. 3014–3022, Nov. 1998.
[40] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, pp. 279–311, 1966.
[41] A.-J. van der Veen, "Algebraic methods for deterministic blind beamforming," *Proc. IEEE*, vol. 86, pp. 1987–2008, Oct. 1998.
[42] A.-J. van der Veen and A. Paulraj, "An analytical constant-modulus algorithm," *IEEE Trans. Signal Processing*, vol. 44, pp. 1136–1155, May 1996.
[43] A.-J. van der Veen, M. C. Vanderveen, and A. J. Paulraj, "Joint angle and delay estimation using shift-invariance properties," *IEEE Signal Processing Lett.*, vol. 4, pp. 142–145, May 1997.
[44] ——, "Joint angle and delay estimation using shift-invariance techniques," *IEEE Trans. Signal Processing*, vol. 46, pp. 405–418, Feb. 1998.
[45] ——, "Estimation of multipath parameters in wireless communications," *IEEE Trans. Signal Processing*, vol. 46, pp. 682–690, Mar. 1998.
[46] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, pp. 1110–1121, May 1991.
[47] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, pp. 2436–2449, Nov. 1991.
[48] X. Wang and H. V. Poor, "Blind multiuser detection: A subspace approach," *IEEE Trans. Inform. Theory*, vol. 42, pp. 677–690, Mar. 1998.
[49] K. T. Wong and M. D. Zoltowski, "Closed-form multi-dimensional multi-invariance ESPRIT," in *Proc. IEEE Int. Conf Acoust., Speech, Signal Process.*, vol. 5, Munich, Germany, Apr. 21–24, 1997, pp. 3489–3492.
[50] M. Zoltowski, "Solving the generalized eigenvalue problem with singular forms," *Proc. IEEE*, vol. 75, pp. 1546–1548, Nov. 1987.

Nicholas D. Sidiropoulos (M'92–SM'99) received the Diploma in electrical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 1988, 1990, and 1992, respectively.

From 1988 to 1992, he was a Fulbright Fellow and a Research Assistant at the Institute for Systems Research (ISR), University of Maryland. From September 1992 to June 1994, he served his military service as a Lecturer in the Hellenic Air Force Academy. From October 1993 to June 1994, he also was a Member of the Technical Staff, Systems Integration Division, G-Systems Ltd., Athens, Greece. He held Postdoctoral (from 1994 to 1995) and Research Scientist (from 1996 to 1997) positions at ISR before joining the Department of Electrical Engineering at the University of Virginia, Charlottesville, in July 1997 as an Assistant Professor. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis. His current research interests are primarily in multi-way analysis and its applications in signal processing for communications and networking.

Dr. Sidiropoulos received the NSF/CAREER award (Signal Processing Systems Program) in June 1998.

Georgios B. Giannakis (F'96) received the Diploma in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1981. From September 1982 to July 1986, he was with the University of Southern California (USC), Los Angeles, where he received the M.Sc. degree in electrical engineering in 1983, the M.Sc. degree in mathematics in 1986, and the Ph.D. degree in electrical engineering in 1986.

After lecturing for one year at USC, he joined the University of Virginia (UVA), Charlottesville, in 1987, where he became a Professor of Electrical Engineering in 1997, and the Graduate Committee Chair and Director of the Communications, Controls, and Signal Processing Laboratory in 1998. Since January 1999 he has been with the University of Minnesota as a professor of Electrical and Computer Engineering. His general interests lie in the areas of signal processing and communications, estimation and detection theory, time-series analysis, and system identification—subjects on which he has published more than 95 journal papers. Specific areas of expertise have included (poly)spectral analysis, wavelets, cyclostationary, and non-Gaussian signal processing with applications to sensor array and image processing. Current research topics focus on transmitter and receiver diversity techniques for equalization of single- and multi-user communication channels, mitigation of rapidly fading wireless channels, compensation of nonlinear amplifier effects, redundant filterbank transceivers for block transmissions, multicarrier, and wideband communication systems.

Dr. Giannakis was awarded the School of Engineering and Applied Science Junior Faculty Research Award from UVA in 1988 and the UVA-EE Outstanding Faculty Teaching Award in 1992. He received the IEEE Signal Processing Society's 1992 Paper Award in the Statistical Signal and Array Processing area and co-authored the 1999 Best Paper Award by Young Author (with M. K. Tsatsanis). He co-organized the 1993 IEEE Signal Processing Workshop on Higher Order Statistics, the 1996 IEEE Workshop on Statistical Signal and Array Processing, and the first IEEE Signal Processing Workshop on Wireless Communications in 1997. He guest (co-)edited two Special Issues on high-order statistics (*International Journal of Adaptive Control and Signal Processing*, and the EURASIP journal *Signal Processing*) and the January 1997 Special Issue on Signal Processing for Advanced Communications of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS, a Secretary of the Signal Processing Conference Board, a Member of the SP Publications Board, and a member and Vice-Chair of the SSAP Technical Committee. He now chairs the Signal Processing for Communications Technical Committee and serves as Editor-in-Chief of the IEEE SIGNAL PROCESSING LETTERS. He is also a Member of the IMS and the European Association for Signal Processing.

Rasmus Bro received the M.Sc. degree in analytical and medical chemistry from the Technical University, Lyngby, Denmark, in 1994. He received the Ph.D. degree in chemometrics from the University of Amsterdam, The Netherlands, in 1998.

From 1994 to 1998, he was first a Research Assistant and then Assistant Professor in the areas of chemometrics and data analysis at the Chemometrics Group, Food Technology, Department of Food and Dairy Science, Royal Veterinary and Agricultural University, Frederiksberg, Denmark. In the Spring of 1995, he spent two months in the Process Analysis Group, Department of Analytical Chemistry, University of Amsterdam. Currently, he is a Research Associate Professor at the Royal Veterinary and Agricultural University with the Chemometrics group. From October 1997 to March 1998, he worked intermediately as a Guest rResearcher at MATFORSK (The Food Research Institute), Aas, Norway. His main research interests are mathematical modeling in a broad sense, mostly associated with so-called multi-way modeling and constrained programming.