

Khatri–Rao Space-Time Codes

Nicholas D. Sidiropoulos, *Senior Member, IEEE*, and Ramakrishna S. Budampati, *Student Member, IEEE*

Abstract—Space-time (ST) coding techniques exploit the spatial diversity afforded by multiple transmit and receive antennas to achieve reliable transmission in scattering-rich environments. ST block codes are capable of realizing full diversity and spatial coding gains at relatively low rates; ST trellis codes can achieve better rate-diversity tradeoffs at the cost of high complexity. On the other hand, V-BLAST supports high rates but has no built-in spatial coding and does not work well with fewer receive than transmit antennas. We propose a novel linear block-coding scheme based on the Khatri-Rao matrix product. The proposed scheme offers flexibility for achieving full-rate or full-diversity, or a desired rate-diversity tradeoff, and it can handle any transmit/receive antenna configuration or signal constellation. The proposed codes are shown to have numerous desirable properties, including guaranteed unique linear decodability, built-in blind channel identifiability, and efficient near-maximum likelihood decoding.

Index Terms—Blind channel identifiability, fading channels, multi-antenna systems, receive diversity, space-time codes, transmit diversity, wireless communications.

I. INTRODUCTION

NEXT-generation wireless systems aim for high rates to support broadband data access. Existing cellular standards do not support the high data rates required for most real-time multimedia services. A new class of wireless communication methods employing multiple transmit and/or receive antennas has recently been developed to achieve higher spectral efficiency in scattering-rich environments. It has been established that channel capacity grows linearly as the number of transmit and receive antennas grow simultaneously [12], [30]. Third-generation cellular standards (e.g., code division multiple access (CDMA 2000) [1] and wideband CDMA [2], [21]) have adopted space-time (ST) coding and modulation techniques that exploit the presence of multiple transmit antennas.

The rate-performance tradeoff lies at the heart of multi-antenna system design. Additional key issues include transmitter and receiver complexity and acquiring and tracking channel state information (CSI) at the receiver. So far, most of the multi-antenna systems have targeted either high performance at relatively low rates or high rates with relatively poor performance. Two paradigms have emerged to date on opposite

ends of this “spectrum”: ST coding and spatial multiplexing. The latter is best embodied by the Vertical—Bell Labs Layered ST (V-BLAST) architecture [14]. A variety of ST coding schemes have been proposed for the quasi-static, flat-fading, multiple-antenna wireless channel. ST block codes based on orthogonal designs (OD) [3], [29] and ST trellis codes [28] are two important classes developed for the case that CSI is available at the receiver. ST-OD codes suffer rate loss for more than 2 (50% for more than 4) transmit antennas. ST trellis codes strike better rate-performance tradeoffs but at the cost of complexity (exponential in the transmission rate and the number of antennas). Both OD and trellis codes require CSI at the receiver. Since neither class of codes is designed to support blind CSI recovery, it is implicitly assumed that accurate CSI may be acquired through training. This requires that the channel remains time-invariant for relatively long transmission epochs. Differential ST modulation (DSTM) [16], [18], [19] circumvents the need for channel estimation and is capable of handling moderate time variation. The drawback is an upfront differential detection penalty, similar to differential phase-shift keying (DPSK), and high decoding complexity at higher rates or with many antennas. Giving up the spatial coding advantage, V-BLAST can handle high rates with reasonable complexity, but the decoding scheme in [14] does not work well with fewer receive than transmit antennas, which is the typical situation in the downlink.

Recently, a wide class of ST block codes has been proposed in [15], under the name linear dispersion (LD) codes. Every linear block code is an LD code. The interest is therefore on the code design procedure, i.e., the selection of code matrices from the general class of LD codes. In [15], code design is approached as an iterative numerical optimization problem with mutual information as the objective function. Because other design considerations are not explicitly accounted for, LD designs in [15] do not necessarily provide full diversity and coding benefits. LD designs also assume that accurate CSI can be acquired through training. LD codes can support high rates without constraints on the number of transmit/receive antennas or the signal constellation.

In this paper, we propose a broad new class of ST codes based on the Khatri–Rao matrix product: *Khatri–Rao Space-Time* (KRST) codes. KRST codes are linear block codes designed to provide the following benefits:

- maximum diversity gain and good coding gain at any transmission rate (note the difference with LD code design, which aims for maximum mutual information);
- ability to easily span the range from full diversity to full transmission rate by simply truncating the code matrix;
- ability to work with any configuration of transmit/receive antennas and any signal constellation;

Manuscript received May 7, 2001; revised May 6, 2002. This work was supported by the National Science Foundation under CAREER award CCR-0096165 and award Wireless CCR-0096164. The associate editor coordinating the review of this paper and approving it for publication was Dr. Naofal M. W. Al-Dhahir.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: nikos@ece.umn.edu; rama@ece.umn.edu).

Publisher Item Identifier 10.1109/TSP.2002.803341.

- unique linear decodability for almost every channel matrix, irrespective of statistics;
- built-in blind CSI recovery *without modifying or interrupting the transmission*;
- effective low-complexity channel tracking after initial CSI has been acquired;
- efficient near maximum likelihood (ML) decoding due to linearity.

The rest of the paper is organized as follows. In Section II, the channel and data model are introduced. The encoding and decoding procedures are presented in Section III, which also includes necessary background on the Khatri-Rao product and code design criteria. The unique linear decodability property of KRST codes is also explained in this section. Blind CSI recovery is discussed in Section IV, while decision-directed channel tracking is discussed in Section V. Section VI provides several examples and comparisons with some competing ST coding techniques, and the main conclusions are summarized in Section VII. The proof of a unique linear decodability result pertaining to our code design is deferred to the Appendix.

II. DATA AND CHANNEL MODEL

Consider the multi-antenna system with M transmit antennas and N receive antennas depicted in Fig. 1. The wireless channel is assumed to be quasi-static and flat fading. The discrete-time baseband-equivalent model for the received data is then given by (cf. e.g., [15])

$$\mathbf{x}_t = \sqrt{\frac{\rho}{M}} \mathbf{H}_t \mathbf{c}_t + \mathbf{n}_t. \quad (1)$$

Here, $\mathbf{c}_t \in \mathbb{C}^{M \times 1}$ denotes the complex transmitted code signal vector with unit power entries¹ ($E[\mathbf{c}_t^H \mathbf{c}_t] = M$), $\mathbf{n}_t \in \mathbb{C}^{N \times 1}$ denotes zero-mean i.i.d. in space and time circular Gaussian ($\mathcal{CN}(0, 1)$) noise, and $\mathbf{x}_t \in \mathbb{C}^{N \times 1}$ denotes the complex received signal vector during one channel use. The channel matrix \mathbf{H}_t has i.i.d. $\mathcal{CN}(0, 1)$ entries, implying that $E[\text{trace}(\mathbf{H}_t \mathbf{H}_t^H)] = MN$. \mathbf{c}_t , \mathbf{n}_t , and \mathbf{H}_t are mutually independent. ρ is the signal-to-noise ratio (SNR) at each receive antenna. This is ensured by the normalization $\sqrt{\rho/M}$.

When the channel is constant for at least K channel uses, we obtain (dropping the block-time dependence for brevity)

$$\mathbf{X} = \sqrt{\frac{\rho}{M}} \mathbf{H} \mathbf{C} + \mathbf{N} \quad (2)$$

where $\mathbf{X} \in \mathbb{C}^{N \times K}$ is the received signal matrix, $\mathbf{C} \in \mathbb{C}^{M \times K}$ is the transmitted code matrix, $\mathbf{N} \in \mathbb{C}^{N \times K}$ is the additive noise matrix (Fig. 2) and $E[\text{trace}(\mathbf{C} \mathbf{C}^H)] = KM$.

III. KHATRI-RAO SPACE-TIME CODES

Adapting the symbol constellation is one simple way to control the transmission rate of linear encoding techniques. Our approach to transmission rate control relies on adjusting the length of the codeblock K via simple truncation of the code matrices. With proper code design, this allows us to span the range from

¹Note that H stands for Hermitian (complex conjugate) transpose, * is reserved for conjugation, and T for transpose.

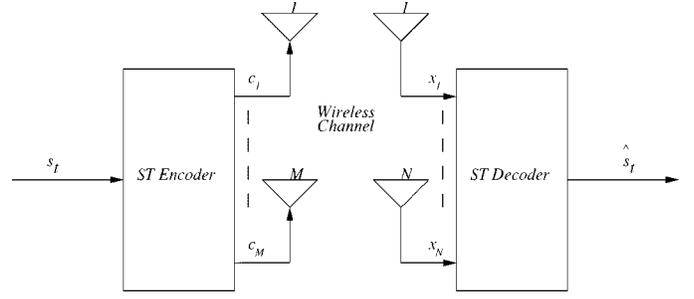


Fig. 1. Wireless channel model.

full rate to full transmit diversity while maintaining maximum possible diversity order for every K in between. The encoding and decoding schemes are discussed next.

A. Encoding Technique

Consider the model in (2). We will now describe the construction of the ST code matrix \mathbf{C} , which is to be transmitted from M antennas over K time slots. The symbol stream s_i [complex symbols chosen from an arbitrary, say ψ -phase shift-keying (PSK)/quadrature amplitude modulation (QAM) constellation and then scaled by $1/(\sqrt{E[s_i s_i^*]})$] is first parsed into $M \times 1$ symbol vectors \mathbf{s}_t . These vectors satisfy the power constraint $E[\mathbf{s}_t^H \mathbf{s}_t] = M$. Each of these symbol vectors is linearly precoded by an $M \times M$ matrix Θ , which is a suitably chosen constellation rotation (CR) matrix [4], [10], [13], [32]. Construction of this Θ will be discussed in Section III-B. The reason for including this Θ is to load each symbol onto each transmit antenna and time slot. This is necessary for diversity purposes, as explained in Section III-B. Next, the $M \times 1$ vector $\Theta \mathbf{s}_t$ is used to form the information-bearing matrix $\mathbf{D}(\Theta \mathbf{s}_t)$, which is a diagonal matrix holding the vector $\Theta \mathbf{s}_t$ on its diagonal. The last step is to post-multiply $\mathbf{D}(\Theta \mathbf{s}_t)$ by an $M \times K$ matrix \mathbf{C}_0^T . The purpose of \mathbf{C}_0^T is to adjust² the time span of the code matrix to within K channel slots. The construction of \mathbf{C}_0^T will be addressed in Section III-B.

The resulting transmitted code matrix is given by

$$\mathbf{C}_t = \mathbf{D}(\Theta \mathbf{s}_t) \mathbf{C}_0^T. \quad (3)$$

If the channel is assumed to be constant for block-time T , then the received data can be modeled as

$$\begin{aligned} \mathbf{X}_t &= \sqrt{\frac{\rho}{M}} \mathbf{H} \mathbf{C}_t + \mathbf{N}_t \\ &= \sqrt{\frac{\rho}{M}} \mathbf{H} \mathbf{D}(\Theta \mathbf{s}_t) \mathbf{C}_0^T + \mathbf{N}_t, \quad t = 1, \dots, T. \end{aligned} \quad (4)$$

In our scheme, each linearly precoded symbol (diagonal element of $\mathbf{D}(\Theta \mathbf{s}_t)$) “rides” onto a rank-one matrix factor that is generated by the outer product of the corresponding column of \mathbf{H} and the associated row of \mathbf{C}_0^T . The rate of transmission is

$$R = \left(\frac{M}{K} \right) \log_2(\psi) \text{bits/channel use}. \quad (5)$$

²Note that, in our construction, K can be $\leq M$ (compression) or $> M$ (expansion); we will focus mainly on the case $K \leq M$ because $K > M$ sacrifices rate without providing any diversity or identifiability benefits.

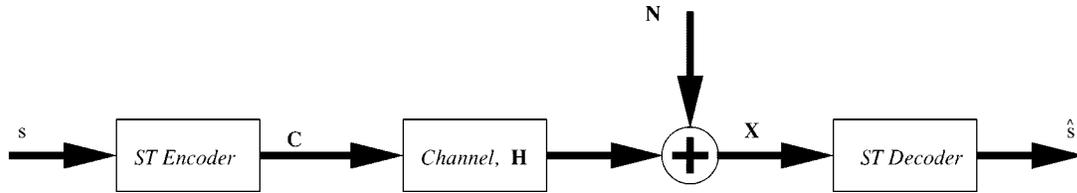


Fig. 2. System model.

B. Performance Analysis and Design Criteria

Recall that the quasi-static flat-fading channel \mathbf{H} has been assumed to have independent $\mathcal{CN}(0,1)$ entries. Consider the case where a maximum likelihood (ML) receiver decodes erroneously in favor of a signal vector $\tilde{\mathbf{s}}_t$ ($M \times K$ code matrix $\mathbf{E}_t = \mathbf{D}(\Theta\tilde{\mathbf{s}}_t)\mathbf{C}_0^T$) when \mathbf{s}_t ($M \times K$ code matrix $\mathbf{C}_t = \mathbf{D}(\Theta\mathbf{s}_t)\mathbf{C}_0^T$) was actually transmitted. The receiver is assumed to have perfect CSI as well as knowledge of \mathbf{C}_0 and Θ . The conditional pairwise error probability can then be approximated by [28]

$$P(\mathbf{C}_t \rightarrow \mathbf{E}_t | \mathbf{H}) \leq \exp\left(\frac{-d^2(\mathbf{C}_t, \mathbf{E}_t)\rho}{4M}\right) \quad (6)$$

where

$$d^2(\mathbf{C}_t, \mathbf{E}_t) = \|\mathbf{H}(\mathbf{C}_t - \mathbf{E}_t)\|_F^2 = \text{trace}[\mathbf{H}(\mathbf{C}_t - \mathbf{E}_t)(\mathbf{C}_t - \mathbf{E}_t)^H \mathbf{H}^H] \quad (7)$$

and $\|\cdot\|_F$ is the Frobenius norm. Define an $M \times M$ matrix \mathbf{A} as

$$\mathbf{A} = (\mathbf{C}_t - \mathbf{E}_t)(\mathbf{C}_t - \mathbf{E}_t)^H. \quad (8)$$

The rank of \mathbf{A} is $\leq \min(M, K)$. \mathbf{A} is a Hermitian positive semi-definite matrix with square root

$$\mathbf{B} = (\mathbf{C}_t - \mathbf{E}_t). \quad (9)$$

The eigenvalues of \mathbf{A} (λ_i 's) are non-negative real numbers. Following the analysis given in [28], we obtain the following upper bound on the average probability of error (averaged over channel statistics)

$$P(\mathbf{C}_t \rightarrow \mathbf{E}_t) \leq \left(\frac{1}{\prod_{i=1}^M \left(1 + \frac{\lambda_i \rho}{4M}\right)}\right)^N. \quad (10)$$

If r is the rank of \mathbf{A} ($=$ rank of \mathbf{B}) and $\lambda_1, \lambda_2, \dots, \lambda_r$ are its nonzero eigenvalues, then at high SNR, (10) reduces to

$$P(\mathbf{C}_t \rightarrow \mathbf{E}_t) \leq \left(\prod_{i=1}^r \lambda_i\right)^{-N} \left(\frac{\rho}{4M}\right)^{-Nr}. \quad (11)$$

Thus, a diversity gain of Nr (*rank criterion* of ST code design) and a coding gain of $(\prod_{i=1}^r \lambda_i)^{1/r}$ (*determinant criterion* of ST code design) are achieved. The diversity gain manifests itself at high SNR. The maximum diversity of a system with fixed (M, N, K) is $N \min(M, K)$, and this is achieved if and only if \mathbf{B} has full rank $\min(M, K)$ ($= K$ if $K \leq M$) for all $\mathbf{C}_t \neq \mathbf{E}_t$. This criterion dictates the choice of Θ and (in part) the choice of \mathbf{C}_0 .

1) *Choice of Θ* : The constellation rotation (CR) matrix Θ , which is used to linearly precode the symbol vector \mathbf{s}_t , is employed to take advantage of the transmit diversity in a multi-antenna environment. According to the rank criterion of ST code design, we are interested in a code structure with the property

$$\begin{aligned} \mathbf{B} &= \mathbf{C}_t - \mathbf{E}_t = \mathbf{D}(\Theta(\mathbf{s}_t - \tilde{\mathbf{s}}_t))\mathbf{C}_0^T \\ &= \text{full rank}, \forall \mathbf{s}_t \neq \tilde{\mathbf{s}}_t. \end{aligned} \quad (12)$$

Now, suppose that \mathbf{C}_0 is chosen to have full rank. Let us also suppose that Θ is such that $\Theta(\mathbf{s}_t - \tilde{\mathbf{s}}_t)$ contains no zeros for all $\mathbf{s}_t \neq \tilde{\mathbf{s}}_t$. Then, the diagonal matrix $\mathbf{D}(\Theta(\mathbf{s}_t - \tilde{\mathbf{s}}_t))$ is non-singular, and the desired full rank of \mathbf{B} follows for all $\mathbf{s}_t \neq \tilde{\mathbf{s}}_t$, implying that the code has maximum diversity advantage $N \min(M, K)$. Hence, the design rule is as follows: For full diversity gain, Θ should be chosen such that $\Theta(\mathbf{s}_t - \tilde{\mathbf{s}}_t)$ contains no zeros for all $\mathbf{s}_t \neq \tilde{\mathbf{s}}_t$, and \mathbf{C}_0 must have full rank. The CR matrices discussed in [4], [10], [13], and [32] satisfy this requirement on Θ . CR matrices are designed using either algebraic number-theoretic constructions or by computer search over compact parameterizations of unitary matrices.

We restrict ourselves to a complex *unitary* constellation rotation matrix Θ , as it is worth using complex unitary rotations when M and/or N are large, and the constellation size is moderate [32]. The specification of such a Θ , which is suitable for pulse amplitude modulation (PAM) and QAM constellations and $M = 2^m$, can be found in [32]

$$\Theta_M := \frac{1}{\sqrt{M}} \mathbf{F}_M \text{diag}(1, \alpha, \dots, \alpha^{M-1}) \quad (13)$$

where \mathbf{F}_M is the $M \times M$ inverse DFT matrix, and $\alpha = \exp(j2\pi/4M)$. For odd M , Θ is obtained through computer search over the unitary parameterization expressed via Givens rotation matrices. Details of this construction can be found in [32]. Note that in our context (which includes post-multiplication by \mathbf{C}_0^T), the CR designs of Θ in [32] need not be optimum from a coding gain viewpoint. However, extensive experimentation has shown that (13) yields good coding gain relative to randomly sampled Θ , which also meets the full diversity gain requirement almost surely. This will be corroborated by simulations in Section VI.

2) *Choice of \mathbf{C}_0* : From here on, we focus on the $K \leq M$ case. \mathbf{C}_0 is chosen to be a $K \times M$ Vandermonde matrix with generators $e^{j2\pi(m-1)/M}$, i.e., $c_{k,m} = e^{j2\pi(m-1)/M(k-1)}$, $m = 1, \dots, M$. This way, \mathbf{C}_0 is a scaled semi-unitary matrix ($\mathbf{C}_0 \mathbf{C}_0^H = M\mathbf{I}$), and it has full row rank, as required for achieving maximum diversity gain. This gives the flexibility to go from $K = M$ (full diversity) to $K = 1$ (full rate) by simple truncation of the code matrix while maintaining maximum

achievable diversity gain anywhere in between. Coupled with a unitary Θ , this choice of \mathbf{C}_0 also assures that the transmit power constraint is satisfied:

$$\begin{aligned}
E[\text{trace}(\mathbf{C}_t^H \mathbf{C}_t)] &= E[\text{trace}(\mathbf{C}_0^* \mathbf{D}^* (\Theta \mathbf{s}_t) \mathbf{D} (\Theta \mathbf{s}_t) \mathbf{C}_0^T)] \\
&= E[\text{trace}(\mathbf{C}_0^T \mathbf{C}_0^* \mathbf{D}^* (\Theta \mathbf{s}_t) \mathbf{D} (\Theta \mathbf{s}_t))] \\
&= E\left[\sum_{m=1}^M \left|\theta_m^T \mathbf{s}_t\right|^2 \mathbf{c}_m^H \mathbf{c}_m\right] \\
&= K \sum_{m=1}^M E\left|\theta_m^T \mathbf{s}_t\right|^2 \\
&= K \sum_{m=1}^M \theta_m^T E(\mathbf{s}_t \mathbf{s}_t^H) (\theta_m^T)^H \\
&= K \sum_{m=1}^M \theta_m^T \mathbf{I}_{M \times M} (\theta_m^T)^H \\
&= KM
\end{aligned}$$

where θ_m^T is the m th row of the unitary matrix Θ , and \mathbf{c}_m is the m th column of \mathbf{C}_0 .

An important consideration behind our choice of \mathbf{C}_0 is that this particular Vandermonde structure guarantees that the equivalent channel matrix used in the decoding³ is full rank for almost every channel matrix \mathbf{H} (cf. Section III-D and the Appendix).

Note that for $K = M$ and $\mathbf{C}_0^T = \mathbf{I}$, our scheme reverts to linear constellation precoding with time-division multiplexing of the antenna elements across the time slots of one code block [32]. However, its rate is then only one symbol per channel use. As the ST-linear constellation precoding (LCP) codes [32] do not work with $K < M$, they do not provide the flexibility to trade off diversity for transmission rate.

C. Decoding Technique

We now discuss the decoding of the received signal matrix. Assuming that the receiver has perfect CSI, we use a basic property of the Khatri–Rao product (KRP) to perform coherent detection of the transmitted symbols.

1) *Khatri–Rao Product (KRP)*: Given two matrices $\mathbf{A}(I \times F)$ and $\mathbf{B}(J \times F)$ with the same number of columns, the KRP $\mathbf{A} \odot \mathbf{B}$ is the $IJ \times F$ matrix defined as

$$\mathbf{A} \odot \mathbf{B} := [\mathbf{a}_1 \otimes \mathbf{b}_1 \cdots \mathbf{a}_F \otimes \mathbf{b}_F] \quad (14)$$

where \mathbf{a}_f is the f th column of \mathbf{A} , similarly for \mathbf{b}_f , and \otimes denotes the Kronecker product of the two column vectors. The KRP has the following key property [5]:

$$\text{vec}(\mathbf{A} \mathbf{D} \mathbf{B}^T) = (\mathbf{B} \odot \mathbf{A}) d(\mathbf{D}) \quad (15)$$

where \mathbf{D} is $F \times F$ diagonal, $\text{vec}(\cdot)$ stacks the columns of its argument, and $d(\cdot)$ extracts the diagonal of its argument and constructs a column vector out of it.

³As we will shortly see, this is the Khatri–Rao product of \mathbf{C}_0 and the actual (random) channel matrix \mathbf{H} .

2) *Vectorized Model*: Using (15), we obtain from (4) the following (noiseless) vectorized model:

$$\begin{aligned}
\text{vec}(\mathbf{X}_t) &= \sqrt{\frac{\rho}{M}} (\mathbf{C}_0 \odot \mathbf{H}) d(\mathbf{D} (\Theta \mathbf{s}_t)) \\
&= \sqrt{\frac{\rho}{M}} (\mathbf{C}_0 \odot \mathbf{H}) \Theta \mathbf{s}_t, \quad t = 1, \dots, T. \quad (16)
\end{aligned}$$

In the presence of noise, a number of decoders can be used to extract the symbols transmitted from the received signal matrix. This is done by writing the received matrix as

$$\begin{aligned}
\mathbf{x}_t := \text{vec}(\mathbf{X}_t) &= \sqrt{\frac{\rho}{M}} (\mathbf{C}_0 \odot \mathbf{H}) \Theta \mathbf{s}_t + \text{vec}(\mathbf{N}_t) \\
&= \mathbf{H}_{\text{eq}} \mathbf{s}_t + \text{vec}(\mathbf{N}_t), \quad t = 1, \dots, T. \quad (17)
\end{aligned}$$

An exhaustive search-based ML detection can be prohibitively complex (exponential in M). An alternative is *sphere decoding* (SD) [9], [31], which can achieve near ML performance at significantly reduced complexity. Block minimum mean square error–decision feedback equalization (MMSE-DFE) [11], [27] can also be used if we have to further reduce the complexity at the expense of performance.

In this paper, we use SD at the receiver end. It has been established that complexity of the original SD, for M transmit antennas and a fixed search radius, is $\mathcal{O}(M^6)$ [9], [31], independent of the constellation size. Efficient implementation reduces the average complexity to $\mathcal{O}(M^3)$ [17]. In our simulations, we will consider QAM or PSK constellations that are scaled to maintain the power constraint $E[\mathbf{s}_t^H \mathbf{s}_t] = M$. Since SD works for QAM or PAM constellations (which are carved from the cubic lattice with integer-valued coordinates), we scale the channel matrix \mathbf{H}_{eq} by $1/(\sqrt{E[s_i s_i^*]})$ and \mathbf{s}_t by $\sqrt{E[s_i s_i^*]}$ so that the signal vector now corresponds to a vector consisting of QAM or PAM symbols carved from this cubic lattice. In our decoding scheme, since \mathbf{H}_{eq} and \mathbf{s}_t can be complex, we rewrite the model in the following real-equivalent form⁴:

$$\begin{aligned}
\begin{bmatrix} \text{Real}(\mathbf{x}_t) \\ \text{Imag}(\mathbf{x}_t) \end{bmatrix} &= \begin{bmatrix} \text{Real}(\mathbf{H}_{\text{eq}}) & -\text{Imag}(\mathbf{H}_{\text{eq}}) \\ \text{Imag}(\mathbf{H}_{\text{eq}}) & \text{Real}(\mathbf{H}_{\text{eq}}) \end{bmatrix} \\
&\quad \times \begin{bmatrix} \text{Real}(\mathbf{s}_t) \\ \text{Imag}(\mathbf{s}_t) \end{bmatrix} + \begin{bmatrix} \text{Real}(\text{vec}(\mathbf{N}_t)) \\ \text{Imag}(\text{vec}(\mathbf{N}_t)) \end{bmatrix}. \quad (18)
\end{aligned}$$

D. Unique Linear Decodability

In the Appendix, we show that the KRST codes have the unique linear decodability property, i.e., the transmitted symbols are guaranteed to be linearly recoverable in the absence of noise for almost every channel matrix \mathbf{H} . In the noiseless case, the transmitted symbol vector \mathbf{s}_t can be recovered by

$$\mathbf{s}_t = \left(\sqrt{\frac{\rho}{M}} (\mathbf{C}_0 \odot \mathbf{H}) \Theta \right)^\dagger \text{vec}(\mathbf{X}_t) \quad (19)$$

where \dagger denotes the matrix pseudo-inverse, provided $\left(\sqrt{\frac{\rho}{M}} (\mathbf{C}_0 \odot \mathbf{H}) \Theta \right)$ is tall or square and has full column rank M . Since Θ is a unitary $M \times M$ matrix, it has full rank

⁴If $x = a + jb$, then $\text{Real}(x) = a$, and $\text{Imag}(x) = b$

M . ($\mathbf{C}_0 \odot \mathbf{H}$) is an $NK \times M$ matrix. We prove in the Appendix that for our particular choice of \mathbf{C}_0 , this KRP is full column rank for almost every \mathbf{H} , provided $NK \geq M$. Note that KRST codes can work with as few as 1 receive antenna if $K = M$.

Unique linear decodability is important for the following two reasons.

- The SD benefits from, and in fact the original SD algorithm requires, unique linear decodability—that is, a full column rank $N_{\text{eq}} \times M_{\text{eq}}$ equivalent channel matrix [9], [31]. Briefly, when $N_{\text{eq}} \geq M_{\text{eq}}$ (and the equivalent channel matrix is full column rank), one works with an M_{eq} -dimensional lattice embedded in $\mathbb{R}^{M_{\text{eq}}}$, whereas when $N_{\text{eq}} < M_{\text{eq}}$ (or the equivalent channel matrix is not full column rank), one works with a *projection* of this lattice onto $\mathbb{R}^{N_{\text{eq}}}$. The latter problem is more complex, and it cannot be handled by the original SD; a conference-paper generalization of SD for this case appears in [8], but there are two drawbacks relative to the full column rank case: i) Complexity increases significantly (this issue is not elaborated in the conference paper [8]); ii) performance is considerably worse compared with the full column rank case.
- Unique linear decodability enables computationally simpler equalization, like nulling and cancelling, or zero-forcing linear inverse in case ML or SD becomes prohibitively complex. This too is a desirable property.

E. Choice of K

The choice of K in KRST codes is dictated by the desired rate-performance tradeoff, subject to the constraint $NK \geq M$. To be more precise, the desired transmission rate R determines K through $R = (M/K) \log_2(\psi)$ bits/channel use, provided the desired rate is achievable for integer $K \geq 1$ satisfying $NK \geq M$. In case $N \geq M$, we are free to choose any positive integer K . Either way, the resulting KRST code will yield full diversity for the given M , N , and K , equal to $N \min(M, K)$. Alternatively, we may choose K for a given diversity order (slope at high SNR) $N \min(M, K)$, again under the constraint $NK \geq M$.

IV. BLIND CSI RECOVERY: PARAFAC ANALYSIS

A key benefit of KRST codes is their built-in blind CSI recovery capability. It is important to note that the encoding technique remains the same as in the known CSI case discussed in Section III-A. The blind-KRST decoding technique makes use of the blind identifiability properties of the parallel factor (PARAFAC) model [6], [23], [25], [26].

A. PARAFAC

PARAFAC analysis is a common name for low-rank decomposition of three- or higher dimensional arrays. Consider an $I \times J \times K$ three-way array $\underline{\mathbf{X}}$ with typical element $x_{i,j,k}$ and the F -component trilinear decomposition

$$x_{i,j,k} = \sum_{f=1}^F a_{i,f} b_{j,f} c_{k,f} \quad (20)$$

for $i = 1, \dots, I$, $j = 1, \dots, J$, and $k = 1, \dots, K$. The three-way array $\underline{\mathbf{X}}$ is expressed as a sum of F rank-one three-way factors in the above equation. The rank of a three-way array $\underline{\mathbf{X}}$ is defined as the minimum number of rank-one (three-way) components needed to decompose $\underline{\mathbf{X}}$. The vectors $\mathbf{a}_f \in \mathbb{C}^{I \times 1}$, $\mathbf{b}_f \in \mathbb{C}^{J \times 1}$, and $\mathbf{c}_f \in \mathbb{C}^{K \times 1}$ are often referred to as loading vectors or score vectors or factor profiles in the PARAFAC literature, depending on the context.

Define an $I \times F$ matrix \mathbf{A} , $J \times F$ matrix \mathbf{B} , and $K \times F$ matrix \mathbf{C} with typical elements $a_{i,f}$, $b_{j,f}$, and $c_{k,f}$, respectively. Furthermore, define $J \times K$ matrices \mathbf{X}_i , $I \times K$ matrices \mathbf{Y}_j , and $I \times J$ matrices \mathbf{Z}_k with corresponding typical elements $\mathbf{X}_i(j, k) := \mathbf{Y}_j(i, k) := \mathbf{Z}_k(i, j) := x_{i,j,k}$. Then, the model in (20) can be written in three equivalent ways:

$$\mathbf{X}_i = \mathbf{B} \mathbf{D}_i(\mathbf{A}) \mathbf{C}^T, \quad i = 1, \dots, I \quad (21)$$

$$\mathbf{Y}_j = \mathbf{A} \mathbf{D}_j(\mathbf{B}) \mathbf{C}^T, \quad j = 1, \dots, J \quad (22)$$

$$\mathbf{Z}_k = \mathbf{A} \mathbf{D}_k(\mathbf{C}) \mathbf{B}^T, \quad k = 1, \dots, K \quad (23)$$

where $\mathbf{D}_i(\mathbf{A})$ is a diagonal matrix constructed out of the i th row of \mathbf{A} . Stacking the matrices in (21), we obtain

$$\mathbf{X}^{(JI \times K)} := \begin{bmatrix} \mathbf{X}_{i=1} \\ \mathbf{X}_{i=2} \\ \vdots \\ \mathbf{X}_{i=I} \end{bmatrix} = \begin{bmatrix} \mathbf{B} \mathbf{D}_1(\mathbf{A}) \\ \mathbf{B} \mathbf{D}_2(\mathbf{A}) \\ \vdots \\ \mathbf{B} \mathbf{D}_I(\mathbf{A}) \end{bmatrix} \mathbf{C}^T = (\mathbf{A} \odot \mathbf{B}) \mathbf{C}^T \quad (24)$$

where the superscript $(JI \times K)$ means that the matrix is of size $JI \times K$, and the j -index (J goes first in the product JJ) runs faster than the i -index along its columns.

Given $\mathbf{A} \in \mathbb{C}^{I \times F}$, $r_{\mathbf{A}} := \text{rank of } \mathbf{A} = r$ iff it contains a collection of r linearly independent columns but no collection of $r+1$ linearly independent columns. $k_{\mathbf{A}} := k$ -rank of $\mathbf{A} = k$ if every k columns are linearly independent, but either $k = F$, or there exists a collection of $k+1$ linearly dependent columns in \mathbf{A} ($k_{\mathbf{A}} \leq r_{\mathbf{A}} \leq \min(I, F)$, $\forall \mathbf{A}$).

A distinguishing feature of the PARAFAC model is its uniqueness. Under mild conditions, the model parameterization is essentially unique, that is, \mathbf{A} , \mathbf{B} , and \mathbf{C} are identifiable without rotational ambiguities. In particular, if

$$k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2F + 2 \quad (25)$$

then \mathbf{A} , \mathbf{B} , and \mathbf{C} are unique up to common permutation and (complex) scaling–counterscaling of columns [23], [25], [26]. If all three matrices have full k -rank,⁵ then

$$\min(I, F) + \min(J, F) + \min(K, F) \geq 2F + 2 \quad (26)$$

is sufficient for identifiability.

When noisy observations $\tilde{x}_{i,j,k}$ are given, the principle of alternating least squares (ALS) can be used to fit the PARAFAC model in (20). The idea behind ALS is to update one matrix, using least squares (LS) conditioned on previously obtained interim estimates for the remaining matrices, and then proceed to update the other matrices. This process is repeated until convergence in least squares fit. A basic trilinear ALS (TALS) algo-

⁵True with probability 1 if drawn from a continuous distribution.

rithm is presented in [25]. The TALS method is guaranteed to converge monotonically. It is conceptually simple and provides good performance [6]. Least squares fitting of (24) (and ML parameter estimation when the noise is modeled as i.i.d. Gaussian and all other parameters are treated as deterministic unknowns) amounts to

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \left\| \begin{bmatrix} \tilde{\mathbf{X}}_1 \\ \vdots \\ \tilde{\mathbf{X}}_I \end{bmatrix} - \begin{bmatrix} \mathbf{B}\mathbf{D}_1(\mathbf{A}) \\ \vdots \\ \mathbf{B}\mathbf{D}_I(\mathbf{A}) \end{bmatrix} \mathbf{C}^T \right\|_F^2 \quad (27)$$

where $\tilde{\mathbf{X}}_i$, $i = 1, \dots, I$ are the noisy slabs. From (27), the conditional least squares update for \mathbf{C} is

$$\hat{\mathbf{C}}^T = \begin{bmatrix} \hat{\mathbf{B}}\mathbf{D}_1(\hat{\mathbf{A}}) \\ \vdots \\ \hat{\mathbf{B}}\mathbf{D}_I(\hat{\mathbf{A}}) \end{bmatrix}^\dagger \begin{bmatrix} \tilde{\mathbf{X}}_1 \\ \vdots \\ \tilde{\mathbf{X}}_I \end{bmatrix} \quad (28)$$

where $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ denote the previously obtained estimates of \mathbf{A} and \mathbf{B} , respectively. The complete symmetry of the trilinear model [cf. (20)] and data reshaping [cf. (21)–(23)] can be used to figure out corresponding conditional LS updates for \mathbf{A} and \mathbf{B} .

B. Blind-KRST Decoding Technique

The blind-KRST decoding technique is based on the following two-step scheme: First, estimate the channel matrix, and then, use this estimate to decode the transmitted signals. The estimation of the channel matrix is done in two stages. First, a PARAFAC model is fitted to the data \mathbf{X}_t , $t = 1, \dots, T$. This is done as follows. Let $\underline{\mathbf{X}}$ be a three-way array of dimensions $T \times N \times K$, whose “slabs” are given by

$$\underline{\mathbf{X}}(t, :, :) = \sqrt{\frac{\rho}{M}} \mathbf{H}\mathbf{D}_t(\mathbf{A})\mathbf{C}_0^T + \mathbf{N}_t, \quad t = 1, \dots, T \quad (29)$$

where the $T \times M$ matrix $\mathbf{A} := (\Theta\mathbf{S})^T$, \mathbf{S} is an $M \times T$ matrix whose columns are the vectors \mathbf{s}_t , $t = 1, \dots, T$, T is the number of slabs in the PARAFAC model, and we have used MATLAB notation, where $\underline{\mathbf{X}}(t, :, :)$ denotes the t th $N \times K$ slab of $\underline{\mathbf{X}}$ perpendicular to the first (row) dimension. Then, the uniqueness condition (25) translates to

$$k_{\mathbf{A}} + k_{\mathbf{H}} + k_{\mathbf{C}_0} = k_{(\Theta\mathbf{S})^T} + k_{\mathbf{H}} + k_{\mathbf{C}_0} \geq 2M + 2. \quad (30)$$

Since the columns of \mathbf{H} are drawn independently from a \mathcal{CN} distribution, it has full k -rank almost surely, and \mathbf{C}_0 has k -rank equal to K by construction. Since Θ is square and full rank, $k_{(\Theta\mathbf{S})^T} = M$ [26], provided \mathbf{S} is full row rank; this is true with very high probability, provided T is large enough. For example, for $M = 4$, $T = 30$, and a BPSK constellation, where the probability of the matrix \mathbf{S} is rank deficient, is roughly $\mathcal{O}(10^{-6})$. Substituting in (30) yields

$$\min(N, M) + K \geq M + 2. \quad (31)$$

Note that blind CSI recovery only requires as few as two receive antennas or as few as two time slots per code block; identifica-

bility can never hold for either $K = 1$ or $N = 1$, however. In addition, note that for small T , and depending on the size of the symbol constellation, there may be non-negligible probability of rank deficiency of \mathbf{S} . However, simulations indicate that uniqueness often holds even with relatively small T . This is likely due to the fact that in our particular context, the matrix \mathbf{C}_0 is known, whereas in (25), all three matrices involved are assumed unknown. Hence, the derivation leading to (31) is, in a sense, pessimistic, which counteracts the effects of smaller T .

The recovery of \mathbf{H} is actually performed by fitting the model in an LS sense using TALS. As \mathbf{C}_0 is known to the receiver, we only need to estimate \mathbf{H} and $\mathbf{A} = (\Theta\mathbf{S})^T$; hence, the conditional least squares update step (28) is performed only for \mathbf{H} and \mathbf{A} , while keeping \mathbf{C}_0 fixed during the iterations. This modified TALS algorithm provides estimates $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{A}}$. The knowledge of \mathbf{C}_0 at the receiver takes care of the permutation ambiguity inherent in blind estimation. The scaling ambiguity ($\tilde{\mathbf{H}} = \mathbf{H}\mathbf{D}_s$, where \mathbf{D}_s is an arbitrary diagonal scaling matrix) that remains can be taken care of by transmitting an identity matrix at the beginning of the transmission burst, i.e.,

$$\underline{\mathbf{X}}(0, :, :) = \sqrt{\frac{\rho}{M}} \mathbf{H}\mathbf{C}_0^T + \mathbf{N}_0. \quad (32)$$

Applying the KRP property discussed in Section III-C using the estimated $\tilde{\mathbf{H}}$, we have

$$\begin{aligned} \underline{\mathbf{X}}(0, :, :) &\cong \sqrt{\frac{\rho}{M}} \tilde{\mathbf{H}}\mathbf{D}_s^{-1}\mathbf{C}_0^T + \mathbf{N}_0 \\ \text{vec}(\underline{\mathbf{X}}(0, :, :)) &\cong \sqrt{\frac{\rho}{M}} (\mathbf{C}_0 \odot \tilde{\mathbf{H}})d(\mathbf{D}_s^{-1}) + \text{vec}(\mathbf{N}_0). \end{aligned} \quad (33)$$

Now, an estimate of \mathbf{D}_s^{-1} is computed using the pseudo-inverse of $(\mathbf{C}_0 \odot \tilde{\mathbf{H}})$, as discussed for the coherent detection, yielding $\hat{\mathbf{D}}_s^{-1}$; then, the final estimate of the channel is obtained:

$$\hat{\mathbf{H}} = \tilde{\mathbf{H}}\hat{\mathbf{D}}_s^{-1}. \quad (34)$$

Once the channel matrix is estimated without any scaling or permutation ambiguity, we can apply the decoding technique explained in the coherent detection case to estimate the transmitted symbols. The complexity of PARAFAC (TALS) is $\mathcal{O}(MNKT)$ per iteration. The typical number of PARAFAC iterations as a function of SNR is as follows: When $M = 4$, $N = 4$, $K = 4$, and $T = 5$, at SNR = 5, 10, and 15, the average number of iterations is 15, 11, and 7, respectively. Note that since we fix the matrix \mathbf{C}_0^T , PARAFAC requires considerably fewer iterations than usual.

V. CHANNEL TRACKING

So far, we have assumed that the channel is time invariant. In practice, however, the channel is usually slowly varying with time. After initial blind acquisition using PARAFAC analysis, it makes sense to revert to computationally simpler decision-directed channel tracking to follow minor channel variations, without incurring the computational expense of blind estimation. This computationally attractive approach also serves to provide an updated scale reference for subsequent periodic blind

re-estimation of the channel matrix. The PARAFAC analysis described in Section IV-A forms the backbone of our proposed tracking technique.

Let us return to (29) and (32), which are reproduced here for convenience.

$$\begin{aligned}\underline{\mathbf{X}}(t, :, :) &= \sqrt{\frac{\rho}{M}} \mathbf{H} \mathbf{D}_t(\mathbf{A}) \mathbf{C}_0^T + \mathbf{N}_t, \quad t = 1, \dots, T \\ \underline{\mathbf{X}}(0, :, :) &= \sqrt{\frac{\rho}{M}} \mathbf{H} \mathbf{C}_0^T + \mathbf{N}_0.\end{aligned}$$

Let \mathbf{H}_{old} be the channel estimate obtained from these equations. If \mathbf{H}_{new} is the new channel matrix at block time $(T+1)$, then the corresponding received signal matrix is

$$\underline{\mathbf{X}}(T+1, :, :) = \sqrt{\frac{\rho}{M}} \mathbf{H}_{\text{new}} \mathbf{D}(\Theta_{\mathbf{s}_{T+1}}) \mathbf{C}_0^T + \mathbf{N}_{T+1}. \quad (35)$$

At the receiver, we use the old channel estimate \mathbf{H}_{old} to decode the transmitted symbols by the coherent detection scheme, i.e., apply SD to the following model:

$$\text{vec}(\underline{\mathbf{X}}(T+1, :, :)) = \sqrt{\frac{\rho}{M}} (\mathbf{C}_0 \odot \mathbf{H}_{\text{old}}) \Theta_{\mathbf{s}_{T+1}} + \text{vec}(\mathbf{N}_{T+1}). \quad (36)$$

This yields an estimate $\tilde{\mathbf{s}}_{T+1}$. Next, we update the matrix \mathbf{A} (cf. (29)) by shifting up the rows by 1, dropping the former first row and inserting $(\Theta_{\tilde{\mathbf{s}}_{T+1}})^T$ as the last row, i.e.,

$$\mathbf{A}(t, :) = \mathbf{A}(t+1, :), \quad t = 1, \dots, (T-1) \quad (37)$$

$$\mathbf{A}(T, :) = (\Theta_{\tilde{\mathbf{s}}_{T+1}})^T \quad (38)$$

then apply PARAFAC analysis to the three-way array $\underline{\mathbf{Q}}$, constructed as

$$\underline{\mathbf{Q}}(t, :, :) = \underline{\mathbf{X}}(t+1, :, :), \quad t = 1, \dots, T. \quad (39)$$

Since \mathbf{C}_0 and the updated \mathbf{A} are fixed, only the third component matrix \mathbf{H} has to be estimated; therefore, TALS reduces to a simple LS problem. Recall (23):

$$\mathbf{Z}_k = \mathbf{A} \mathbf{D}_k(\mathbf{C}) \mathbf{B}^T, \quad k = 1, \dots, K.$$

The compact model representation discussed in (24) corresponding to the above set of equations is

$$\mathbf{Z}^{(IK \times J)} = \begin{bmatrix} \mathbf{Z}_{k=1} \\ \mathbf{Z}_{k=2} \\ \vdots \\ \mathbf{Z}_{k=K} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \mathbf{D}_1(\mathbf{C}) \\ \mathbf{A} \mathbf{D}_2(\mathbf{C}) \\ \vdots \\ \mathbf{A} \mathbf{D}_K(\mathbf{C}) \end{bmatrix} \mathbf{B}^T = (\mathbf{C} \odot \mathbf{A}) \mathbf{B}^T. \quad (40)$$

Following this approach of ‘‘matricization,’’ the three-way array $\underline{\mathbf{Q}}$ is matricized into a $TK \times N$ matrix \mathbf{P} . In the noiseless case, we obtain

$$\mathbf{P} = (\mathbf{C}_0 \odot \mathbf{A}) \mathbf{H}^T. \quad (41)$$

In the presence of noise, the LS estimate of \mathbf{H}_{new} is given by

$$\tilde{\mathbf{H}}_{\text{new,LS}} = ((\mathbf{C}_0 \odot \mathbf{A})^\dagger \mathbf{P})^T. \quad (42)$$

$\tilde{\mathbf{H}}_{\text{new,LS}}$ now replaces \mathbf{H}_{old} , and the procedure is repeated again so that the channel is tracked continuously. Note that $\tilde{\mathbf{H}}_{\text{new,LS}}$ from (42) can now be used in place of \mathbf{H}_{old} in (36) to get a better estimate of the transmitted symbol vector \mathbf{s}_{T+1} . This vector estimate can again be used to update the last row of matrix \mathbf{A} in (38), which can then be used to get a better estimate of the channel. This cyclic process can be performed iteratively until convergence,⁶ but the main drawback is that SD has to be performed during each cycle. This makes the overall iterative process prohibitively complex, especially for real-time application; hence, we do not advocate further iteration at this point, especially for slowly varying channels.

VI. SIMULATION RESULTS AND COMPARISONS

The proposed KRST codes are flexible for going from full-rate codes to full-diversity codes. The fact that they perform well at high rates prompts a comparison with the recently proposed LD codes [15]. KRST codes are also compared with ST-LCP codes [32] as both codes employ the CR matrix Θ . The blind CSI recovery scheme is compared with the DSTM scheme [18]–[20]. Throughout the simulations, gray mapping is used to calculate the bit error rate (BER), and SD is used at the receiver for KRST, LD, and ST-LCP codes. The enumeration-based differential ML receiver proposed in [18], [19] is employed for DSTM.

A. KRST: $M = 4, N = 4$

Let us first look at the flexibility that KRST codes offer. We are interested in producing codes that help in achieving a desired tradeoff between rate and diversity performance. Consider the case where we have four transmit and four receive antennas. For K (number of slots per code block) = 1, we achieve full-rate, and for $K = 4$, we achieve full-diversity. The symbols to be transmitted are chosen from the QPSK constellation. The CR matrix Θ for $M = 4$ is given by [32]:

$$\Theta = \frac{1}{\sqrt{4}} \mathbf{F}_4 \text{diag}(1, \alpha, \alpha^2, \alpha^3) \quad (43)$$

where $\alpha = \exp(j2\pi/16)$. \mathbf{C}_0 is chosen as described in Section III-B2. K is the number of rows in \mathbf{C}_0 . Since the rest of the design remains the same, rate-diversity control through simple truncation (selection of K) provides significant code design and adaptation flexibility.

The SNR versus BER plot for various K is given in Fig. 3. The simulation result is based on 10^4 independent random realizations of the channel matrix \mathbf{H} , each used for 10^4 channel time slots. It is observed that as we go from $K = 1$ ($R = 8$) to $K = 4$ ($R = 2$), performance improves as expected. An interesting observation is that instead of using full-rate codes with poor performance or using full-diversity codes with low rate, it makes more sense to use codes that offer good performance at reasonably high rates, like the codes for $K = 2$ or $K = 3$ in a practical setting.

⁶Convergence in LS fit is guaranteed, as this is an ALS procedure; see the discussion in Section IV-A.

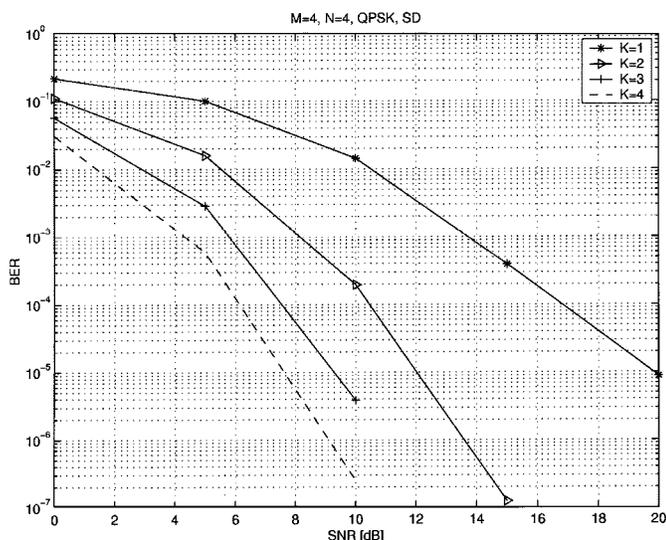


Fig. 3. Khatri-Rao space-time codes: $M = 4, N = 4, \text{QPSK}$.

B. KRST versus LD: $M = 2, N = 2$

Linear dispersion (LD) codes [15] are linear block codes that transmit substreams of data in linear combinations over space and time. LD codes are capable of achieving high rates and are in fact optimized for maximum information rate. KRST codes are linear block codes that are also capable of achieving high rates, but they aim for maximum diversity and blind identifiability.

Consider the case where we have two transmit and two receive antennas. Two time slots are used per code block, i.e., $K = 2$. In order to maintain $R = 4$, the symbols to be transmitted are chosen from a QPSK constellation for the LD code (code design: [15, p. 31, eqs. (31) and (34)]). For KRST codes, we use

$$\Theta = \frac{1}{\sqrt{2}} \mathbf{F}_2 \text{diag}(1, \alpha) \tag{44}$$

where $\alpha = \exp(j2\pi/8)$. For KRST with $K = 2$, the symbol constellation is 16QAM, whereas for $K = 1$, we use QPSK. The comparison between the two KRST codes and the LD code is given in Fig. 4. It is observed that KRST outperforms the LD code at high SNR.

C. KRST versus LD: $M = 3, N = 1$

One more comparison between KRST and LD codes is made here. In this case, we have three transmit antennas and one receive antenna. The symbols are chosen from QPSK constellation. Since we have an odd number of transmit antennas, we have to use computer search to obtain the constellation rotation matrix Θ [32]. This Θ is given by (45), shown at the bottom of

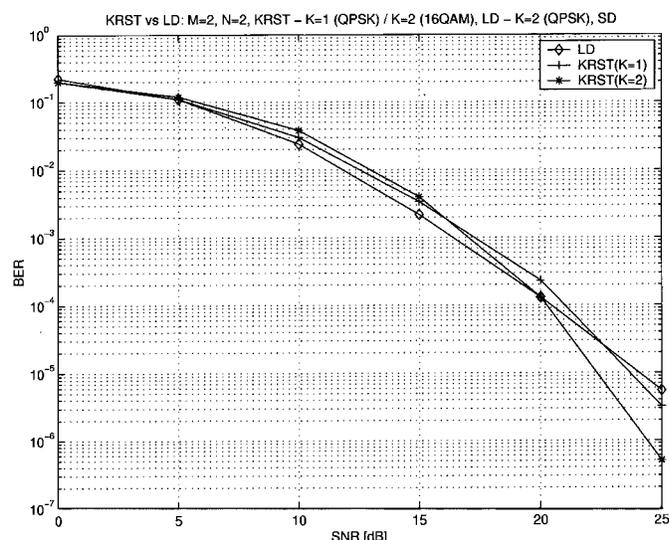


Fig. 4. Comparison of KRST and LD codes: $M = 2, N = 2, \text{KRST}-K = 1 (\text{QPSK}) / K = 2 (16\text{QAM}), \text{LD}-K = 2 (\text{QPSK}), R = 4$.

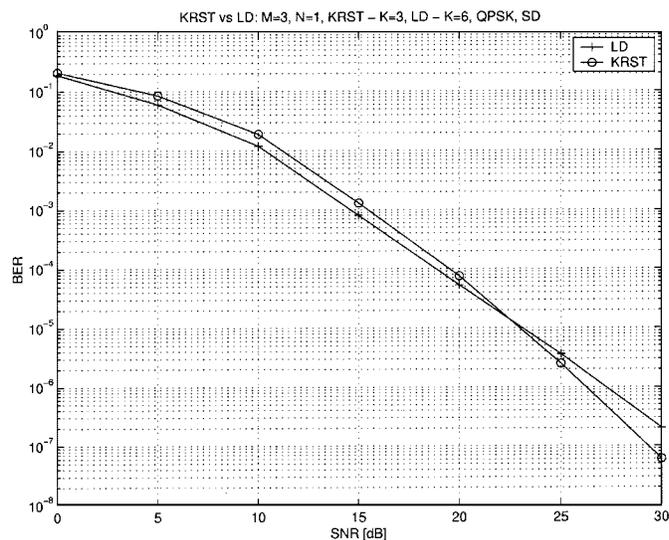


Fig. 5. Comparison of KRST and LD codes: $M = 3, N = 1, \text{KRST}-K = 3, \text{LD}-K = 6, \text{QPSK}, R = 2$.

the page. KRST uses $K = 3$ in order to maintain the rate $R = 2$. The LD code (code design: [15, p. 38, eq (39)]) with $K = 6$ is used for comparison, as it gives better performance than the LD code with $K = 4$ [15]. The comparison is given in Fig. 5. Again, at high SNR, the KRST code gives lower BER than the LD code. The KRST code offers full diversity $NK = 3$, whereas it can be shown that the error matrix \mathbf{B} [cf. (9)] for the LD code is rank deficient for $\mathbf{s}_t = \sqrt{1/2}[1-j, -1-j, -1-j, 1+j, -1-j, -1-j]^T$ and $\tilde{\mathbf{s}}_t = \sqrt{1/2}[1+j, -1-j, -1+j, 1+j, -1+j, -1-j]^T$.

$$\Theta = \begin{bmatrix} 0.6867 & 0.5133 - 0.1125j & -0.4275 + 0.2643j \\ -0.3578 - 0.3076j & 0.6962 - 0.1720j & -0.0110 - 0.5128j \\ 0.1895 + 0.5195j & 0.2418 - 0.3891j & 0.6959 \end{bmatrix}. \tag{45}$$

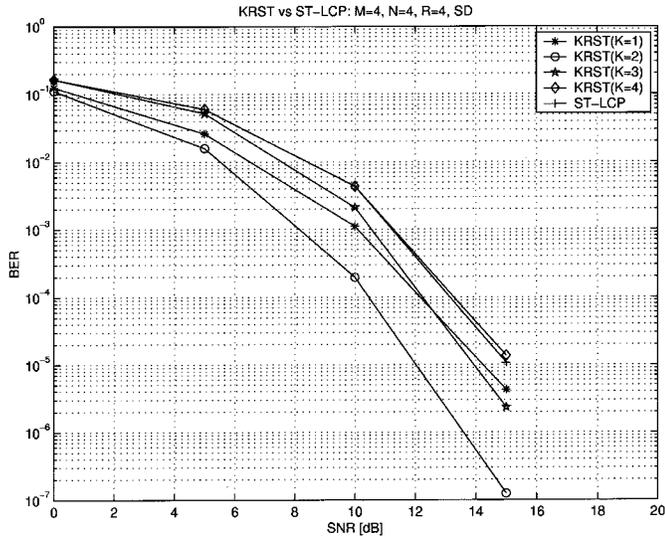


Fig. 6. Comparison of KRST and ST-LCP codes: $M = 4$, $N = 4$, $R = 4$.

D. KRST versus ST-LCP: $M = 4$, $N = 4$

A comparison between KRST codes and ST—linear constellation precoding (ST-LCP) codes [32] is made as both codes enjoy maximum diversity KN ($K = M$ for ST-LCP), and both employ the CR matrix Θ . If $M = 4$ and $N = 4$, ST-LCP codes can achieve $R = 4$ if 16QAM symbols are transmitted. KRST codes can achieve the same rate using $K = 1, 2, 3$, or 4 and BPSK, QPSK, 8QAM, or 16QAM symbols, respectively. The CR matrix Θ to be used is the same as the one used in Section VI-A. It is observed from Fig. 6 that the KRST code with $K = 2$ outperforms all the other codes and gains about 3 dB over the ST-LCP code.

E. Blind CSI Recovery vs DSTM: $M = 4$, $N = 4$

The blind CSI recovery scheme is compared with differential ST modulation (DSTM) [18]–[20]. In order to have a meaningful comparison, we employ the same number of transmit and receive antennas and maintain the same rate of transmission. Since $K = M$ in DSTM, we also use $K = M$ for the KRST code. In this simulation, 10^5 independent random realizations of \mathbf{H} are used, and the channel is assumed to remain constant for a duration of 1000 block-time slots. The case where the channel varies every time slot is considered in Section VI-H. Consider the case where we have four transmit and four receive antennas. In order to maintain a rate $R = 1$, DSTM has to use 8PSK constellation (code design: [20, table 3, p. 166]), whereas KRST uses BPSK. Thirty slabs are used to fit the PARAFAC model. A comparison of these two codes and a training-based scheme is presented in Fig. 7. The training-based scheme first obtains a LS channel estimate based on training data of duration 500 block-time slots. This is followed by payload KRST transmission for the remaining 500 block-time slots where the channel remains constant. In order to maintain $R = 1$, we transmit QPSK symbols during this time. The receiver uses the LS channel estimate coupled with SD to detect the transmitted symbols. Observe that all three designs give similar performance, but the training-based scheme assumes that

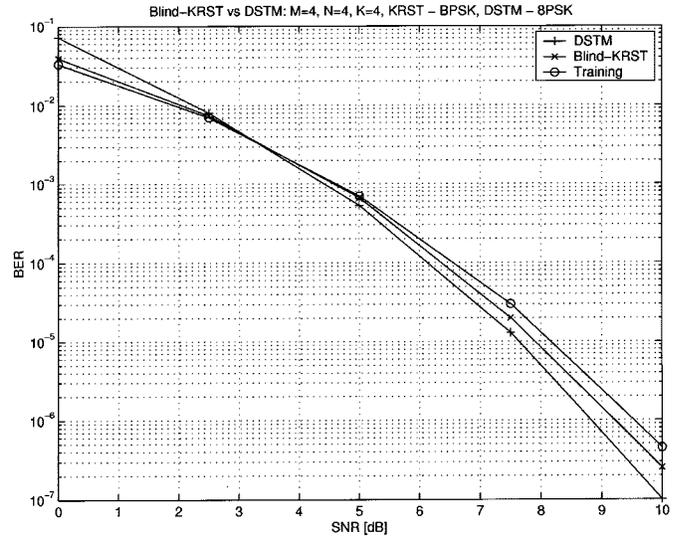


Fig. 7. Comparison of blind-KRST and DSTM codes: $M = 4$, $N = 4$, $K = 4$, blind-KRST—BPSK, DSTM—8PSK, $R = 1$.

the channel remains constant for a very long time, which is usually not the case in a practical setting. The KRST advantage is that the blind receiver for KRST uses SD, which has lower complexity compared with the enumeration-based differential ML receiver employed in DSTM. The latter has an estimator-correlator interpretation and its complexity increases exponentially in M and R . Blind-KRST can achieve higher rates by simply selecting $K = 2$ or 3, whereas DSTM can achieve higher rates only by increasing the constellation size, which further increases the decoding complexity. A simpler near-ML receiver based on lattice decoding has recently been proposed for *diagonal* DSTM codes in [7]. The performance of this algorithm is quite close to ML, and its complexity is polynomial in M and R . However, the DSTM codes used in this simulation are not diagonal (DSTM codes are not diagonal in general), and therefore, this simpler algorithm cannot be used. Note that a constant channel is the ideal scenario for DSTM; hence, we have the slight advantage of DSTM over KRST in Fig. 7. When the channel varies continuously even at a moderate rate (e.g., due to Doppler), the situation can be reversed; see, e.g., the results in Fig. 11, where KRST clearly outperforms DSTM.

F. Blind CSI Recovery: Number of Slabs

The accuracy of our blind CSI recovery technique depends on the number of slabs in the PARAFAC model. Consider the case $M = 4$, $N = 4$, and $K = 4$. The transmitted symbols are selected from the BPSK constellation. Fig. 8 illustrates a comparison between two different slab-sizes $T = 30$ and $T = 5$. Note that $T = 5$ is not appreciably worse than $T = 30$ in terms of final BER. This suggests that the algorithm can handle even moderate mobility without significant degradation. Fig. 9 illustrates the same point for $M = 4$, $N = 4$, $K = 3$, BPSK, and $T = 5, 10$, and 30.

G. Channel Tracking

The performance of the channel tracking scheme for slowly varying channels, which was discussed in Section V, is now

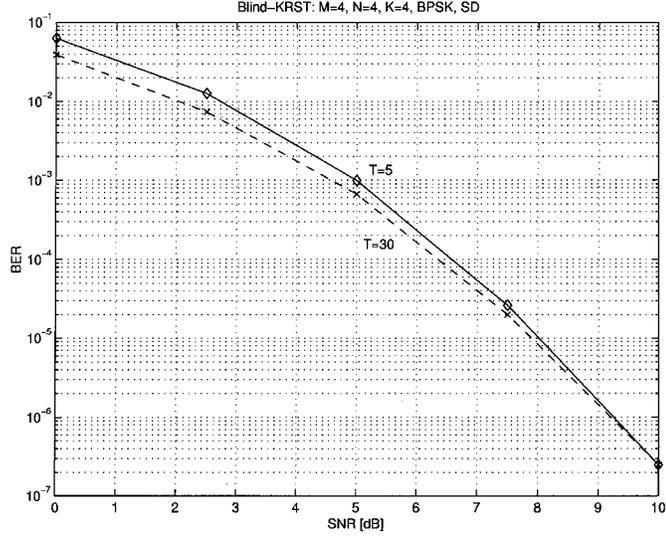


Fig. 8. Number of slabs in blind-KRST: $M = 4$, $N = 4$, $K = 4$, $T = 5(30)$, BPSK.

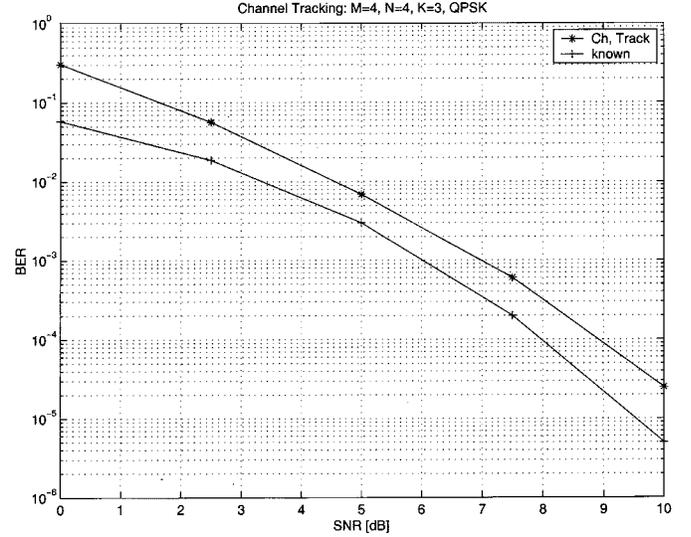


Fig. 10. Channel tracking: $M = 4$, $N = 4$, $K = 3$, QPSK.

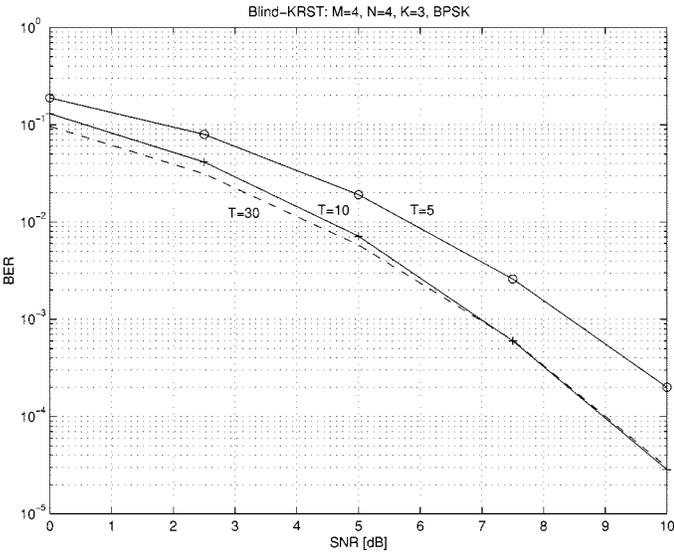


Fig. 9. Number of slabs in blind-KRST: $M = 4$, $N = 4$, $K = 3$, $T = 5(10, 30)$, BPSK.

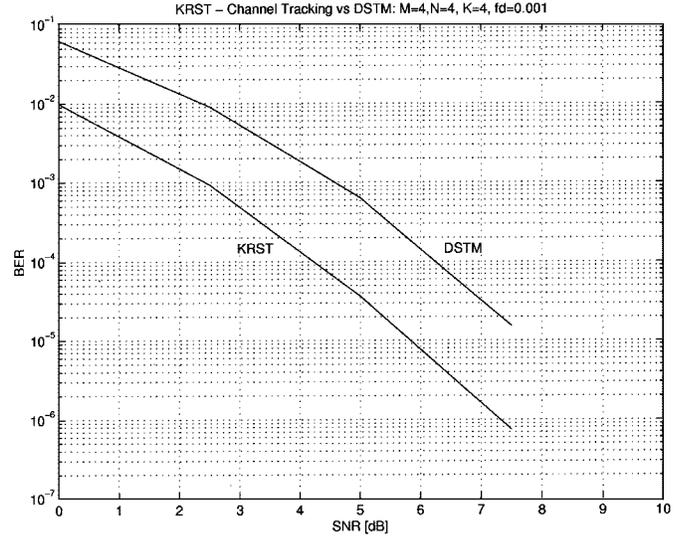


Fig. 11. Comparison of KRST channel tracking and DSTM codes: $M = 4$, $N = 4$, $K = 4$, $f_d = 0.001$.

compared with that of the coherent detection scheme that assumes perfect CSI at the receiver. Fig. 10 illustrates the comparison for the case where we have $M = 4$, $N = 4$, and $K = 3$. The transmitted symbols are selected from the QPSK constellation. In this simulation, 10^7 independent random realizations of the channel matrix \mathbf{H} are used, and the channel taps are modified at each block-time instant p as

$$h_{m,n}(p) = h_{m,n}(p-1)e^{j2\pi f_d} \quad (46)$$

where f_d is the sum of the carrier frequency offset, and the Doppler shift at the n th receive antenna and is taken to be 0.001. It is observed that the performance of the channel tracking scheme degrades by only a few decibels relative to the fully coherent case.

H. Time-Selective Fading Channels $M = 4$, $N = 4$

Consider the case when frequency-offset induced time-selective channels are present. These channels are allowed to vary as fast as one symbol duration. The channel from the m th transmit antenna to the n th receive antenna is modeled as (cf. e.g., [24])

$$h_{m,n}(t) = h_{m,n}(0)e^{j2\pi f_d t} \quad (47)$$

where f_d is the sum of the carrier frequency offset and the Doppler shift at the n th receive antenna. We will employ the KRST channel tracking scheme in this scenario and compare the result with that obtained using DSTM. Consider the case where $M = 4$, $N = 4$, and $f_d = 0.001$. Once again, rate $R = 1$ is maintained by using 8PSK for DSTM and BPSK for KRST. It is observed from Fig. 11 that KRST outperforms DSTM by about 2 dB in this case.

VII. CONCLUSIONS

A novel linear space-time block coding technique based on the Khatri-Rao matrix product has been proposed. It can be applied to any configuration of transmit and receive antennas and any symbol constellation. The performance analysis and design criteria were described, and it has been shown that KRST codes offer flexibility, which can be used to achieve full-rate or full-diversity or desired rate-diversity tradeoffs in between these two extremes. The proposed codes have attractive features like guaranteed unique linear decodability and built-in blind channel identifiability. KRST codes were compared with LD codes, and it has been shown that KRST codes outperform LD codes at high SNR. From the comparison with ST-LCP codes, it has been observed that KRST codes achieve the same rate using a lower order constellation, yielding better performance. The results show that the performance of the blind-KRST decoding scheme is comparable with that of DSTM, but KRST is simpler to decode optimally. The proposed decision-directed channel tracking procedure offers a simple and effective alternative to continuous blind channel re-estimation, bringing the complexity of maintaining CSI well under that of ML block decoding.

APPENDIX

We will need the following well-known Lemma.

A. Lemma 1

Consider an analytic function $f(\mathbf{x})$ of several variables $\mathbf{x} = [x_1, \dots, x_n]^T \in \mathbb{C}^n$. If f is nontrivial in the sense that there exists $\mathbf{x}_0 \in \mathbb{C}^n$ such that $f(\mathbf{x}_0) \neq 0$, then the zero set of $f(\mathbf{x})$

$$\mathcal{Z} := \{\mathbf{x} \in \mathbb{C}^n | f(\mathbf{x}) = 0\}$$

is of measure (Lebesgue measure in \mathbb{C}^n) zero. A simple proof of this lemma is given in [22].

B. Proposition 1

If $KN \geq M$ and \mathbf{C}_0 is chosen to be a $K \times M$ Vandermonde matrix with generators $e^{j2\pi(m-1)/M}$, i.e., $c_{k,m} = e^{j2\pi(m-1)/M(k-1)}$, $m = 1, \dots, M$ (as in Section III-B2), then $r_{\mathbf{C}_0 \odot \mathbf{H}} = k_{\mathbf{C}_0 \odot \mathbf{H}} = M$ for almost every \mathbf{H} (where, for any given matrix \mathbf{A} , $r_{\mathbf{A}} = \text{rank of } \mathbf{A}$, and $k_{\mathbf{A}} = k\text{-rank of } \mathbf{A}$).

C. Proof

First, note that $\mathbf{C}_0 \odot \mathbf{H}$ and $\mathbf{H} \odot \mathbf{C}_0$ are the same modulo a permutation of rows; hence, $r_{\mathbf{C}_0 \odot \mathbf{H}} = r_{\mathbf{H} \odot \mathbf{C}_0}$. It suffices to show that the upper square part of $\mathbf{H} \odot \mathbf{C}_0$ is nonsingular, i.e., its determinant is nonzero for almost every \mathbf{H} . This determinant is a polynomial function of the entries of \mathbf{H} and, hence, is analytic. It suffices to show that it is also nontrivial. This only requires finding a specific \mathbf{H} for which the said determinant is nonzero. The key idea is as follows. Select \mathbf{H} to be a Vandermonde matrix with generators $e^{j2\pi(K(m-1)/M)}$, i.e., $h_{n,m} = e^{j2\pi(K(m-1)/M)(n-1)}$. This yields a Vandermonde Khatri-Rao product $\mathbf{H} \odot \mathbf{C}_0$ with generators $e^{j2\pi(m-1)/M}$, $m = 1, \dots, M$,

whose upper square part is nonsingular. Invoking the analytic function Lemma, $r_{\mathbf{C}_0 \odot \mathbf{H}} = k_{\mathbf{C}_0 \odot \mathbf{H}} = M$ for almost every \mathbf{H} (i.e., except for a measure zero subset of \mathbb{C}^{NM}). ■

REFERENCES

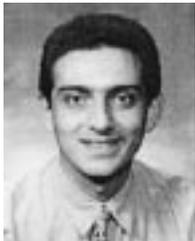
- [1] "CDMA 2000 Candidate Submission," TIA 45.5 Subcommittee, 1998.
- [2] "Space-time block coded transmit antenna diversity for WCDMA," Texas Instruments Inc., Helsinki, Finland, UMTS SMG2-L1, Tech. Doc. 662/98, 1998.
- [3] S. M. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.
- [4] J. Boutros and E. Viterbo, "Signal space diversity: A power and bandwidth efficient diversity technique for the Rayleigh fading channel," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1453–1467, July 1998.
- [5] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, pp. 772–781, Sept. 1978.
- [6] R. Bro, "PARAFAC: Tutorial and applications," *Chemometr. Intell. Lab. Syst.*, vol. 38, pp. 149–171, 1997.
- [7] K. L. Clarkson, W. Sweldens, and A. Zheng, "Fast multiple-antenna differential decoding," *IEEE Trans. Commun.*, vol. 49, pp. 253–261, Feb. 2001.
- [8] M. O. Damen, K. Abed-Meriam, and J.-C. Belfiore, "A generalized lattice decoder for asymmetric space-time communication architecture," in *Proc. ICASSP*, vol. 5, June 2000, pp. 2581–2584.
- [9] M. O. Damen, A. Chkief, and J.-C. Belfiore, "Lattice codes decoder for space-time codes," *IEEE Commun. Lett.*, vol. 4, pp. 161–163, May 2000.
- [10] V. M. DaSilva and E. S. Sousa, "Fading-resistant modulation using several transmitter antennas," *IEEE Trans. Commun.*, vol. 45, pp. 1236–1244, Oct. 1997.
- [11] A. Duel-Hallen, "A family of multiuser decision-feedback detectors for asynchronous code-division multiple-access channels," *IEEE Trans. Commun.*, vol. 43, pp. 421–434, Feb./Mar./Apr. 1995.
- [12] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.
- [13] X. Giraud, E. Boutillon, and J.-C. Belfiore, "Algebraic tools to build modulation schemes for fading channels," *IEEE Trans. Inform. Theory*, vol. 43, pp. 938–952, May 1997.
- [14] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electron. Lett.*, vol. 17, Nov. 1998.
- [15] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, Aug. 2000, submitted for publication.
- [16] B. Hassibi, A. Shokrollahi, B. Hochwald, and W. Sweldens, "Representation theory for high-rate multiple-antenna code design," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2335–2367, Sept. 2001.
- [17] B. Hassibi and H. Vikalo, "On the expected complexity of sphere decoding," in *35th Asilomar Conf. Signals, Syst., Comput.*, vol. 2, Nov. 2001, pp. 1051–1055.
- [18] B. Hochwald and W. Sweldens, "Differential unitary space time modulation," *IEEE Trans. Commun.*, vol. 48, pp. 2041–2052, Dec. 2000.
- [19] B. L. Hughes, "Differential space-time modulation," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2567–2578, Nov. 2000.
- [20] —, "Further results on differential space-time modulation," in *Proc. IEEE Sensor Array Multichan. Signal Process. Workshop*, Cambridge, MA, Mar. 2000, pp. 163–167.
- [21] L. Jalloul, K. Rohani, K. Kuchi, and J. Chen, "Performance analysis of CDMA transmit diversity methods," in *Proc. Veh. Technol. Conf.*, Fall 1999, pp. 1326–1330.
- [22] T. Jiang, N. D. Sidiropoulos, and J. M. F. ten Berge, "Almost sure identifiability of multidimensional harmonic retrieval," *IEEE Trans. Signal Processing*, vol. 49, pp. 1849–1859, Sept. 2001.
- [23] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Appl.*, vol. 18, pp. 95–138, 1977.
- [24] Z. Liu, G. B. Giannakis, and B. L. Hughes, "Double differential space-time block coding for time-selective fading channels," *IEEE Trans. Commun.*, vol. 49, pp. 1529–1539, Sept. 2001.
- [25] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind parafac receivers for DS-CDMA systems," *IEEE Trans. Signal Processing*, vol. 48, pp. 810–823, Mar. 2000.

- [26] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Signal Processing*, vol. 48, pp. 2377–2388, Aug. 2000.
- [27] A. Stamoulis, G. B. Giannakis, and A. Scaglione, "Block FIR decision-feedback equalizers for filterbank precoded transmissions with blind channel estimation capabilities," *IEEE Trans. Commun.*, vol. 49, pp. 69–83, Jan. 2001.
- [28] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.
- [29] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456–1467, July 1999.
- [30] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Euro. Trans. Telecom.*, vol. 10, pp. 585–595, Nov. 1999.
- [31] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1639–1642, July 1999.
- [32] Y. Xin, Z. Wang, and G. B. Giannakis, "Space-time diversity systems based on linear constellation precoding," *IEEE J. Select. Areas Commun.*, Mar. 2001, submitted for publication.



Ramakrishna S. Budampati (S'01) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, in 2000 and the M.S. degree in electrical engineering from the University of Minnesota, Minneapolis, in January 2002. He is currently pursuing the Ph.D. degree in electrical engineering at the University of Minnesota, Minneapolis.

His research interests include space-time coding, multiuser detection, and multicarrier and blind channel estimation algorithms.



Nicholas D. Sidiropoulos (M'92–SM'99) received the Diploma degree 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 (UMCP), 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), UMCP. From September 1992 to June 1994, he served his military service as a Lecturer with the Hellenic Air Force Academy. From October 1993 to June 1994, he was a Member of the Technical Staff, Systems Integration Division, G-Systems Ltd., Athens, Greece. From 1994 to 1995, he held a Postdoctoral position and, from 1996 to 1997, a Research Scientist position at ISR-UMCP before joining the Department of Electrical Engineering, 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 networking.

Dr. Sidiropoulos is a member of the Signal Processing for Communications Technical Committee (SPCOM-TC) of the IEEE SP Society and currently serves as Associate Editor for *IEEE TRANSACTIONS ON SIGNAL PROCESSING* and the *IEEE SIGNAL PROCESSING LETTERS*. He received the NSF/CAREER award in the Signal Processing Systems Program in June 1998.