Self-Matching Space-Time Block Codes for Matrix Kalman Estimator-Based ML Detector in MIMO Fading Channels

Stephen Lam, K. N. (Kostas) Plataniotis, *Senior Member, IEEE*, and Subbarayan Pasupathy, *Life Fellow, IEEE*

Abstract—This paper presents a unifying framework for designing a joint channel-estimation-and-data-detection (CE/DD) scheme and space-time block code (STBC) that improves the performances in a multiple-input–multiple-output (MIMO) slow flat Rayleigh fading channel. Modeling the channel using the continuous-fading model, a matrix state-space model, which naturally represents the temporal and spatial dimensions of a MIMO system, is introduced. A consistent and novel matrix CE/DD scheme is developed using a matrix Kalman filter and a matrix normalized-innovations-based maximum-likelihood detector. In MIMO CE/DD in multiplicative fading, symmetric STBCs (S-STBCs) cause isometric data sequences, which lead to a detection error floor. Motivated by the minimization of the probability of error, two asymmetric STBCs are introduced to be used with these S-STBCs to mitigate isometry. To further improve detection performance, a self-matching STBC (SM-STBC), which mitigates isometry using asymmetry, improves estimation performance using training, and improves detection performance by adapting its code properties, is introduced. This SM-STBC generalizes a limited version that was previously proposed. A comprehensive analysis, which is supported by some simulation studies, indicates that the proposed framework of transceiver and STBC designs offers excellent detection performance.

Index Terms—Asymmetric space-time block code (STBC), continuous-fading model (CFM), innovations, matrix Kalman filtering, self-matching STBC (SM-STBC).

I. INTRODUCTION

MULTI-INPUT–MULTI-OUTPUT (MIMO) systems are gaining in popularity due to their higher capacity [1]–[12]. Since perfect channel state information (CSI), which is necessary for coherent detection, is generally unavailable and differential demodulation reduces capacity and degrades detection performance, a joint channel-estimation-and-data-detection (CE/DD) scheme is needed [2], [13]–[16]. The CE/DD schemes in [2] and [13] estimate the channel using pilots and assume that the fading coefficients remain constant between successive pilot matrices. This assumption is even more restrictive than the block-fading model (BFM), where

the fading coefficients remain constant for one block only. Thus, the channel estimates are already outdated when detection occurs, and the detection performance degrades. To minimize probability of error P_e, a decision-directed CE/DD scheme, where the channel is modeled using the continuous-fading model (CFM) and the estimates are recursively updated, is needed. In the decision-directed CE/DD schemes in [15] and [16], the CSI is represented as vectors in the channel estimations (CEs) but as matrices in the data detections (DDs). This inconsistency between the CSI representations requires the CSI to be converted back and forth between their vector and matrix forms. In addition, the vector form of the CSI does not depict the temporal and spatial nature of MIMO systems. For a consistent and natural MIMO representation, a matrix state-space model (MSSM), where the temporal and spatial dimensions are intuitively represented by matrices, and its corresponding matrix CE/DD scheme, where matrix quantities are consistently used, are needed.

A novel decision-directed CE/DD scheme is proposed and analyzed in this paper: Using an MSSM based on the CFM, the proposed scheme recursively estimates the channel using the matrix Kalman filter (MKF) and detects the data using the normalized-innovations-based maximum likelihood (ML) detector. An ordinary Kalman filter (KF) only works with vector state-space models (VSSMs). Hence, this paper introduces the technique that was originally developed in [17] and the definition of the covariance of a matrix in [18] to derive the MKF for the MSSM. The DD is shown to be a Mahalanobis (i.e., weighted Euclidean) distance square DD, where the weighting factor represents the effective signal-to-noise ratio (SNR) and is calculated recursively by the MKF.

The P_e of the CE/DD is derived and is shown to be linked to the properties of not only the model and CE/DD but also to the space-time block code (STBC). Commonly used linear unitary STBCs are shown to be optimal when CSI is known [5], [11], [12]. However, they are "symmetric," and symmetry causes isometric data sequences and detection error floor in CE/DD in multiplicative fading [19]–[21]. To minimize P_e, pilot training can be used to mitigate isometry, but it reduces transmission rate [21]. This paper introduces information-bearing asymmetric STBCs (A-STBCs), which were developed from symmetric STBCs (S-STBCs), and shows how A-STBCs, in conjunction with S-STBCs, can be used to mitigate isometry. Hence, A-STBC can be seen as a generalized form of training. To further minimize P_e, a self-matching STBC (SM-STBC), which

is asymmetric and adapts its code properties according to the estimation performance to reduce estimation and detection error, is proposed. The SM-STBC in this paper generalizes the one in [13] by replacing the transmission-rate-reducing fixed unitary matrix with the information-bearing A-STBC.

The unifying framework of CE/DD and STBC designs based on CFM in order to minimize P_e is the major contribution in this paper. The channel and system models are derived in Section II. In Sections III and IV, a CE/DD is proposed, and its corresponding P_e is derived. Properties of STBCs affecting P_e, including symmetry, are explained in Section V. Section VI presents the use of A-STBCs and SM-STBCs to reduce P_e. Simulation results are discussed in Section VII, and conclusions are drawn in Section VIII.

II. SYSTEM MODEL

In the succeeding sections, *scalars* are in italic, **vectors** are in lowercase bold, **MATRICES** are in uppercase bold, $\mathbf{I}$ is the identity matrix, $\mathbf{0}$ is the zero matrix, $E\{\cdot\}$ is the expectation, $\cdot^*$ is the conjugate, $\cdot^H$ is the Hermitian transpose, $\mathrm{rank}(\cdot)$ is the rank, $\mathrm{tr}(\cdot)$ is the trace, $\det(\cdot)$ is the determinant, $\mathrm{diag}(\cdot)$ is the diagonal matrix, $\mathrm{vec}(\cdot)$ is the vectorization operation[1] [17], $\otimes$ is the Kronecker product [22], $\|\cdot\|_F$ is the Frobenius norm [5], $\mathbf{M} \sim \mathcal{CN}(\bar{\mathbf{M}}, \boldsymbol{\Sigma})$ denotes a circularly symmetric (proper) complex Gaussian distribution with mean $\bar{\mathbf{M}} = E\{\mathbf{M}\}$ and covariance $\boldsymbol{\Sigma} = \mathrm{cov}(\mathbf{M}) = E\{\mathrm{vec}(\mathbf{M} - \bar{\mathbf{M}})\mathrm{vec}(\mathbf{M} - \bar{\mathbf{M}})^H\}$ [18], [23], [24], a matrix $\mathbf{M}$ is unitary if $\mathbf{M}^H\mathbf{M} = \mathbf{I}$, a $r \times c$ STBC $\mathbf{C}$ is unitary if $\mathbf{C}^H\mathbf{C} = r\mathbf{I}$, $\mathbf{M}_1^L$ is the sequence of matrices $[\mathbf{M}(1), \ldots, \mathbf{M}(L)]$, $\mathbf{M}_1^L\mathbf{N}$ is $[\mathbf{M}(1)\mathbf{N}, \ldots, \mathbf{M}(L)\mathbf{N}]$, and $\mathbf{N}\mathbf{M}_1^L$ is $[\mathbf{N}\mathbf{M}(1), \ldots, \mathbf{N}\mathbf{M}(L)]$.

In flat Rayleigh fading, the theoretical capacity of a MIMO system with M transmit and N receive antennas grows linearly with $\min(M, N)$ [5], [25], [26]. Hence, MIMO systems are preferred over single-input–single-output (SISO) systems for high-data-rate applications. Consider the MIMO system in Fig. 1 for slow flat Rayleigh fading; matched filtering and symbol-rate sampling are performed to obtain sufficient statistics, and the observation of the nth receive antenna at the kth symbol is

$$y_n(k) = \sum_{m=1}^{M} x_m(k)h_{m,n}(k) + v_n(k) \qquad \forall n = 1, \ldots, N \tag{1}$$

where $x_m(k)$ is the transmitted symbol from the mth antenna, $h_{m,n}(k) \sim \mathcal{CN}(0, 1)$ is the fading coefficient from the mth transmit antenna to the nth receive antenna, and $v_n(k) \sim \mathcal{CN}(0, r)$ is the additive white Gaussian noise (AWGN) [21], [27]. Instead of assuming that the fading is time invariant, this paper considers a more realistic scenario: The MN fading coefficients are temporally correlated but spatially independent and identically distributed (IID), i.e., $E\{h_{a,b}(k)h_{c,d}^*(k + \tau)\} = \delta(a - c)\delta(b - d)J_0(2\pi\tau f_D T_s)$ $\forall 1 \le a$, $c \le M$ and $\forall 1 \le b, d \le N$, where $\delta(k)$ is 1 if $k = 0$ and 0 otherwise, and $f_D T_s$ is the normalized fading rate [16]. The assumption of spatial independence holds when the antennas are sufficiently spaced [28].

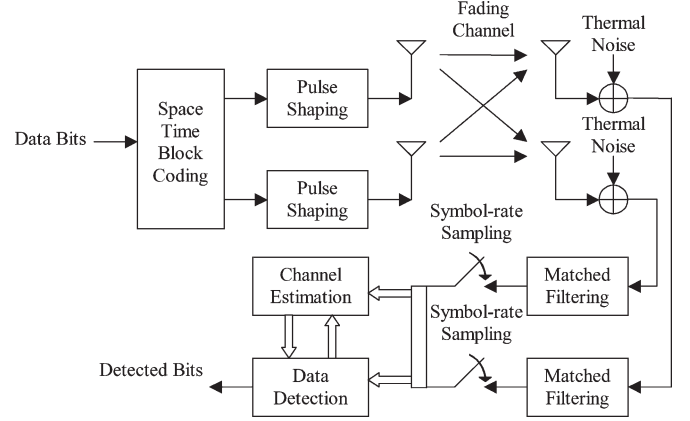[1]The vector is formed by stacking the columns of the matrix together.

Fig. 1. System diagram of the MIMO CE/DD in Rayleigh fading channels.

In addition, the AWGN samples are IID, i.e., $E\{v_b(k)v_d^*(k + \tau)\} = \delta(b - d)\delta(\tau)r$. It is assumed that $E\{|x_m(k)|^2\} = 1$; thus, the SNR is $\rho = M/r$. With the matrices that are defined in column 1 of Table I, the observations over N receive antennas and T symbol durations in (1) are collected as [21], [29]

$$\mathbf{Y}(l) = \mathbf{X}(l)\mathbf{H}(l) + \mathbf{V}(l). \tag{2}$$

Remark—Fading Model: The CFM, not BFM, is used in this paper to derive (2) [30]. BFM assumes that each fading coefficient is constant over the block of T symbols (i.e., $T < T_c$, where T_c is the channel coherence time) and adjacent blocks experience uncorrelated fading [1], [7]. However, in a realistic channel, fading changes gradually without piecewise jumps [14], [16]. Table I compares the definitions of the matrices in (2) for both CFM and BFM, and it shows that the BFM is simply a special case of the CFM when $h_{m,n}((l+1)T - 1) = h_{m,n}((l + 1)T - 2) = \cdots = h_{m,n}(lT + 1) = h_{m,n}(lT) \equiv h_{m,n} \forall m, n$. ∎

Based on the availability of the CSI, three receiver schemes can be used: 1) coherent (perfect CSI); 2) differential (no CSI); and 3) CE/DD (estimated CSI). Realistically, perfect CSI is unavailable, but differential demodulation increases P_e and reduces capacity [5], [13], [14], [25]. The P_e and capacity of the CE/DD scheme approach those of the coherent scheme when the estimation error is small [13], [15], [16], [25]. Thus, the CE/DD in Fig. 1 is considered in this paper.

III. ESTIMATION-ASSISTED RECEIVER

A. CE

Since $h_{m,n}(k)$ and $v_n(k)$, which are the elements of $\mathbf{H}(l)$ and $\mathbf{V}(l)$ in (2), are circularly symmetric complex Gaussian random processes, a second-order linear filter is used to estimate the CSI. Among the various options, a recursive version is preferred over a batch version to track the temporal variation of fading effectively [15], [16], [31]. The KF, which is a recursive adaptive second-order linear estimator, is used in this paper because it is optimal in the minimum mean square error (mse) sense [15], [16], [31], [32].

TABLE I
DEFINITIONS OF THE MATRIX QUANTITIES IN OBSERVATION (2) FOR BOTH CFM AND BFM

	Continuous Fading Model (CFM)	Block Fading Model (BFM)
$\mathbf{Y}(l)$	$\begin{bmatrix} y_1((l+1)T-1) & \dots & y_N((l+1)T-1) \\ \vdots & \ddots & \vdots \\ y_1(lT) & \dots & y_N(lT) \end{bmatrix}$	$\begin{bmatrix} y_1((l+1)T-1) & \dots & y_N((l+1)T-1) \\ \vdots & \ddots & \vdots \\ y_1(lT) & \dots & y_N(lT) \end{bmatrix}$
$\mathbf{X}(l)$	$\text{diag}\left(\begin{bmatrix} x_1((l+1)T-1) \\ \vdots \\ x_1(lT) \end{bmatrix}\right) \cdots \text{diag}\left(\begin{bmatrix} x_M((l+1)T-1) \\ \vdots \\ x_M(lT) \end{bmatrix}\right)$	$\begin{bmatrix} x_1((l+1)T-1) & \dots & x_M((l+1)T-1) \\ \vdots & \ddots & \vdots \\ x_1(lT) & \dots & x_M(lT) \end{bmatrix}$
$\mathbf{H}(l)$	$\begin{bmatrix} h_{1,1}((l+1)T-1) & \dots & h_{1,N}((l+1)T-1) \\ \vdots & \ddots & \vdots \\ h_{1,1}(lT) & \dots & h_{1,N}(lT) \\ \vdots & & \vdots \\ h_{M,1}((l+1)T-1) & \dots & h_{M,N}((l+1)T-1) \\ \vdots & \ddots & \vdots \\ h_{M,1}(lT) & \dots & h_{M,N}(lT) \end{bmatrix}$	$\begin{bmatrix} h_{1,1} & \dots & h_{1,N} \\ \vdots & \ddots & \vdots \\ h_{M,1} & \dots & h_{M,N} \end{bmatrix}$
$\mathbf{V}(l)$	$\begin{bmatrix} v_1((l+1)T-1) & \dots & v_N((l+1)T-1) \\ \vdots & \ddots & \vdots \\ v_1(lT) & \dots & v_N(lT) \end{bmatrix}$	$\begin{bmatrix} v_1((l+1)T-1) & \dots & v_N((l+1)T-1) \\ \vdots & \ddots & \vdots \\ v_1(lT) & \dots & v_N(lT) \end{bmatrix}$

1) MSSM: The use of KF requires a state equation that approximates the temporal variation of fading using a rational hypermodel [15], [16]. The second-order autoregressive (AR-2) model is used due to its excellent tracking performance and reasonable complexity [15], [16], [31]–[34]:

$$h_{m,n}(k) = -a_1 h_{m,n}(k-1) - a_2 h_{m,n}(k-2) + w_{m,n}(k) \tag{3}$$

where $w_{m,n}(k) \sim \mathcal{CN}(0,\beta)$ is the driving noise, $E\{w_{a,b}(k)w_{c,d}^*(k+\tau)\} = \delta(a-c)\delta(b-d)\delta(\tau)\beta$, $a_1 = -2r_d\cos(\sqrt{2}\pi f_D T_s)$, $a_2 = r_d^2$, $r_d = 1 - 0.2\pi f_D T_s$, and $\beta = ([(1+a_2)^2 - a_1^2](1-a_2))/(1+a_2)$. Recursively applying (3), the fading coefficients over T symbol durations are related as follows [21], [29]:

$$\mathbf{h}_{m,n}(l) = \tilde{\mathbf{A}}\mathbf{h}_{m,n}(l-1) + \mathbf{w}_{m,n}(l) \tag{4}$$

where, for illustration, the specifics of $\tilde{\mathbf{A}}$, $\mathbf{h}_{m,n}(l)$, and $\mathbf{w}_{m,n}(l)$ for $T = 2$ and $T = 4$ are given in Table II. Then, the MN fading paths that are described by (4) can be collected as [21], [29]

$$\underbrace{\begin{bmatrix} \mathbf{h}_{1,1}(l) & \cdots & \mathbf{h}_{1,N}(l) \\ \vdots & \ddots & \vdots \\ \mathbf{h}_{M,N}(l) & \cdots & \mathbf{h}_{M,1}(l) \end{bmatrix} \cdots}_{\mathbf{H}(l)}$$

$$= \underbrace{(\mathbf{I} \otimes \tilde{\mathbf{A}})}_{\mathbf{A}} \mathbf{H}(l-1) + \underbrace{\begin{bmatrix} \mathbf{w}_{1,1}(l) & \cdots & \mathbf{w}_{1,N}(l) \\ \vdots & \ddots & \vdots \\ \mathbf{w}_{M,1}(l) & \cdots & \mathbf{w}_{M,N}(l) \end{bmatrix}}_{\mathbf{W}(l)} \tag{5}$$

where $\mathbf{H}(0)$ is the initial condition. The observation (2), the state (5), the initial condition ($\mathbf{H}(0)$), and the statistics ($E(\mathbf{H}(l)) = \mathbf{0}$, $\text{cov}(\mathbf{H}(l))$, $E(\mathbf{V}(l)) = \mathbf{0}$, $\text{cov}(\mathbf{V}(l)) = r\mathbf{I} = $ $\mathbf{R}$, $\mathbf{A}$, $E(\mathbf{W}(l)) = \mathbf{0}$, $\text{cov}(\mathbf{W}(l)) = \mathbf{Q})$ altogether constitute the MSSM. The matrix equations describe the dynamics of the MSSM, and the definition of the covariance of a matrix allows us to describe the statistics of the MSSM [18]. Since the model statistics will be used in the CE/DD, its value and accuracy will affect the performance of the CE/DD.

Remark—Matrix Nature of the State-Space Model: Many state-space models have been used in the literature: Some employ matrices in the observation equation and DD but vectors in the state equation and CE [15], [16]; some employ vectors in both (i.e., a vectorized model) [35]; some formulate matrix models by imposing restrictions on the number of receive antennas, the order of hypermodel, or the temporal correlation of fading [2], [13], [16], [28], [35], [36]; some avoid the spatial dimension, thus transmit diversity, by using only one receive antenna [2], [35]; and some avoid the temporal dimension, thus temporal diversity, by using vector modulation instead of matrix modulation [37], [38]. In MIMO systems, the data are sent and received via multiple antennas and symbol durations. A matrix captures this by designating its row and column dimensions as the time and space dimensions. Thus, the MSSM in this paper allows a more intuitive, insightful, consistent, and compact formulation of the MIMO system and the resulting CE/DD scheme [17]. In addition, it can be used with any number of transmit and receive antennas, different hypermodels, time-variant fading channels, and diversity-achieving matrix modulations. ∎

Remark—Varying Mobile Environment: In a realistic mobile environment, $f_D T_s$ can vary over time. From (5), the value of $f_D T_s$ changes the MSSM, and the model affects the CE/DD performance; mismatched $f_D T_s$ is undesirable [32], [39]. Using the technique in [40], $f_D T_s$ can be estimated accurately and periodically, and the parameters $\mathbf{A}$ and $\mathbf{Q}$ can be easily updated to adapt to the varying environment. ∎

2) MKF: The use of MSSM dictates that a matrix estimator is desired. Since ordinary KF only works with VSSMs, an

TABLE II
SPECIFICS OF THE QUANTITIES IN STATE (4) FOR $T = 2$ AND $T = 4$

	$T = 2$	$T = 4$
$\mathbf{h}_{m,n}(l)$	$\begin{bmatrix} h_{m,n}(2l+1) \\ h_{m,n}(2l) \end{bmatrix}$	$\begin{bmatrix} h_{m,n}(4l+3) \\ h_{m,n}(4l+2) \\ h_{m,n}(4l+1) \\ h_{m,n}(4l) \end{bmatrix}$
$\tilde{\mathbf{A}}$	$\begin{bmatrix} a_1^2 - a_2 & a_1 a_2 \\ -a_1 & -a_2 \end{bmatrix}$	$\begin{bmatrix} a_1^4 - 3a_1^2 a_2 + a_2^2 & a_1^3 a_2 - 2a_1 a_2^2 & 0 & 0 \\ -a_1^3 + 2a_1 a_2 & -a_1^2 a_2 + a_2^2 & 0 & 0 \\ a_1^2 - a_2 & a_1 a_2 & 0 & 0 \\ -a_1 & -a_2 & 0 & 0 \end{bmatrix}$
$\mathbf{w}_{m,n}(l)$	$\begin{bmatrix} w_{m,n}(2l+1) - a_1 w_{m,n}(2l) \\ w_{m,n}(2l) \end{bmatrix}$	$\begin{bmatrix} w_{m,n}(2l+3) - a_1 w_{m,n}(2l+2) + (a_1^2 - a_2)w_{m,n}(2l+1) \cdots \\ + (-a_1^3 + 2a_1 a_2)w_{m,n}(2l) \\ w_{m,n}(2l+2) - a_1 w_{m,n}(2l+1) + (a_1^2 - a_2)w_{m,n}(2l) \\ w_{m,n}(2l+1) - a_1 w_{m,n}(2l) \\ w_{m,n}(2l) \end{bmatrix}$

MKF is needed. This paper introduces the procedure that was originally developed in [17] to the area of communications and uses the procedure to formulate an MKF specifically to estimate $\mathbf{H}(l)$. The procedure does not lose any information or impose any restriction on the MSSM. First, using the vectorization operator and the Kronecker product [17], [22], (2) and (5) are converted to the *vectorized representation* [35]. Since the matrix quantities are now represented as vectors, the ordinary KF recursion is then applied to the vectorized equations. Converting the vector quantities in the resulting recursion back to matrices, the following MKF is derived [17], [29]:

$$\left[\hat{\mathbf{H}}(l+1|l), \mathbf{P}(l+1|l) \right] \cdots$$
$$= \mathcal{MKF}\left(\hat{\mathbf{H}}(l|l-1), \mathbf{P}(l|l-1), \mathbf{X}(l), \mathbf{Y}(l) \right)$$

$$\tilde{\mathbf{Z}}(l|l-1) = \mathbf{Y}(l) - \mathbf{X}(l)\hat{\mathbf{H}}(l|l-1)$$

$$\mathbf{P}_{\tilde{\mathbf{Z}}}(l|l-1) = (\mathbf{I} \otimes \mathbf{X}(l))\,\mathbf{P}(l|l-1)\,(\mathbf{I} \otimes \mathbf{X}(l))^{\mathrm{H}} + \mathbf{R}$$

$$\mathbf{K}(l) = \mathbf{P}(l|l-1)\,(\mathbf{I} \otimes \mathbf{X}(l))^{\mathrm{H}}\,\mathbf{P}_{\tilde{\mathbf{Z}}}^{-1}(l|l-1)$$

$$\hat{\mathbf{H}}(l+1|l) = \mathrm{vec}^{-1}\left\{ (\mathbf{I} \otimes \mathbf{A})\left[\mathrm{vec}\left(\hat{\mathbf{H}}(l|l-1) \right) \cdots \right. \right.$$
$$\left. \left. + \mathbf{K}(l)\mathrm{vec}\left(\tilde{\mathbf{Z}}(l|l-1) \right) \right] \right\}$$

$$\mathbf{P}(l+1|l) = (\mathbf{I} \otimes \mathbf{A})\left\{ [\mathbf{I} - \mathbf{K}(l)\,(\mathbf{I} \otimes \mathbf{X}(l))]\,\mathbf{P}(l|l-1) \right\} \cdots$$
$$(\mathbf{I} \otimes \mathbf{A})^{\mathrm{H}} + \mathbf{Q} \qquad (6)$$

where $\mathbf{R} = \mathrm{cov}(\mathbf{V}(l)) = r\mathbf{I}$, and $\mathbf{Q} = \mathrm{cov}(\mathbf{W}(l))$. The initial condition and its error covariance are $\hat{\mathbf{H}}(0|0)$ and $\mathbf{P}(0|0) = \mathrm{cov}(\mathbf{H}(0) - \hat{\mathbf{H}}(0|0))$, respectively. Without *a priori* information, $\hat{\mathbf{H}}(0|0)$ is assumed to match the MSSM statistics, i.e., $\hat{\mathbf{H}}(0|0) = \mathrm{E}\{\mathbf{H}(0)\} = \mathbf{0}$. Thus, $\mathbf{P}(0|0) = \mathrm{cov}(\mathbf{H}(0)) \approx \mathbf{I}$.

B. DD

Given the channel state prediction $\hat{\mathbf{H}}(l|l-1)$ and its error covariance $\mathbf{P}(l|l-1)$ that is calculated by the MKF (6), the ML DD is then given as follows [29]:

$$\hat{\mathbf{X}}(l) = \arg\min_{\mathbf{\Xi} \in \mathbf{\Omega_X}} \left\{ \tilde{Z}_P\left(\mathbf{Y}(l), \mathbf{\Xi}, \hat{\mathbf{H}}(l|l-1), \mathbf{P}(l|l-1) \right) \right\} \qquad (7)$$

where

$$\tilde{Z}_P(\mathbf{Y}, \mathbf{X}, \hat{\mathbf{H}}, \mathbf{P}) = \left\{ \mathrm{vec}(\mathbf{Y} - \mathbf{X}\hat{\mathbf{H}})^{\mathrm{H}} \right.$$
$$\left. \times \left((\mathbf{I} \otimes \mathbf{X})\mathbf{P} \cdot (\mathbf{I} \otimes \mathbf{X})^{\mathrm{H}} + \mathbf{R} \right)^{-1} \mathrm{vec}(\mathbf{Y} - \mathbf{X}\hat{\mathbf{H}}) \right\}$$

is the normalized-innovations calculated by the MKF, and $\mathbf{\Omega_X}$ is the set of possible matrices for $\mathbf{X}(l)$. The ML DD is hereafter called the *matrix normalized-innovations-based DD (MNI-DD)*.

Remark—Matrix Normalized Innovations: The metric that is used in (7) is the Mahalanobis distance square, which is the Euclidean distance square that is weighted adaptively by the innovation covariance [41]. Hence, the MNI-DD generalizes the usual Euclidean distance-square-based DD. The innovation covariance represents the effective SNR, which consists of the covariances of the thermal noise and the estimation error. ∎

Remark—Operation of the CE/DD: For CE/DD, the MKF operates in a decision-directed mode, so $\mathbf{X}(l)$ in (6) is replaced by $\hat{\mathbf{X}}(l)$, which is provided by the DD [15], [16]. The MKF provides the MNI-DD with $\hat{\mathbf{H}}(l|l-1)$ for the detection of $\mathbf{X}(l)$, and the MNI-DD provides the MKF with $\hat{\mathbf{X}}(l)$ for the estimation of $\mathbf{H}(l+1|l)$. The CE/DD recursively estimates the channel and detects the data, and the channel estimates are always up to date, unlike those in [2] and [13]. ∎

IV. PROBABILITY OF DETECTION ERROR P_e

For the CE/DD scheme (6) and (7), the P_e in terms of bit error rate (BER) is approximated by the union bound [2] $P_e \leq \sum_{\alpha \neq \beta} P_{\alpha \to \beta}$ (# error bits in the error event/# bits per code matrix), where $P_{\alpha \to \beta}$ is the pairwise error probability (PEP) when $\mathbf{X}_\alpha$ is sent but $\mathbf{X}_\beta$ is detected (see the Appendix) [38], i.e.,

$$P_{\alpha \to \beta} = \frac{1}{\left(1 - \frac{\lambda_{\alpha\beta+}}{\lambda_{\alpha\beta-}}\right)^{2MN-1}} \sum_{l=0}^{MN-1} \binom{2MN-1}{l} \cdot \left(-\frac{\lambda_{\alpha\beta+}}{\lambda_{\alpha\beta-}}\right)^l \qquad (8)$$

and $\lambda_{\alpha\beta+}$, $\lambda_{\alpha\beta-}$ are the positive and negative eigenvalues of $\mathrm{cov}(\boldsymbol{\Upsilon})\boldsymbol{\Theta}_{\alpha,\beta}$, respectively, where

$$
\mathrm{cov}(\boldsymbol{\Upsilon})\boldsymbol{\Theta}_{\alpha,\beta} = \begin{bmatrix} \breve{\mathbf{X}}_\alpha \mathrm{cov}(\mathbf{H}(l)) \, \breve{\mathbf{X}}_\alpha^{\mathrm{H}} + \mathbf{R} & \breve{\mathbf{X}}_\alpha \mathrm{cov}\left(\hat{\mathbf{H}}(l|l-1)\right) \\ \mathrm{cov}\left(\hat{\mathbf{H}}(l|l-1)\right)^{\mathrm{H}} \breve{\mathbf{X}}_\alpha^{\mathrm{H}} & \mathrm{cov}\left(\hat{\mathbf{H}}(l|l-1)\right) \end{bmatrix}
$$
$$
\cdot \begin{bmatrix} \mathbf{P}_{\tilde{\mathbf{Z}},\beta}^{-1} - \mathbf{P}_{\tilde{\mathbf{Z}},\alpha}^{-1} & \mathbf{P}_{\tilde{\mathbf{Z}},\alpha}^{-1}\breve{\mathbf{X}}_\alpha - \mathbf{P}_{\tilde{\mathbf{Z}},\beta}^{-1}\breve{\mathbf{X}}_\beta \\ \breve{\mathbf{X}}_\alpha^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{Z}},\alpha}^{-1} - \breve{\mathbf{X}}_\beta^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{Z}},\beta}^{-1} & \breve{\mathbf{X}}_\beta^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{Z}},\beta}^{-1}\breve{\mathbf{X}}_\beta - \breve{\mathbf{X}}_\alpha^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{Z}},\alpha}^{-1}\breve{\mathbf{X}}_\alpha \end{bmatrix} \quad (9)
$$

$\mathbf{P}_{\tilde{\mathbf{Z}},\alpha} = \breve{\mathbf{X}}_\alpha \mathbf{P}(l|l-1)\breve{\mathbf{X}}_\alpha^{\mathrm{H}} + \mathbf{R}$; $\mathbf{P}_{\tilde{\mathbf{Z}},\beta} = \breve{\mathbf{X}}_\beta \mathbf{P}(l|l-1)\breve{\mathbf{X}}_\beta^{\mathrm{H}} + \mathbf{R}$; $\breve{\mathbf{X}}_\alpha = \mathbf{I} \otimes \mathbf{X}_\alpha$; and $\breve{\mathbf{X}}_\beta = \mathbf{I} \otimes \mathbf{X}_\beta$ [42]. The asymptotic bound of (8) is [38]

$$
\tilde{P}_{\alpha\to\beta} = \lim_{\rho\to\infty} P_{\alpha\to\beta} = \binom{2MN-1}{MN-1}\left(-\frac{\lambda_{\alpha\beta+}}{\lambda_{\alpha\beta-}}\right)^{-MN}.
$$

Remark—Complete and Consistent Package: The MSSM, MKF, MNI-DD, and P_e all naturally fit together by consistently employing the same matrix quantities. P_e is linked, either explicitly or implicitly, to the following: 1) *Modulation*: Weighted distances between $\mathbf{X}_\alpha$ and $\mathbf{X}_\beta$ appear in (9). Hence, properties of the modulation will affect P_e. More discussion follows in Section VI. 2) *Channel model*: The innovation covariance $\mathbf{P}_{\tilde{\mathbf{Z}}}(l|l-1)$ is calculated recursively by the MKF according to the MSSM. Thus, P_e will be affected by the model parameters, e.g., $\mathbf{R}$, $\mathbf{A}$, $\mathbf{Q}$, and the initial conditions $\mathbf{H}(0|0)$ and $\mathbf{P}(0|0)$. Thus, mismatched parameters and initial conditions may increase P_e. 3) *CE performance*: $\mathbf{P}(l|l-1)$, which is calculated by the MKF, indicates the theoretical estimation performance. Thus, factors that degrade the theoretical estimation performance, e.g., high $f_D T_s$, will increase P_e. 4) *Detection of past codewords*: P_e is derived, assuming that $\hat{\mathbf{H}}l|l-1)$ is accurate. If $\hat{\mathbf{X}}(l-1) \neq \mathbf{X}(l-1)$, then $\hat{\mathbf{H}}(l|l-1)$ is inaccurate, and an error is more likely to occur with the detection of $\mathbf{X}(l)$. ∎

V. PERFORMANCE-RELATED PROPERTIES OF STBCs

As discussed in Section IV, the properties of the data matrices, which can be uncoded or coded, will affect P_e. In an uncoded system, deep fading might render the transmitted symbols undetectable. The use of space-time codes exploits transmit diversity to mitigate deep fading [37]. Two types of space-time codes are available: 1) block and 2) trellis, and the former is the focus of this paper. An STBC is a set of $T \times M$ complex matrices $\boldsymbol{\Omega}_{\mathbf{C}} = \{\mathbf{C} \in \mathbb{C}^{T\times M}\}$, where each matrix contains the coded symbols $x_m(k)$'s to be sent over M transmit antennas and T symbol periods. To accommodate continuous fading in (2), each codeword $\mathbf{C}(T \times M)$ needs to be mapped uniquely into a transmit matrix $\mathbf{X}(l)(T \times MT)$. Since the performance of the CE/DD scheme should lie between that of the coherent receiver and that of the differential demodulator, the coherent and differential STBCs serve as the basis for the discussion of code design.

To simplify the STBC design process, BFM and coherent detection are commonly assumed [1], [2], [5], [8], [10]–[12], [30]. Many design criteria (e.g., rank, trace, and determinant)

have been proposed to minimize P_e [2], [5], [8], [10], [11], [30]. Linear unitary STBC minimizes P_e when BFM and coherent detection are assumed [5], [8], [11], [12]. A well-known example for $M = T = 2$ is the Alamouti STBC [1]. Group STBCs, either cyclic or dicyclic, are used for differential demodulation, which requires that the channel be constant for $2T$ symbol durations (i.e., $2T < T_c$) [4], [6], [9]. Group STBCs, which are square but nonlinear, can also be used for coherent detection [7]. However, the absence of CSI increases P_e by a factor of 2^{MN} [14].

Remark—Square Versus Rectangular STBCs: To minimize P_e, maximal transmit diversity is desired, i.e., $T \geq M$ [2], [5], [10], [30]. STBCs can be square ($T = M$) or rectangular ($T > M$) [1], [4]–[6], [9]. So far, only square examples (Alamouti and group) have been given. If a square STBC achieves maximal transmit diversity in BFM, then it also achieves maximal transmit diversity in CFM [30]. Hence, square STBCs serve as a reasonable benchmark. ∎

Remark—P_e of Square Unitary STBCs: To gain better insights about P_e, simplification is needed. Assuming that BFM ($\mathrm{cov}(\mathbf{H}(l)) = \mathbf{I}$, $\mathbf{P}(l|l-1) = \sigma^2\mathbf{I}$, $\mathrm{cov}(\hat{\mathbf{H}}(l|l-1)) = (1-\sigma^2)\mathbf{I}$) and square unitary STBCs ($M = T$, $\mathbf{X}^{\mathrm{H}}\mathbf{X} = \mathbf{X}\mathbf{X}^{\mathrm{H}} = M\mathbf{I}$) are used [43, App. D]

$$
-\frac{\lambda_{\alpha\beta+}}{\lambda_{\alpha\beta-}} = 1 + \frac{T_1}{2MN} + \sqrt{\left(\frac{T_1}{2MN}\right)^2 + \frac{T_1}{MN}} \quad (10)
$$

where $T_1 = \mathrm{tr}\left(\mathrm{cov}\left(\boldsymbol{\Upsilon}\right)\boldsymbol{\Theta}_{\alpha,\beta}\right) = (1-\sigma^2)(N/r + \sigma^2 M)d_s^2$, $d_s^2 = \mathrm{tr}[(\mathbf{X}_\alpha - \mathbf{X}_\beta)^{\mathrm{H}}(\mathbf{X}_\alpha - \mathbf{X}_\beta)] = \|\mathbf{X}_\alpha - \mathbf{X}_\beta\|_{\mathrm{F}}^2$, and σ^2 is the (theoretical) estimation error variance (i.e., estimation mse) [42], [44].

When (10) increases, P_e (8) decreases. Thus, d_s^2, which is the Euclidean distance square between codewords in the "space-time sphere," should be maximized [19], [45], [46]. This criterion is similar to that in SISO systems, where the distances between signal points on the complex plane are maximized [39]. Generally, as M increases, the code rate of a linear square unitary STBC decreases, so there is proportionally more room for the codewords to spread out in the space-time sphere [5]. Thus, as M increases, it should be easier to obtain a larger d_s^2. Using a smaller constellation for the symbols in a linear unitary STBC or using a group STBC with fewer codewords also increases d_s^2. P_e is related to d_s^2 because the CE/DD scheme is also based on Euclidean-distance square: The MKF (6) minimizes the mse between $\mathbf{H}(l)$ and $\hat{\mathbf{H}}(l|l-1)$, and the MNI-DD (7) minimizes the Mahalanobis distance square of the innovations.

Keeping M, N, and d_s^2 constant, T_1 is maximal when $\sigma^2 = 0$. As $\sigma^2 > 0$ increases, T_1 decreases; thus, P_e increases. So, square unitary STBCs work well in coherent detection, but their performances suffer even with a small estimation error.

If (10) ≥ 4, then as MN increases, $(-\lambda_{\alpha\beta+}/\lambda_{\alpha\beta-})^{-MN}$ decreases faster than $\binom{2MN-1}{MN-1}$ increases. First, as N varies, (10) remains constant. Thus, when (10) ≥ 4, increasing N reduces $\tilde{P}_e$ (8) through the exponent. Second, as M increases, (10) generally increases. Hence, as long as (10) ≥ 4, increasing M reduces $\tilde{P}_e$ through both the exponent and (10). ∎

An STBC $\boldsymbol{\Omega_\mathbf{C}}$ is said to be *symmetric* if a unitary rotation leaves the STBC invariant [19], i.e.,

$$\exists \mathbf{U} \neq \mathbf{I}, \qquad \mathbf{U}^\mathrm{H}\mathbf{U} = \mathbf{I}\{\forall \mathbf{C} \in \boldsymbol{\Omega_\mathbf{C}} : \mathbf{C}\mathbf{U}^\mathrm{H} \in \boldsymbol{\Omega_\mathbf{C}}\}. \quad (11)$$

It is easy to see that every linear unitary STBC, if the symbols are drawn from a rotationally invariant constellation, is symmetric [29], [39]. Moreover, all group STBCs are symmetric by definition. Even though all examples that are presented thus far (Alamouti and group) are square and symmetric, an S-STBC can be rectangular, and a square STBC can be asymmetric.

Lemma 1: When an S-STBC is used (i.e., $\mathbf{X}(l), \mathbf{X}(l)\mathbf{U}^\mathrm{H} \in \boldsymbol{\Omega_\mathbf{X}}$), the MKF (6) introduces a counterunitary transform in $\hat{\mathbf{H}}(l+1|l)$ in response to a unitary transform in $\mathbf{X}(l)$. That is, for some $\mathbf{U}$, where $\mathbf{U}^\mathrm{H}\mathbf{U} = \mathbf{U}\mathbf{U}^\mathrm{H} = \mathbf{I}$ [29]

$$\left[\mathbf{U}\hat{\mathbf{H}}(l+1|l), (\mathbf{I} \otimes \mathbf{U})\mathbf{P}(l+1|l)(\mathbf{I} \otimes \mathbf{U})^\mathrm{H}\right]$$
$$= \mathcal{MKF}\left(\mathbf{U}\hat{\mathbf{H}}(l|l-1), (\mathbf{I} \otimes \mathbf{U})\mathbf{P}(l|l-1)(\mathbf{I} \otimes \mathbf{U})^\mathrm{H}, \dots \right.$$
$$\left. \mathbf{X}(l)\mathbf{U}^\mathrm{H}, \mathbf{Y}(l)\right).$$

Lemma 2: When an S-STBC is used, the MNI-DD (7) introduces a counterunitary transform in $\hat{\mathbf{X}}(l)$ in response to a unitary transform in $\hat{\mathbf{H}}(l|l-1)$. For some $\mathbf{U}$, $\mathbf{U}^\mathrm{H}\mathbf{U} = \mathbf{U}\mathbf{U}^\mathrm{H} = \mathbf{I}$ [29]

$$\hat{\mathbf{X}}(l)\mathbf{U}^\mathrm{H} = \arg\min_{\boldsymbol{\Xi} \in \boldsymbol{\Omega_\mathbf{X}}}\left\{\tilde{Z}_P\left(\mathbf{Y}(l), \boldsymbol{\Xi}, \mathbf{U}\hat{\mathbf{H}}(l|l-1), \dots \right.\right.$$
$$\left.\left. (\mathbf{I} \otimes \mathbf{U})\mathbf{P}(l|l-1)(\mathbf{I} \otimes \mathbf{U})^\mathrm{H}\right)\right\}.$$

Proposition 3: Assume that an S-STBC is used and $\mathbf{Y}_1^L = [\mathbf{Y}(1), \dots, \mathbf{Y}(L)]$ is observed. By Lemmas 1 and 2, given the initial condition $\hat{\mathbf{H}}(0|0)$, the CE/DD produces the detected code and channel estimate sequences, i.e.,

$$\left(\hat{\mathbf{X}}_1^L = \left[\hat{\mathbf{X}}(1), \dots, \hat{\mathbf{X}}(L)\right], \dots \right.$$
$$\left. \hat{\mathbf{H}}_1^L = \left[\hat{\mathbf{H}}(1|0), \dots, \hat{\mathbf{H}}(L|L-1)\right]\right).$$

Given $\mathbf{U}\hat{\mathbf{H}}(0|0)$, the CE/DD produces the unitary transformed counterparts $(\hat{\mathbf{X}}_1^L\mathbf{U}^\mathrm{H}, \mathbf{U}\hat{\mathbf{H}}_1^L)$.

Corollary 4: In most realistic applications, the actual initial condition is unavailable. Setting it to match the statistical property of the channel ($\hat{\mathbf{H}}(0|0) = \mathbf{0}$), from Proposition 3, both $\hat{\mathbf{X}}_1^L$ and $\hat{\mathbf{X}}_1^L\mathbf{U}^\mathrm{H}$ are likely equal and cannot be differentiated by the CE/DD.

Since the model in (2) is assumed to be intersymbol-interference free, the normalized innovations in (7) is a white process, and the path metric is the sum of the branch metrics: $\bar{p}(\mathbf{X}_1^L) = \sum_{l=1}^L \tilde{p}(\hat{\mathbf{X}}(l)) = \sum_{l=1}^L \tilde{Z}_P(\mathbf{Y}(l), \hat{\mathbf{X}}(l), \hat{\mathbf{H}}(l\,|\,l-1), \mathbf{P}(l|l-1))$. Two sequences of codewords are *isometric* if their

path metrics are identical. From Corollary 4, $\hat{\mathbf{X}}_1^L$ and $\hat{\mathbf{X}}_1^L\mathbf{U}^\mathrm{H}$ are isometric because

$$\bar{p}\left(\hat{\mathbf{X}}_1^L\right) = \sum_{l=1}^L \tilde{Z}_P\left(\mathbf{Y}(l), \hat{\mathbf{X}}(l), \hat{\mathbf{H}}(l|l-1), \mathbf{P}(l|l-1)\right)$$
$$= \sum_{l=1}^L \tilde{Z}_P\left(\mathbf{Y}(l), \hat{\mathbf{X}}(l)\mathbf{U}^\mathrm{H}, \mathbf{U}\hat{\mathbf{H}}(l|l-1), \right.$$
$$\left. \mathbf{U}\mathbf{P}(l|l-1)\mathbf{U}^\mathrm{H}\right)$$
$$= \bar{p}\left(\hat{\mathbf{X}}_1^L\mathbf{U}^\mathrm{H}\right). \quad (12)$$

Hence, S-STBCs (11) cause isometry and detection ambiguity in CE/DD. Selection of the wrong isometric sequence results in detection error floor ($\mathbf{X}(l) \neq \mathbf{X}(l)\mathbf{U}^\mathrm{H}$) and increases the estimation error ($\|\mathbf{H}(l) - \mathbf{U}\hat{\mathbf{H}}(l|l)\|_\mathrm{F} \gg \|\mathbf{H}(l) - \hat{\mathbf{H}}(l|l)\|_\mathrm{F}$) [21], [29], [39], [47]. A large number of STBCs used in the literature, e.g., all the STBCs shown thus far, are symmetric. Thus, isometry affects a wide variety of applications.

VI. STBC DESIGN

A. A-STBCs

Since isometry is induced by the *symmetry* of an STBC, it is natural to consider A-STBC as a solution. To ensure that $\mathbf{X}_1^L$ and $\mathbf{X}_1^L\mathbf{U}^\mathrm{H}$ are not isometric, it is sufficient that the subsequences $[\mathbf{X}(a), \mathbf{X}(b)]$ and $[\mathbf{X}(a)\mathbf{U}^\mathrm{H}, \mathbf{X}(b)\mathbf{U}^\mathrm{H}]$, for some $1 \leq a < b \leq L$, are not isometric [21], [29], i.e.,

$$\forall\left(\mathbf{X}(a) \in \boldsymbol{\Omega}_{\mathbf{X}(a)}, \mathbf{X}(b) \in \boldsymbol{\Omega}_{\mathbf{X}(b)}\right), \dots$$
$$\nexists(\mathbf{U}^\mathrm{H} \neq \mathbf{I})\left\{\mathbf{X}(a)\mathbf{U}^\mathrm{H} \in \Omega_{\mathbf{X}(a)}, \mathbf{X}(b)\mathbf{U}^\mathrm{H} \in \Omega_{\mathbf{X}(b)}\right\}. \quad (13)$$

Two easy ways to design a pair of P-ary S-STBC $\boldsymbol{\Omega}_{\mathbf{C}(a)}$ and P-ary A-STBC $\boldsymbol{\Omega}_{\mathbf{C}(b)}$, so that transmit matrices $\mathbf{X}(a)$ and $\mathbf{X}(b)$ together break isometry, as indicated in (13), are given in the following:

Algorithm 1: Retain the linear unitary STBC structure but employ asymmetric signal constellation (ASC) [39]: 1) Select a linear unitary STBC structure. 2) For $\boldsymbol{\Omega}_{\mathbf{C}(a)}$, employ regular $\sqrt[Q]{P}$-phase-shift keying (PSK) for the Q symbols. 3) For $\boldsymbol{\Omega}_{\mathbf{C}(b)}$, employ asymmetric $\sqrt[Q]{P}$-PSK. For example, using the Alamouti's code structure in [1] ($Q = 2$), binary phase-shift keying (BPSK) $\{1, -1\}$, and asymmetric BPSK $\{1, j\}$ [39], the following pair of 4-ary STBCs $\boldsymbol{\Omega}_{\mathbf{C}(a)}$ and $\boldsymbol{\Omega}_{\mathbf{C}(b)}$ break isometry, as described in (13). Thus

$$\boldsymbol{\Omega}_{\mathbf{C}(a)} = \left\{\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix}\right\}$$

$$\boldsymbol{\Omega}_{\mathbf{C}(b)} = \left\{\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & j \\ j & 1 \end{bmatrix}, \begin{bmatrix} j & -1 \\ 1 & -j \end{bmatrix}, \begin{bmatrix} j & j \\ j & -j \end{bmatrix}\right\}.$$

Remark—Tradeoff With Algorithm 1: Both $\boldsymbol{\Omega}_{\mathbf{C}(a)}$ and $\boldsymbol{\Omega}_{\mathbf{C}(b)}$ contain the same number of codewords. However, due
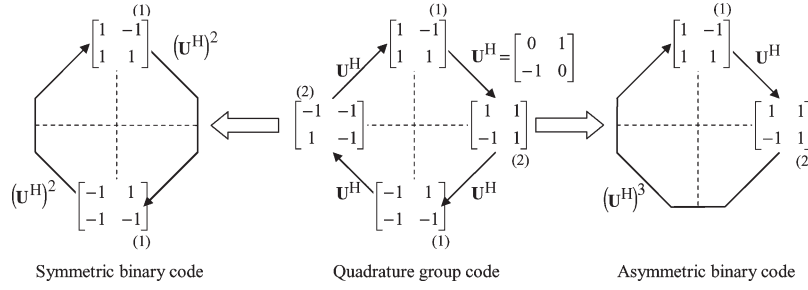
Fig. 2.   Construction of an A-STBC from a group code.

to the employment of ASC, the codewords of the A-STBC are not uniformly distributed in the space-time sphere, and $\mathbf{\Omega}_{\mathbf{C}(b)}$ has a smaller d_s^2. This contributes unfavorably to the detection of $\mathbf{C}(b)$. However, since isometry is broken, the sequence of codewords and channel estimates are uniquely identified. Hence, the performance improvement of the channel estimation and detection of past codewords contributes favorably to the detection of the entire sequence. Since A-STBC is only used *once* in each sequence, as long as d_s^2 is not too small to break isometry, the use of A-STBCs improves overall detection performance.   ∎

Algorithm 2: Group STBC can be seen as *generalized PSK* in the space-time domain [46]. Hence, [39, Alg. 1] can be modified to construct a pair of P-ary S-STBC and P-ary A-STBC from a P^2-ary unitary group STBC: 1) Select a P^2-ary unitary group STBC (either cyclic or dicyclic). 2) Arrange the codewords in a circle where each codeword is a $\mathbf{U}^\mathrm{H}$ rotated version of the previous one. 3) Label the codewords sequentially with $1, \ldots, P$. 4) Select a set of codewords with the same label to obtain the P-ary S-STBC. 5) Select one codeword from each set to obtain the P-ary A-STBC.

For example, Fig. 2 illustrates how a pair of symmetric 2-ary STBC and asymmetric 2-ary STBC can be constructed from a 4-ary cyclic group STBC.

Remark—Tradeoff With Algorithm 2: Since the resulting P-ary STBCs are subsets of the P^2-ary group STBC, the code rate is reduced. This can be compensated by using a group STBC with higher P but smaller d_s^2. Again, the smaller d_s^2 is usually not an issue.   ∎

Asymmetry breaks isometry and provides the ML DD with the correct channel estimate $\hat{\mathbf{H}}(l|l-1)$ instead of the unitarily rotated version $\mathbf{U}\hat{\mathbf{H}}(l|l-1)$. This leads to the mitigation of the detection error floor. Pilot training, which breaks isometry, is equivalent to having only one element in the A-STBC $\mathbf{\Omega}_{\mathbf{X}(b)}$. Hence, training can be seen as a special case of A-STBC. Since all S-STBCs that are shown thus far are square, the examples of A-STBCs that are shown here are also square. However, A-STBCs can also be rectangular, and the analysis in this paper will also apply.

B. Estimation-Based STBCs

1) Code Design Criteria: As discussed in Section IV, in addition to isometry, estimation error also contributes to increasing P_e. To minimize $\tilde{P}_{\alpha \to \beta}$, due to the negative exponent, it is necessary that $\lambda_{\alpha\beta+}/-\lambda_{\alpha\beta-} \gg 1$, or equivalently, $\lambda_{\alpha\beta+} +$

$\lambda_{\alpha\beta-} \gg 0$. Thus, $MN(\lambda_{\alpha\beta+} + \lambda_{\alpha\beta-}) = \mathrm{tr}(\mathrm{cov}(\mathbf{\Upsilon})\mathbf{\Theta}_{\alpha,\beta}) = T_1$ must be maximized, i.e.,

$$
\begin{aligned}
T_1 = \mathrm{tr}\Bigg[& \left(\breve{\mathbf{X}}_\alpha \mathrm{cov}\left(\mathbf{H}(l)\right) \breve{\mathbf{X}}_\alpha^\mathrm{H} + \mathbf{R} \right) \left(\mathbf{P}_{\breve{\mathbf{Z}},\beta}^{-1} - \mathbf{P}_{\breve{\mathbf{Z}},\alpha}^{-1} \right) \\
& + \breve{\mathbf{X}}_\alpha \mathrm{cov}\left(\hat{\mathbf{H}}(l|l-1) \right) \breve{\mathbf{X}}_\alpha^\mathrm{H} \mathbf{P}_{\breve{\mathbf{Z}},\alpha}^{-1} \\
& - \breve{\mathbf{X}}_\alpha \mathrm{cov}\left(\hat{\mathbf{H}}(l|l-1) \right) \breve{\mathbf{X}}_\beta^\mathrm{H} \mathbf{P}_{\breve{\mathbf{Z}},\beta}^{-1} \\
& - \breve{\mathbf{X}}_\beta \mathrm{cov}\left(\hat{\mathbf{H}}(l|l-1) \right) \breve{\mathbf{X}}_\alpha^\mathrm{H} \mathbf{P}_{\breve{\mathbf{Z}},\beta}^{-1} \\
& + \breve{\mathbf{X}}_\beta \mathrm{cov}\left(\hat{\mathbf{H}}(l|l-1) \right) \breve{\mathbf{X}}_\beta^\mathrm{H} \mathbf{P}_{\breve{\mathbf{Z}},\beta}^{-1} \Bigg].
\end{aligned} \tag{14}
$$

Because of its complexity, (14) is simplified by assuming that BFM ($\mathrm{cov}(\mathbf{H}(l)) = \mathbf{I}$, $\mathbf{P}(l|l-1) = \sigma^2\mathbf{I}$, and $\mathrm{cov}(\hat{\mathbf{H}}(l|l-1)) = (1-\sigma^2)\mathbf{I}$, where σ^2 is the theoretical estimation mse) and unitary STBCs are used. Using the matrix inversion lemma, (14) reduces to [29, App. I.1]

$$
\begin{aligned}
\left(\frac{1}{r^2 + r\sigma^2 T} \right) \Bigg\{ & r(1-\sigma^2)\mathrm{tr}\left(\left(\breve{\mathbf{X}}_\alpha - \breve{\mathbf{X}}_\beta \right)^\mathrm{H} \left(\breve{\mathbf{X}}_\alpha - \breve{\mathbf{X}}_\beta \right) \right) \\
& + \sigma^2 \left[MNT^2 - \mathrm{tr}\left(\left(\breve{\mathbf{X}}_\alpha^\mathrm{H} \breve{\mathbf{X}}_\beta \right) \left(\breve{\mathbf{X}}_\alpha^\mathrm{H} \breve{\mathbf{X}}_\beta \right)^\mathrm{H} \right) \right] \Bigg\}.
\end{aligned} \tag{15}
$$

Using the matrix identity $\ln(\det(\mathbf{M})) = \mathrm{tr}(\ln(\mathbf{M}))$ and the Taylor expansion $\ln(1 + x) \approx x$ [5], it is easy to show that (15) resembles the KL distance criterion in [48].

From (15), the first term dominates when $\sigma^2 \to 0$, and the second term dominates when $\sigma^2 \to 1$. When the CSI is known ($\sigma^2 = 0$), (15) becomes

$$
\frac{1}{r}\mathrm{tr}\left[\left(\breve{\mathbf{X}}_\alpha - \breve{\mathbf{X}}_\beta \right)^\mathrm{H} \left(\breve{\mathbf{X}}_\alpha - \breve{\mathbf{X}}_\beta \right) \right] \tag{16}
$$

which is the trace criterion (d_s^2) for designing coherent STBCs [8]. This suggests that square linear unitary STBCs should be used when CSI is known [5], [8], [30]. Let λ_n be the eigenvalues of the matrix $[(\breve{\mathbf{X}}_\alpha - \breve{\mathbf{X}}_\beta)^\mathrm{H}(\breve{\mathbf{X}}_\alpha - \breve{\mathbf{X}}_\beta)]$ and $d_{\min}^2$ be the minimum of these eigenvalues; then, $MNd_{\min}^2 \leq \sum_{n=1}^{MN} \lambda_n$. Hence, maximizing (16) is equivalent to maximizing $d_{\min}^2$.

When the CSI is ignored ($\sigma^2 = 1$), (15) becomes

$$
\frac{1}{r^2 + rT}\left\{ MNT^2 - \mathrm{tr}\left[\left(\breve{\mathbf{X}}_\alpha^\mathrm{H} \breve{\mathbf{X}}_\beta \right) \left(\breve{\mathbf{X}}_\alpha^\mathrm{H} \breve{\mathbf{X}}_\beta \right)^\mathrm{H} \right] \right\}. \tag{17}
$$

In (17), $\breve{\mathbf{X}}_\alpha^{\mathrm{H}}\breve{\mathbf{X}}_\beta$ resembles the differential demodulator that suggests that differential demodulation and group STBC should be used when CSI is ignored [4], [6]. Let λ_n be the eigenvalues of the matrix $\left[(\breve{\mathbf{X}}_\alpha^{\mathrm{H}}\breve{\mathbf{X}}_\beta) \quad (\breve{\mathbf{X}}_\alpha^{\mathrm{H}}\breve{\mathbf{X}}_\beta)^{\mathrm{H}}\right]$ and $\delta_{\max}^2$ be the maximum of these eigenvalues; then, $\sum_{n=1}^{MN} \lambda_n \leq MN\delta_{\max}^2$. So, maximizing (17) is equivalent to minimizing $\delta_{\max}^2$.

From (16) and (17), one might propose to first employ group STBC and differential demodulation when σ^2 is large and then switch to square linear unitary STBC once σ^2 is small enough. However, differential demodulation provides only the unitary rotational differences between consecutive codewords (which can be seen as the space-time extension of the phase differences between consecutive symbols); thus, the sets $(\hat{\mathbf{X}}_1^L, \hat{\mathbf{H}}_1^L)$ and $(\mathbf{X}_1^L \mathbf{U}^{\mathrm{H}}, \mathbf{U}\hat{\mathbf{H}}_1^L)$ remain isometric [29], [39]. So, without isometry-breaking A-STBCs or training, the CSI remain ambiguous, and CE/DD cannot switch from differential demodulation to coherent detection. Thus, a single type of STBC that performs well in various $0 < \sigma^2 < 1$ should be employed.

2) SM-STBCs: First, the competing criteria (16) and (17) suggest that the successful STBC should adapt its code properties to σ^2. Hence, this STBC is called the *SM-STBC*. Second, Section IV suggests that the SM-STBC should also reduce the estimation error. Finally, Section VI-A suggests that the SM-STBC should break isometry. Again, to reduce the complexity that is involved in designing the SM-STBC, BFM is assumed. The SM-STBC is designed as follows: 1) *Embedded training*: Training has been used traditionally to reduce estimation error: pilot matrices [2] and codes with embedded training [49]. To minimize P_e, STBCs with embedded training are considered. Hence, the SM-STBC $(T \times M)$ is divided into a "training" part $(T_t \times M)$ and an "information" part $(T_d \times M)$, where $T = T_t + T_d$. According to [2] and [50], the training matrix should be unitary (i.e., $T_t \geq M$, $T \geq M$). 2) *Information*: Since square linear unitary STBCs perform well in coherent detection, they are adopted for the information part (i.e., $T_d = M$, $T \geq 2M$) [5], [8], [11], [12], [30]. 3) *Asymmetry*: Embedded training with a fixed matrix reduces P_e by reducing σ^2 and breaking isometry, but it also reduces transmission rate [29], [39]. Since training is a limiting case of asymmetry, the training part is replaced by information-bearing A-STBCs in Section VI-A. 4) *Self-matching*: When σ^2 is small, training is less important, and most of the signal energy should be used for information. Training becomes significant when σ^2 increases, and energy should be allocated appropriately between the training and information parts. The distribution should be controlled by a weighting factor, and (15) suggests that the factor should be $\mathbf{P}(l|l-1) = \sigma^2\mathbf{I}$. From (16) and (17), the SM-STBC should adapt its $d_{\min}^2$ and $\delta_{\max}^2$ according to σ^2: Maximization of $d_{\min}^2$ should be emphasized when σ^2 is small, and minimization of $\delta_{\max}^2$ should be emphasized when σ^2 is large.

Remark—Validity of the Code Design Criterion: At the heart of the SM-STBC is the square information part. If the square information part achieves full transmit diversity in BFM, then it also achieves full transmit diversity in CFM [30]. Hence, even though the code design criterion (15) is suboptimal (i.e., derived using the BFM), it is a reasonable approach. ∎

Following the preceding design process, the following $T \times M$ unitary SM-STBC is proposed:

$$\mathbf{C}_{\mathrm{SM}}\left(\{b_p\}_{p=1}^P, \{c_q\}_{q=1}^Q\right)$$
$$= \sqrt{\frac{u+v}{1+\sigma^2}} \cdot \left[\frac{\sigma}{\sqrt{u}}\left(\mathbf{U}\left(\{b_p\}_{p=1}^P\right)\right)^{\mathrm{H}} \middle| \frac{1}{\sqrt{v}}\left(\mathbf{V}\left(\{c_q\}_{q=1}^Q\right)\right)^{\mathrm{H}}\right]^{\mathrm{H}}$$

$$(18)$$

where $\mathbf{U}(\cdot)$ is an $(T-M) \times M$ unitary A-STBC, $\mathbf{V}(\cdot)$ is an $M \times M$ unitary S-STBC, $\{b_p\}_{p=1}^P$ and $\{c_q\}_{q=1}^Q$ are independent symbols, $u = \|\mathbf{U}(\cdot)\|_{\mathrm{F}}^2 = (T-M)M$, and $v = \|\mathbf{V}(\cdot)\|_{\mathrm{F}}^2 = M^2$.

Remark—Differences Between the SM-STBCs in (18) and [13]: The SM-STBC in (18) generalizes the previously proposed solution in [13] by using an A-STBC instead of a fixed unitary matrix for the training part. Since $T \geq 2M$, the data rate of the SM-STBC in [13] is reduced by at least half. Since the SM-STBC in (18) uses information-bearing A-STBC, it breaks isometry and decreases P_e without reducing the code rate substantially. In addition, the modulation and CE/DD schemes in [13] are ineffective: 1) Extra pilot matrices are used for CE, and the embedded training property of the SM-STBC is never exploited. The embedded training property of (18) is explicitly and effectively utilized in this paper. 2) The CE is a function of the pilot matrix, which must precede the information-bearing codewords, and the DD is a function of this CE. Thus, the channel estimates that are used in DD are always outdated. The CE/DD scheme that is proposed in this paper is derived based on the CFM, and the channel estimates that are used for DD are kept up to date recursively using the MKF. 3) The steady-state σ^2 is used in [13]. The MKF in this paper easily calculates σ^2 for every l (*a priori* if desired), which accounts for the fact that the (theoretical) estimation performance changes over time. ∎

Remark—Open-Loop Design of the CE/DD and SM-STBC Scheme: To adapt the SM-STBC in (18), knowledge of the estimation performance at every l is required by not only the receiver but the transmitter as well. In realistic applications, the actual estimation error square $\mathrm{vec}(\tilde{\mathbf{H}}(l|l-1))\mathrm{vec}(\tilde{\mathbf{H}}(l|l-1))^{\mathrm{H}}$, where $\tilde{\mathbf{H}}(l|l-1) = \mathbf{H}(l) - \hat{\mathbf{H}}(l|l-1)$, cannot be calculated since $\mathbf{H}(l)$ is unknown. Generally, this quantity may be approximated at the transmitter using some elaborate procedure only if $\hat{\mathbf{H}}(l|l-1)$ is fed back. However, using the MSSM, the MKF recursively calculates the theoretical estimation error covariance $\mathrm{cov}(\tilde{\mathbf{H}}(l|l-1)) = \mathbf{P}(l|l-l) = \sigma^2\mathbf{I}$ at every l. From (6), this calculation does not require knowledge of $\mathbf{Y}(l)$ or $\hat{\mathbf{H}}(l|l-1)$. Thus, it can be easily calculated at the transmitter with no feedback from the receiver. In fact, $\mathbf{P}(l|l-1)$ can even be calculated *a priori*. Hence, the use of the model eliminates the need for a feedback loop or some elaborate procedure. So, the simpler open-loop model-based CE/DD and SM-STBC scheme in this paper can take advantage of certain performance gains that are generally accessible only to more complex closed-loop systems, such as those in [51] and [52]. ∎

Let $d_v^2 = \min \mathrm{eig}[(\mathbf{V}_\alpha - \mathbf{V}_\beta)^{\mathrm{H}}(\mathbf{V}_\alpha - \mathbf{V}_\beta)]$, and assuming that $\mathbf{U}(\cdot)$ is detected correctly, the $d_{\min}^2$ and $\delta_{\max}^2$ of the

TABLE III
SUMMARY OF THE DEFAULT VALUES FOR THE SIMULATION PARAMETERS

Parameter	Value
center frequency	$f_c = 1.8 \times 10^9$ Hz (high tier IS-136) [54]
symbol duration	$T_s = 4.12 \times 10^{-5}$ s [54]
pulse shaping	square root raised cosine with a 0.35 roll-off [54]
frame length	162 symbols [54]
Rayleigh fading generator	modified Jakes model [55]
hypermodel	AR-2 model [32]–[34]
normalized fading rate	$f_D T_s = 0.00637$ (i.e. slow fading) [27], [32], [56]
vehicular speeds	92.8 km/h
signal constellation	QPSK
# transmit antennas	$M = 2$
# receive antennas	$N = 2$
training matrices	unitary matrices [2], [51]
linear unitary STBC	Alamouti STBC [1]
differential STBC	Cyclic group STBC [4], [6], [9]
asymmetric STBC	A-STBC in the example of *Algorithm 1*
self-matching STBC	SM-STBC in (18) with $T = 4$
# Monte Carlo runs	5000

SM-STBC in (18) are [29, App. I.2] given as follows:

$$d_{\min}^2 = \frac{T}{M}\left(\frac{d_v^2}{1+\sigma^2}\right) \quad \text{and}$$
$$\delta_{\max}^2 = T^2\left(1 - \frac{\sigma^2}{(1+\sigma^2)^2}\frac{d_v^2}{M}\right). \tag{19}$$

Thus, the properties of the SM-STBC $(d_{\min}^2, \delta_{\max}^2)$ are functions of the property of the square unitary STBC $\mathbf{V}(\cdot)$ (d_v^2). From (19), the SM-STBC adjusts $(T/M)(d_v^2) \geq d_{\min}^2 \geq (T/M)(d_v^2/2)$ and $T^2 \leq \delta_{\max}^2 \leq T^2(1 - (1/4)(d_v^2/M))$ according to $0 \leq \sigma^2 \leq 1$. Since $T \geq 2M$, $d_{\min}^2 \geq d_v^2$. Thus, in CE/DD, the SM-STBC that is expected to perform better than the square unitary STBC, i.e., $\mathbf{V}(\cdot)$, does. When $\sigma^2 \to 0$, the SM-STBC emphasizes increasing $d_{\min}^2 \to (T/M)d_v^2$ and compromises on reducing $\delta_{\max}^2$ by increasing $\delta_{\max}^2 \to T^2$; when $\sigma^2 \to 1$, the SM-STBC emphasizes reducing $\delta_{\max}^2 \to T^2(1 - (1/4)(d_v^2/M))$ and compromises on increasing $d_{\min}^2$ by reducing $d_{\min}^2 \to (T/M)(d_v^2/2)$.

Remark—P_e of SM-STBCs: Assuming that BFM and SM-STBCs are used, $-(\lambda_{\alpha\beta+}/\lambda_{\alpha\beta-})$ in (8) also becomes (10), where $T_1 = (TN/M)((\sigma^4 T + r - r\sigma^4)/((r^2 + r\sigma^2 T)(1+\sigma^2)^2))d_v^2$ [43, App. H]. The P_e of the SM-STBC is a function of the distance property of the square unitary code $\mathbf{V}(\cdot)$. Thus, the design of a SM-STBC can be reduced to the design of a square unitary STBC. Since the design of square unitary STBCs has been studied very well, the design of an SM-STBC is made very easy. ∎

VII. RESULT

Table III summarizes the default values of the system and simulation parameters: $f_D T_s = 0.00637$, $M = N = 2$, QPSK is used for the symbols in linear unitary STBCs, unitary matrices are used for pilot and embedded training, the A-STBC in the example of Algorithm 1 is used when an A-STBC is needed, and 5000 independent iterations are performed for each experiments. Table IV summarizes the properties of the "benchmark STBCs": Alamouti with QPSK (coherent and CE/DD), 16-ary cyclic group STBC (differential), and SM-STBC in (18) with

$T = 4$, Alamouti with QPSK for $\mathbf{V}(\cdot)$, and either fixed unitary matrix or an A-STBC (correctly detected) for $\mathbf{U}(\cdot)$ (CE/DD). When formulating SM-STBC using Alamouti, the code rate reduces by 50% since T increases by 100%. The values in the parenthesis in Table IV illustrates how some of the lost code rate can be recovered by replacing fixed unitary matrix with information-bearing A-STBC.

The theoretical P_e at different σ^2's of various receiver and STBC schemes are compared in Fig. 3 by fixing the SNR at 20 dB and assuming that isometry is mitigated. The P_e of the CE/DD with Alamouti STBC is excellent (10^{-7}) when $\sigma^2 \approx 0$, but it degrades rapidly as σ^2 increases (Section V). Ignoring CSI, differential demodulation and cyclic STBC offer much poorer (three orders of magnitude higher) but stable $P_e = 10^{-4}$ across all σ^2's. The SM-STBC (with CE/DD) adapts its code properties to σ^2 and offers better P_e than Alamouti (with CE/DD) or cyclic STBC (with differential demodulation) (Section VI-B-2). Since $d_{\min}^2 \geq d_v^2 \ \forall \sigma^2$, in CE/DD, the P_e of the SM-STBC with Alamouti is always lower than that of Alamouti, even when CSI is known (∼2.5 orders of magnitude lower at $\sigma^2 = 0.2$ and ∼0.8 order of magnitude lower at $\sigma^2 = 0$). When $\sigma^2 \to 0$, the P_e of the SM-STBC in CE/DD approaches that of the coherent limit (known CSI and Alamouti) and eventually surpasses it as $d_{\min}^2 \to 2d_v^2$.

The experimental P_e and estimation mse of the CE/DD for various STBCs at different SNRs are shown in Figs. 4 and 5. These results are compared against the P_e of the coherent receiver when the CSI is known and the mse of the CE when the data are known. Using Alamouti STBC, the CE/DD experiences an error floor $(P_e = 0.5)$ due to isometry (Section V). Ignoring the CSI, differential demodulation with cyclic STBC can only reduce the error floor $(P_e = 7 \times 10^{-3})$ but not eliminate it since the first block of every frame remains ambiguous [29], [39]. The error floor is eliminated when a pilot matrix is used with differential demodulation, but the P_e is still much higher than the coherent limit (∼2 orders of magnitude higher at 15 dB). Thus, a CE/DD scheme with an isometry-breaking solution should be employed.

Using pilot matrices to break isometry, the P_e of the CE/DD with Alamouti can be further reduced (by 0.5 order of

TABLE IV
Various Benchmark STBCs and Their Properties. $\boldsymbol{\Delta}_{\alpha\beta} = \mathbf{C}_\alpha - \mathbf{C}_\beta$ Where $\mathbf{C}_\alpha, \mathbf{C}_\beta \in \boldsymbol{\Omega}_\mathbf{C}, \mathbf{C}_\alpha \neq \mathbf{C}_\beta$. N/A Indicates That These Properties Are not Defined for the STBC. The Properties of the SM-STBC Are Calculated, Assuming That Isometry has Already Been Broken Because $\mathbf{U}(\cdot)$ Is Known or Detected Perfectly. The Values in the Parenthesis Show the Code Size, and the Code Rate of the SM-STBC With A-STBC Before Isometry Is Broken

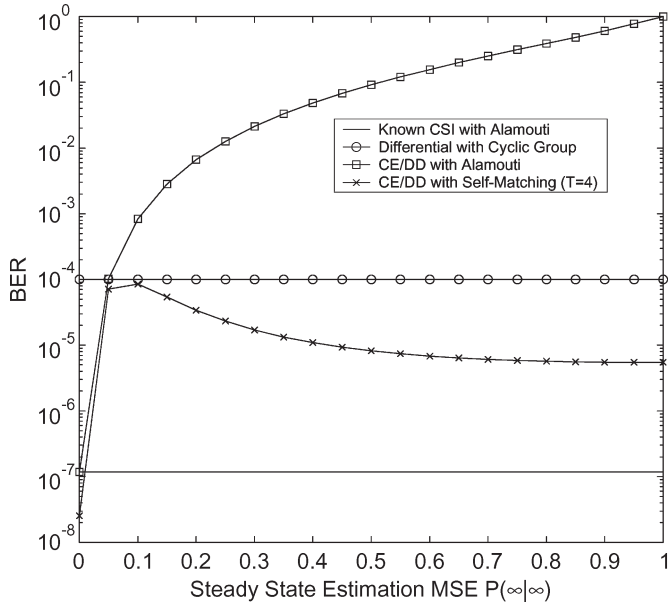|		Alamouti	Self-Matching	Cyclic Group
	$T \times M$	2×2	4×2	2×2
	size of STBC	16	16 (64)	16
	code rate (bits/channel use)	2	1 (1.5)	2
	linearity	Yes	No	No
	unitarity	Yes	Yes	Yes
	symmetry	Yes	No	Yes
	rank = $\mathrm{rank}(\boldsymbol{\Delta}_{\alpha\beta})$ [2], [30]	2	2	2
5 dB	$\det = \det(\boldsymbol{\Delta}_{\alpha\beta}^{\mathrm{H}}\boldsymbol{\Delta}_{\alpha\beta})$ [8], [10]	4	8.4016	1.3726
	$d_s^2 = \mathrm{tr}(\boldsymbol{\Delta}_{\alpha\beta}^{\mathrm{H}}\boldsymbol{\Delta}_{\alpha\beta})$ [8]	4	5.7971	2.3431
	$d_{\min}^2 = \min\mathrm{eig}\left(\boldsymbol{\Delta}_{\alpha\beta}^{\mathrm{H}}\boldsymbol{\Delta}_{\alpha\beta}\right)$ [13]	2	2.8986	0.3045
	$\delta_{\max}^2 = \max\mathrm{eig}\left((\mathbf{C}_\alpha^{\mathrm{H}}\mathbf{C}_\beta)(\mathbf{C}_\alpha^{\mathrm{H}}\mathbf{C}_\beta)^{\mathrm{H}}\right)$ [13]	4	12.8074	4
15 dB	$\det = \det(\boldsymbol{\Delta}_{\alpha\beta}^{\mathrm{H}}\boldsymbol{\Delta}_{\alpha\beta})$	4	14.7929	1.3726
	$d_s^2 = \mathrm{tr}(\boldsymbol{\Delta}_{\alpha\beta}^{\mathrm{H}}\boldsymbol{\Delta}_{\alpha\beta})$	4	7.6923	2.3431
	$d_{\min}^2 = \min\mathrm{eig}\left(\boldsymbol{\Delta}_{\alpha\beta}^{\mathrm{H}}\boldsymbol{\Delta}_{\alpha\beta}\right)$	2	3.8462	0.3045
	$\delta_{\max}^2 = \max\mathrm{eig}\left((\mathbf{C}_\alpha^{\mathrm{H}}\mathbf{C}_\beta)(\mathbf{C}_\alpha^{\mathrm{H}}\mathbf{C}_\beta)^{\mathrm{H}}\right)$	4	15.4083	4

Fig. 3. Comparison of the theoretical P_e at various σ^2's, in BER, among different receiver and STBC schemes, fixing SNR at 20 dB.
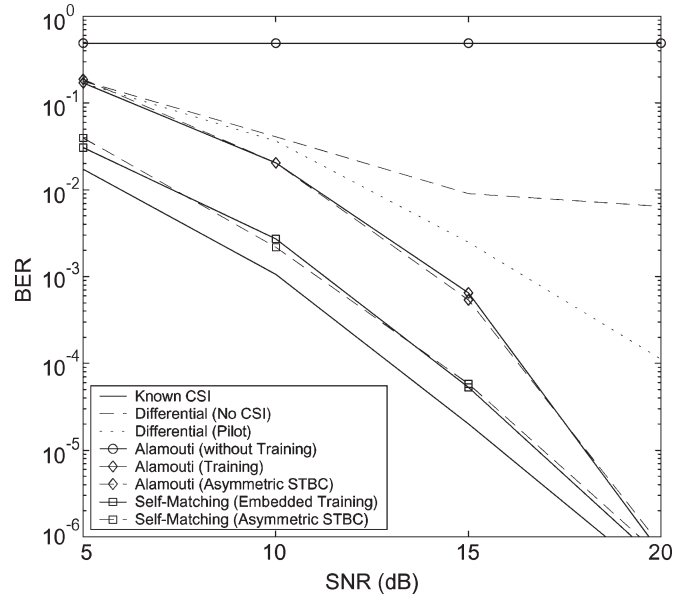

Fig. 4. Experimental BER at various SNR of 1) coherent detector when CSI is known; 2) ordinary differential demodulator; 3) differential demodulator with one pilot matrix; 4) CE/DD, Alamouti with QPSK, and no isometry-breaking solution; 5) CE/DD, Alamouti with QPSK, and pilot training; 6) CE/DD, Alamouti with QPSK, and A-STBC to break isometry; 7) CE/DD, SM-STBC with fixed $\mathbf{U}(\cdot)$; and 8) CE/DD, SM-STBC with A-STBC.

magnitude at 15 dB). When isometry is mitigated, the performance of the CE/DD with Alamouti improves substantially (P_e and mse are reduced by three and two orders of magnitude at 15 dB). When transmission-rate-reducing pilots are replaced by information-bearing A-STBCs to break isometry, the performance does not degrade. This confirms that asymmetry breaks isometry and generalizes training (Section VI-A). Assuming that isometry is broken, as SNR increases, the mse decreases, and the P_e of the CE/DD with Alamouti approaches that of the coherent limit. However, at lower SNR (< 10 dB), the mse is larger, and the P_e of the CE/DD with Alamouti is only marginally lower than that of the differential demodulation.

This confirms that Alamouti performs well only when the mse is negligible (Section V).

To further improve P_e, the SM-STBC in (18) is proposed. Since the design of the SM-STBC is reduced to the design of the linear square unitary STBC $\mathbf{V}(\cdot)$ (Section VI-B2), $\mathbf{V}(\cdot)$ is fixed as Alamouti with QPSK. Experimentally, it is found that a 2×2 $\mathbf{U}(\cdot)$ is enough training for this particular setup, so $T = 4$ [29]. With a fixed $\mathbf{U}(\cdot)$, the SM-STBC performs much better than its square linear unitary STBC counterpart (P_e is one order of magnitude lower at 10 and 15 dB, and mse is 0.5 order of
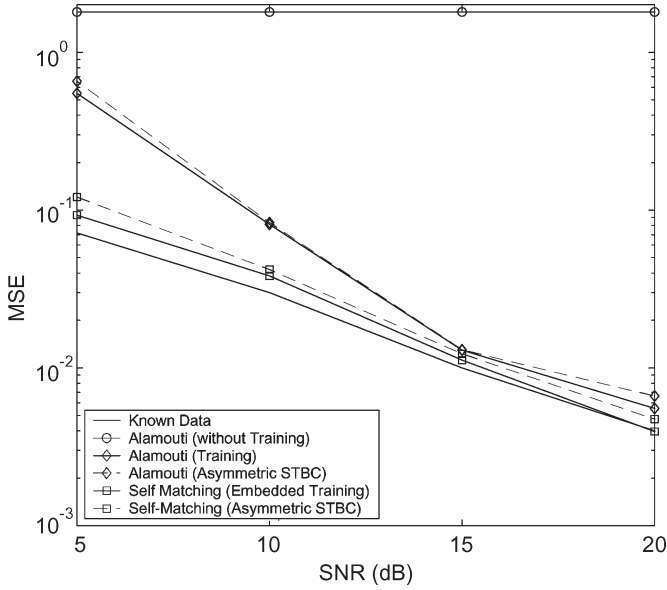
Fig. 5.   Experimental estimation mse at various SNRs of 1) channel estimator when data are known; 2) CE/DD, Alamouti with QPSK, and no isometry-breaking solution; 3) CE/DD, Alamouti with QPSK, and pilot training; 4) CE/DD, Alamouti with QPSK, and A-STBC to break isometry; 5) CE/DD, SM-STBC with fixed $\mathbf{U}(\cdot)$; and 6) CE/DD, SM-STBC with A-STBC.

magnitude lower at 10 dB). The tradeoff for the improvement in P_e is the reduction in code rate (reduced by 50%). However, the use of Alamouti with QPSK in CE/DD requires an isometry-breaking solution. The use of pilot reduces transmission rate, and the use of A-STBC requires parallel MKFs [29], [39]. The use of asymmetric SM-STBC does not induce isometry, so neither pilots nor parallel MKFs are needed. To compensate for the reduction in code rate, the fixed $\mathbf{U}(\cdot)$ is replaced by an A-STBC (6.2.2) (code rate increases from 1 to 1.5). The performance of the SM-STBC with A-STBC is shown to be the same as that of the SM-STBC with fixed $\mathbf{U}(\cdot)$.

Table V compares the theoretical and experimental P_e of various STBCs for different M's and with different constellations. The theoretical P_e (Section IV) agree with the experimental P_e. Comparing the P_e among the square unitary STBCs, Alamouti with QPSK consistently performs better than Alamouti with 16QAM since the former has a much larger d_s^2 (Section V). In addition, increasing $M = 2$ to $M = 4$ improves P_e, due to the larger d_s^2. Comparing the P_e among the rectangular unitary STBCs, due to the larger d_v^2, the 8×4 SM-STBC outperforms the 4×2 SM-STBC with QPSK, and the latter outperforms the 4×2 SM-STBC with 16QAM (Section VI-B2). Comparing the P_e between each square unitary STBC and its rectangular SM-STBC counterpart, the latter outperforms the former due to its code-adapting, isometry-breaking, and embedded training properties (Section VI-B2).

VIII. CONCLUSION

The major contribution of this paper is providing a unified framework to model a MIMO system and, subsequently, to derive a model-based CE/DD scheme and to design STBCs. Using the more realistic CFM, a complete MSSM (its dynamics and statistics) for a MIMO system is derived without imposing restrictive assumptions about the system. The MSSM employs matrix quantities consistently, which intuitively represent the spatial and temporal dimensions of a MIMO system. This natural and compact representation provides insights that are lost in other models [2], [13], [15]–[17], [28], [35]–[38]. From the MSSM, a consistent matrix CE/DD scheme is derived. Using the technique in [17], the MKF is introduced to the area of communications and formulated specifically to estimate the channel. The ML DD is the MNI-DD where the metric is the Mahalanobis distance square of the innovations weighted by the effective SNR. Motivated by the principle of minimizing P_e, the A-STBC is designed to mitigate isometry, which causes detection error floor in fading. The SM-STBC is proposed to further minimize P_e by adapting its code properties to estimation performance, breaking isometry with asymmetry, and reducing estimation error with its embedded training ability. Generally, the factor that is used to scale the SM-STBC, which is required at the transmitter, can only be calculated in a closed-loop system. However, using the MSSM, the factor can be calculated *a priori* for every block without any feedback. In addition to being effective and simple, the Kalman recursion is a good choice because the normalized innovations, which are used for DD, and the estimation error covariance, which is used to scale the SM-STBC, is readily provided by the MKF. However, the strength of the CE/DD and STBC design framework lie in the modeling. So, other filters can also be derived from the MSSM for CE. This modeling-based transceiver and modulation design philosophy can be applied not only to SISO and MIMO systems but also to many other systems and channels, such as cooperative network and Ricean fading channel.

APPENDIX
DERIVATION OF THE PAIRWISE PROBABILITY OF ERROR

From (7), given $\hat{\mathbf{H}}(l|l-1)$ and $\mathbf{P}(l|l-1)$, following the technique in [2], the PEP is

$$P_{\alpha\to\beta} = P\Bigg\{ \tilde{Z}_P\left(\mathbf{Y}(l), \mathbf{X}_\beta, \hat{\mathbf{H}}(l|l-1), \mathbf{P}(l|l-1)\right) \cdots$$

$$< \tilde{Z}_P\left(\mathbf{Y}(l), \mathbf{X}_\alpha, \hat{\mathbf{H}}(l|l-1), \mathbf{P}(l|l-1)\right) \Bigg\}$$

$$= P\left\{ \mathbf{\Upsilon}^{\mathrm{H}} \mathbf{\Theta}_{\alpha,\beta} \mathbf{\Upsilon} < 0 \right\} \tag{20}$$

where

$$\mathbf{\Upsilon} = \begin{bmatrix} \mathrm{vec}\left(\mathbf{Y}(l)\right) \\ \mathrm{vec}\left(\hat{\mathbf{H}}(l|l-1)\right) \end{bmatrix}$$

$$\mathbf{\Theta}_{\alpha,\beta} = \begin{bmatrix} \mathbf{P}_{\tilde{\mathbf{Z}},\beta}^{-1} - \mathbf{P}_{\tilde{\mathbf{Z}},\alpha}^{-1} & \mathbf{P}_{\tilde{\mathbf{Z}},\alpha}^{-1}\breve{\mathbf{X}}_\alpha - \mathbf{P}_{\tilde{\mathbf{Z}},\beta}^{-1}\breve{\mathbf{X}}_\beta \\ \breve{\mathbf{X}}_\alpha^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{Z}},\alpha}^{-1} - \breve{\mathbf{X}}_\beta^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{Z}},\beta}^{-1} & \breve{\mathbf{X}}_\beta^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{Z}},\beta}^{-1}\breve{\mathbf{X}}_\beta - \breve{\mathbf{X}}_\alpha^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{Z}},\alpha}^{-1}\breve{\mathbf{X}}_\alpha \end{bmatrix}.$$

TABLE V
COMPARISON BETWEEN THE THEORETICAL AND EXPERIMENTAL P_e's OF VARIOUS STBCs. THE 4×4 LINEAR UNITARY STBC, WHICH WAS FOUND IN [5], CONTAINS ONLY THREE COMPLEX SYMBOLS SINCE LINEARITY AND UNITARITY RESTRICT THE MAXIMUM ACHIEVABLE CODE RATE OF A SQUARE STBC. AN EXPERIMENTAL BER OF 0 INDICATES THAT NO ERROR RESULTED IN THE 5000 INDEPENDENT ITERATIONS. THE CODE RATES OF THE SM-STBC WITH FIXED $\mathbf{U}(\cdot)$ ARE SHOWN FIRST, FOLLOWED, IN PARENTHESIS, BY THOSE OF THE SM-STBC WITH A-STBC

	$T \times M$	2×2	2×2	2×2	4×4	4×2	4×2	8×4
	code	Alamouti	Alamouti	Alamouti	Unitary	Self-Matching	Self-Matching	Self-Matching
	sub-code $\mathbf{V}(\cdot)$					Alamouti	Alamouti	4×4 Unitary
	constellation	BPSK	QPSK	16QAM	QPSK	QPSK	16QAM	QPSK
	bits/channel use	1	2	4	1.5	1 (1.5)	2 (2.5)	0.75 (1)
	$d^2_{\min}$ or d^2_v	4	2	0.4	2.6667	2	0.4	2.6667
10 dB	experimental BER	4×10^{-4}	2×10^{-2}	2×10^{-1}	6×10^{-3}	2×10^{-3}	2×10^{-1}	2×10^{-3}
	theoretical BER	4×10^{-4}	2×10^{-2}	1×10^{-1}	5×10^{-3}	1×10^{-3}	1×10^{-1}	1×10^{-3}
15 dB	experimental BER	1×10^{-5}	6×10^{-4}	1×10^{-1}	3×10^{-5}	4×10^{-5}	3×10^{-2}	3×10^{-6}
	theoretical BER	8×10^{-6}	8×10^{-4}	7×10^{-2}	2×10^{-5}	2×10^{-5}	2×10^{-2}	3×10^{-6}
20 dB	experimental BER	0	7×10^{-7}	1×10^{-2}	0	4×10^{-7}	3×10^{-4}	0
	theoretical BER	2×10^{-7}	8×10^{-7}	1×10^{-2}	2×10^{-8}	4×10^{-7}	4×10^{-4}	4×10^{-9}

From [2, eq. (30)], the characteristic function of the random variable in (21), i.e., $\mathbf{\Upsilon}^{\mathrm{H}} \mathbf{\Theta}_{\alpha,\beta} \mathbf{\Upsilon}$, is

$$\phi(s) = \frac{1}{\det\left(\mathbf{I} + s, \mathrm{cov}(\mathbf{\Upsilon})\mathbf{\Theta}_{,\beta}\right)}. \tag{21}$$

From [38, App. A], (21) simplifies to

$$\phi(s) = \left[\frac{1}{(1 + \lambda_{\alpha\beta+s})(1 + \lambda_{\alpha\beta-s})}\right]^{MN} \tag{22}$$

where $\lambda_{\alpha\beta+}$ and $\lambda_{\alpha\beta-}$ are the positive and negative eigenvalues of $\mathrm{cov}(\mathbf{\Upsilon})\mathbf{\Theta}_{\alpha,\beta}$, respectively [42]. From (22) and [38, eqs. (7) and (8)], $P_{\alpha\to\beta}$ is then given by (8).

ACKNOWLEDGMENT

REFERENCES

[1] S. M. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[2] J.-C. Guey, M. P. Fitz, M. R. Bell, and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 4, p. 527, Apr. 1999.

[3] B. Hassibi and B. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1804–1824, Jul. 2002.

[4] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2041–2052, Dec. 2000.

[5] A. Hottinen, O. Tirkkonen, and R. Wichman, *Multi-antenna Transceiver Techniques for 3G and Beyond*. Hoboken, NJ: Wiley, 2003.

[6] B. L. Hughes, "Differential space-time modulation," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2567–2578, Nov. 2000.

[7] B. L. Hughes, "Optimal space-time constellations from groups," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 401–410, Feb. 2003.

[8] D. M. Ionescu, "On space-time code design," *IEEE Trans. Wireless Commun.*, vol. 2, no. 1, pp. 20–28, Jan. 2003.

[9] A. Shokrollahi, B. Hassibi, B. M. Hochwald, and W. Sweldens, "Representation theory for high-rate multiple-antenna code design," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2335–2367, Sep. 2001.

[10] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time codes for high data-rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[11] O. Tirkkonen and A. Hottinen, "Improved MIMO performance with non-orthogonal space-time block codes," in *Proc. IEEE GlobeCom*, Nov. 2001, vol. 2, pp. 1122–1126.

[12] O. Tirkkonen and A. Hottinen, "Square-matrix embeddable space-time block codes for complex signal constellations," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 384–395, Feb. 2002.

[13] E. Baccarelli and M. Biagi, "Performance and optimized design of space-time codes for MIMO wireless systems with imperfect channel estimates," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2911–2923, Oct. 2004.

[14] B. Hochwald and T. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.

[15] C. Komninakis, C. Fragouli, A. H. Sayed, and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1065–1076, May 2002.

[16] Z. Liu, X. Ma, and G. B. Giannakis, "Space-time coding and Kalman filtering for time-selective fading channels," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 183–186, Feb. 2002.

[17] D. Choukroun, H. Weiss, I. Y. Bar-Itzhack, and Y. Oshman, "Kalman filtering for matrix estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 147–150, Jan. 2006.

[18] D. H. Nissen, "A note on the variance of a matrix," *Econometrica*, vol. 36, no. 3/4, pp. 603–604, Jul.–Oct. 1968.

[19] G. D. Forney, Jr., "Geometrically uniform codes," *IEEE Trans. Inf. Theory*, vol. 37, no. 5, pp. 1241–1260, Sep. 1991.

[20] S. Gazor and H. Rad, "Space-time coding ambiguities in joint adaptive channel estimation and detection," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 372–384, Feb. 2004.

[21] S. Lam, K. Plataniotis, and S. Pasupathy, "Isometry during MIMO joint channel estimation and data detection in fading channels," in *Proc. 22nd Biennial Symp. Commun.*, Jun. 2004, pp. 251–253.

[22] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. CAS-25, no. 9, pp. 772–781, Sep. 1978.

[23] E. A. Lee and D. G. Messerschmitt, *Digital Communication*, 2nd ed. Norwell, MA: Kluwer, 1994.

[24] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

[25] T. Yoo and A. Goldsmith, "Capacity of fading MIMO channels with channel estimation error," in *Proc. IEEE ICC*, Jun. 2004, pp. 808–813.

[26] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[27] J. K. Cavers, "On the validity of slow and moderate fading models for matched filter detection of Rayleigh fading signals," *Can. J. Electr. Comput. Eng.*, vol. 17, no. 4, pp. 183–189, 1992.

[28] A. F. Naguib, V. Tarohkh, N. Seshadri, and A. R. Calderbank, "A space-time coding modem for high-data-rate wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1459–1478, Oct. 1998.

[29] S. Lam. (2005, Jun.). "Asymmetric and self-matching modulations for estimation-assisted detection in fading channels," Ph.D. dissertation,

Univ. Toronto, Toronto, ON, Canada. [Online]. Available: http://www.lib.umi.com/cr/utoronto/fullcit?pNR02771

[30] W. Su, Z. Safar, and K. J. R. Liu, "Diversity analysis of space-time modulation over time-correlated Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1832–1839, Aug. 2004.

[31] G. Auer, H. Bogucka, J. Bonnet *et al.*, (2004, Jul.). "Identification of radio-link technologies," *WINNER Information Society Technologies*. [Online]. Available: https://www.ist-winner.org/DeliverableDocuments/D2.1.pdf

[32] L. Lindbom, A. Ahlén, M. Sternad, and M. Falkenström, "Tracking of time-varying mobile radio channels—Part II: A case study," *IEEE Trans. Commun.*, vol. 50, no. 1, pp. 156–167, Jan. 2002.

[33] P. Sharma, "Predictive models for wireless fading channels," Ph.D. dissertation, Univ. Massachusetts, Lowell, 2003.

[34] P. H.-Y. Wu and A. Duel-Hallen, "Multiuser detectors with disjoint Kalman channel estimators for synchronous CDMA mobile radio channels," *IEEE Trans. Commun.*, vol. 48, no. 5, pp. 752–756, May 2000.

[35] J. Giese and M. Skoglund, "Space-time constellation design for partial CSI at the receiver," in *Proc. IEEE Int. Symp. Inf. Theory*, Sep. 2005, pp. 2213–2217.

[36] H. Zhu, B. Farhang-Boroujeny, and C. Schlegel, "Pilot embedding for joint channel estimation and data detection in MIMO communication systems," *IEEE Commun. Lett.*, vol. 7, no. 1, pp. 30–32, Jan. 2003.

[37] R. Knopp and P. A. Humblet, "On coding for block fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 189–205, Jan. 2000.

[38] X. Zhu and R. D. Murch, "Performance analysis of maximum likelihood detection in a MIMO antenna system," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 187–191, Feb. 2002.

[39] S. Lam, K. Plataniotis, and S. Pasupathy, "Isometric data sequences and data modulation schemes in fading," *IEEE Trans. Commun.*, vol. 52, no. 3, pp. 406–415, Mar. 2004.

[40] L. Krasny, H. Arslan, D. Koilpillai, and S. Chennakeshu, "Doppler spread estimation in mobile radio systems," *IEEE Commun. Lett.*, vol. 5, no. 5, pp. 197–199, May 2001.

[41] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Hoboken, NJ: Wiley, 2004.

[42] S. J. Grant and J. K. Cavers, "Performance enhancement through joint detection of cochannel signals using diversity arrays," *IEEE Trans. Commun.*, vol. 46, no. 8, pp. 1038–1049, Aug. 1998.

[43] S. Lam, K. Plataniotis, S. Pasupathy. (2005, Sep.). *Design of estimation-assisted detector and space-time block codes for mimo systems*, Dept. Elect. Comput. Eng., Univ. Toronto, Toronto, ON, Canada, Tech. Rep. version 1. [Online]. Available: http://www.comm.utoronto.ca/~kostas/slam/files/mimo-report.pdf

[44] F. A. Graybill, *Introduction to Matrices With Applications in Statistics*. Belmont, CA: Wadsworth, 1967.

[45] D. M. Ionescu, "New results on space-time code design criteria," in *Proc. IEEE WCNC*, 1999, vol. 2, pp. 684–687.

[46] H. Schulze, "Geometrical properties of orthogonal space-time codes," *IEEE Commun. Lett.*, vol. 7, no. 1, pp. 64–66, Jan. 2003.

[47] P. Kyritsi, R. A. Valenzuela, and D. C. Cox, "Channel and capacity estimation errors," *IEEE Commun. Lett.*, vol. 6, no. 12, pp. 517–519, Dec. 2002.

[48] M. J. Borran, A. Sabharwal, and B. Aazhang, "On design criteria and construction of noncoherent space-time constellations," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2332–2351, Oct. 2003.

[49] M. Brehler and M. K. Varanasi, "Training-codes for the noncoherent multi-antenna block-Rayleigh-fading channel," in *Proc. Conf. Inf. Sci. Syst.*, Mar. 2003, pp. 12–14.

[50] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[51] G. Jöngren, M. Skoglund, and B. Ottersten, "Design of channel-estimate-dependent space-time block codes," *IEEE Trans. Commun.*, vol. 52, no. 7, pp. 1191–1203, Jul. 2004.

[52] T. Yoo, E. Yoon, and A. Goldsmith, "MIMO capacity with channel uncertainty: Does feedback help?" in *Proc. IEEE GlobeCom*, Nov. 2004, pp. 96–100.

[53] W.-R. Wu and Y.-M. Tsuie, "An LMS-based decision feedback equalizer for IS-136 receivers," *IEEE Trans. Veh. Technol.*, vol. 51, no. 1, pp. 130–143, Jan. 2002.

[54] P. Dent, G. E. Bottomley, and T. Croft, "Jakes fading model revisited," *Electron. Lett.*, vol. 29, no. 13, pp. 1162–1163, Jun. 1993.

[55] H. Zamiri-Jafarian and S. Pasupathy, "Adaptive MLSDE using EM algorithm," *IEEE Trans. Commun.*, vol. 47, no. 8, pp. 1181–1193, Aug. 1999.

Stephen Lam was born in Hong Kong on June 16, 1973. He received the B.A.Sc. degree in engineering science (computer) from Simon Fraser University, Burnaby, BC, Canada, in 1997 and the M.A.Sc. and Ph.D. degrees in electrical and computer engineering (communications) from the University of Toronto, Toronto, ON, Canada, in 1999 and 2005, respectively. He is currently a Postdoctoral Research Fellow with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, under the supervision of Prof. Plataniotis and Prof. Pasupathy. His current research interests include the area of receiver designs, specifically, joint channel-estimation-and-data-detection schemes, for digital wireless and mobile communication channels.

K. N. (Kostas) Plataniotis (S'90–M'92–SM'03) received the B.Eng. degree in computer engineering from the University of Patras, Patras, Greece, in 1988 and the M.S. and Ph.D. degrees in electrical engineering from Florida Institute of Technology (Florida Tech), Melbourne, in 1992 and 1994, respectively.

He is currently an Associate Professor with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada. His research interests include multimedia systems, signal and image processing, signal processing, communications systems, and biometrics.

Dr. Plataniotis is a Registered Professional Engineer in the Province of Ontario and a member of the Technical Chamber of Greece. He is an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS and the IEEE SIGNAL PROCESSING LETTERS and the Image Processing Area Editor for the IEEE Signal Processing Society's e-letter. He was the Technical Program Co-Chair for IEEE ICME-2006 and the Vice-Chair of IEEE ITSC-2006. He was the recipient of the 2005 IEEE Canada's Outstanding Engineering Educator Award "for contributions to engineering education and inspirational guidance of graduate students" and a corecipient of the 2006 IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award.
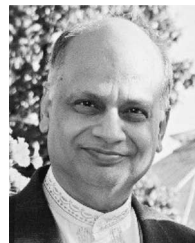
Subbarayan Pasupathy (M'73–SM'81–F'91–LF'07) was born in Chennai (Madras), Tamilnadu, India. He received the B.E. degree in telecommunications from the University of Madras, Chennai, India, the M.Tech. degree in electrical engineering from Indian Institute of Technology, Chennai, and the M.Phil. and Ph.D. degrees in engineering and applied science from Yale University, New Haven, CT.

He is currently a Professor with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, where he has been a Faculty Member since 1972. He has also served as the Chairman of the Communications Group and as the Associate Chairman of the Department of Electrical Engineering, University of Toronto. His research over the last 30 years has mainly been in statistical communication theory and signal processing and their applications to digital communications.

Prof. Pasupathy is a registered Professional Engineer in the Province of Ontario. During 1982–1989, he was an Editor for *Data Communications and Modulation* for the IEEE TRANSACTIONS ON COMMUNICATIONS. He has also served as a Technical Associate Editor for the *IEEE Communications Magazine* (1979–1982) and as an Associate Editor for the *Canadian Electrical Engineering Journal* (1980–1983). He wrote a regular humor column entitled "Light Traffic" for the *IEEE Communications Magazine* during 1984–1998. He was the recipient of the Canadian Award in Telecommunications in 2003 from the Canadian Society of Information Theory. He was elected Fellow of the IEEE in 1991 "for contributions to bandwidth-efficient coding and modulation schemes in digital communication" and was elected as a Fellow of the Engineering Society of Canada in 2004. He has been identified as a "highly cited researcher" by ISI Web of Knowledge, and his name is listed in ISIHighlyCited.com.