Noncoherent Space–Time Coding: An Algebraic Perspective

Hesham El Gamal, *Senior Member, IEEE*, Defne Aktas, *Member, IEEE*, and Mohamed Oussama Damen, *Senior Member, IEEE*

Abstract—The design of space–time signals for noncoherent block-fading channels where the channel state information is not known *a priori* at the transmitter and the receiver is considered. In particular, a new algebraic formulation for the diversity advantage design criterion is developed. The new criterion encompasses, as a special case, the well-known diversity advantage for unitary space–time signals and, more importantly, applies to arbitrary signaling schemes and arbitrary channel distributions. This criterion is used to establish the optimal diversity-versus-rate tradeoff for *training based* schemes in block-fading channels. Our results are then specialized to the class of *affine* space–time signals which allows for a low complexity decoder. Within this class, space–time constellations based on the threaded algebraic space–time (TAST) architecture are considered. These constellations achieve the optimal diversity-versus-rate tradeoff over noncoherent block-fading channels and outperform previously proposed codes in the considered scenarios as demonstrated by the numerical results. Using the analytical and numerical results developed in this paper, nonunitary space–time codes are argued to offer certain advantages in block-fading channels where the *appropriate use* of coherent space–time codes is shown to offer a very efficient solution to the noncoherent space–time communication paradigm.

Index Terms—Diversity-versus-rate tradeoff, generalized likelihood ratio test (GLRT) receivers, noncoherent channels, space–time coding.

I. INTRODUCTION

PIONEERING information-theoretic studies that quantify the significant capacity gains allowed by spatial diversity have sparked significant interest in the design and analysis of space–time codes (STCs) for multiple-input multiple-output (MIMO) fading channels. Earlier works on MIMO systems have primarily focused on the coherent scenario where the channel state information (CSI) is assumed to be known *a priori* at the receiver (e.g., [1]–[3]). While this assumption may be realistic in slow fading channels, where the channel can be estimated accurately with only a marginal loss in throughput,

it is clearly not appropriate for faster fading channels. Fortunately, more recent information-theoretic calculations suggest that high capacities are still possible with multiple-antenna systems in the noncoherent scenario if the channel coherence time is not very small (e.g., [4], [5]). These works have also established the *asymptotic* (at large signal-to-noise ratio (SNR) regime) optimality of unitary space–time signaling schemes in the *ergodic*, i.e., delay-unlimited, scenario. Most of the followup works have focused on the design of **unitary** noncoherent space–time signals for ergodic as well as delay-limited fading channels (e.g., [6], [7]). In fact, the unitary signaling paradigm seems to have been accepted as a *de facto* standard for noncoherent channels. Here, we revisit this conclusion.

In this paper, we reformulate the notion of diversity advantage in space–time channels based on an algebraic perspective. As shown in the sequel, the new formulation encompasses, as a special case, the well-known diversity advantage design criterion for unitary codes, and, more importantly, covers more general signaling schemes and arbitrary channel distributions. Our approach is motivated by the intractability of the probability of error analysis in this general scenario. The proposed formulation is general for block-fading channels with arbitrary delay constraints, and hence, bridges the gap between the quasi-static and ergodic models. We then use the new diversity criterion to characterize the optimal diversity-versus-rate tradeoff for training-based signaling schemes. Training-based schemes have also been considered in recent independent works [8]–[10] (even the codes proposed in [7] can be considered as unitary training-based schemes). The excellent performance of training-based affine space–time constellations was first reported by Brehler and Varanasi in [8]. This class of affine space–time signals is important for practical systems since they can be decoded with a low-complexity algorithm. Independently, affine constellations that achieve full algebraic diversity were proposed later in [9]. Here, we generalize the results in [9] and establish the diversity-versus-rate tradeoff for this class of affine space–time constellations. Our characterization shows that in general the optimal number of active transmit antennas depends on the delay constraint (i.e., the number of blocks per codeword), operating SNR, and the receiver complexity. More specifically, we argue that the optimal number of active antennas in ergodic channels [5] is not necessarily optimal in delay-limited channels. We present exemplary affine constellations based on the threaded algebraic space–time (TAST) constellations [9] that achieve the optimal diversity-versus-rate tradeoff. We use numerical results to argue in favor of those *nonunitary* space–time constellations, and hence, demonstrate

H. El Gamal is with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: helgamal@ece.osu.edu).

D. Aktas is with the Department of Electrical and Electronics Engineering, Bilkent University, Bilkent, Ankara 06800, Turkey (e-mail: daktas@ee.bilkent.edu.tr).

M. O. Damen is with the Electrical and Computer Engineering Department, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: modamen@ece.uwaterloo.ca).

that the appropriate use of coherent STCs allows for a very efficient solution to the noncoherent space–time communication paradigm, an observation which is made independently in a parallel work [10].

The rest of the paper is organized as follows. Section II introduces the assumptions and notations used throughout the paper. The algebraic formulation for the diversity advantage is developed in Section III. Section IV-A utilizes the new criterion to characterize the diversity-versus-rate tradeoff for training-based schemes. The tradeoff characterization for affine signaling schemes and the new TAST constellations are presented in Section IV-B. Section V presents a simulation study that demonstrates the significant gains offered by the proposed constellations. Section VI offers some concluding remarks.

II. SYSTEM MODEL

We consider an $L_t \times L_r$ MIMO frequency-nonselective fading channel. In order to capture the delay constraints, we adopt the simplified block-fading model in which the channel coefficients are assumed to be fixed across a fading block of T_c symbol periods and change independently from one block to the next. The information is coded across L_t transmit antennas and T_s symbol periods. The number of fading blocks per codeword is, therefore, given by $M = T_s/T_c$. For simplicity, we assume here that M is an integer. We will use the notation $(\cdot)[m]$ to denote the vector or matrix of values of parameter $(\cdot)$ for the mth fading block. Mathematically, the $L_r \times T_s$ matrix of discrete time received signals is given by

$$\boldsymbol{Y} = \sqrt{\frac{\rho}{L_t}} \boldsymbol{H} \boldsymbol{S}_i + \boldsymbol{N} \tag{1}$$

with

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}[1] & \boldsymbol{Y}[2] & \cdots & \boldsymbol{Y}[M] \end{bmatrix} \tag{2}$$

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}[1] & \boldsymbol{H}[2] & \cdots & \boldsymbol{H}[M] \end{bmatrix} \tag{3}$$

$$\boldsymbol{N} = \begin{bmatrix} \boldsymbol{N}[1] & \boldsymbol{N}[2] & \cdots & \boldsymbol{N}[M] \end{bmatrix} \tag{4}$$

and

$$\boldsymbol{S}_i = \begin{bmatrix} \boldsymbol{S}_i[1] & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_i[2] & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{S}_i[M] \end{bmatrix}. \tag{5}$$

In this notation, $\boldsymbol{H}[m]$ is the $L_r \times L_t$ channel matrix in the mth block, and hence, the (i,j) element of $\boldsymbol{H}[m]$ denotes the complex channel gain between the ith receive and jth transmit antennas. Similarly, $\boldsymbol{N}[m]$ is the $L_r \times T_c$ additive white Gaussian noise matrix in which the (i,k) entry is the noise sample seen by the ith receive antenna at the kth time interval in the mth block. The block-diagonal $ML_t \times T_s$ transmission matrix $\boldsymbol{S}_i$ is drawn from the space–time constellation $\{\boldsymbol{S}_i\}_{i=1}^{2^{\eta T_s}}$, where η is the throughput in bits per channel use. The channel gains and noise samples are modeled as independent zero-mean complex Gaussian random variables with unit variance. The SNR at each receive antenna is, therefore, given by ρ. We hasten to stress that our development of the algebraic diversity advantage is independent of the fading channel distribution, as shown in the sequel.

In this paper, we focus our attention on the noncoherent scenario where the channel coefficients are known *a priori* at neither the transmitter nor the receiver. We further adopt the slightly pessimistic assumption that the channel distribution is not known at the receiver, and hence, resort to the following decoding rule inspired by the generalized likelihood ratio test (GLRT) receiver

$$\hat{\boldsymbol{S}} = \arg \min_{\{\boldsymbol{S}_i\}, \boldsymbol{H}} \left\| \boldsymbol{Y} - \sqrt{\frac{\rho}{L_t}} \boldsymbol{H} \boldsymbol{S}_i \right\| \tag{6}$$

where $\| \cdot \|$ denotes the Frobenius norm.

III. DIVERSITY ADVANTAGE: AN ALGEBRAIC PERSPECTIVE

One of the fundamental parameters that governs the asymptotic performance of STCs in the high-SNR regime is the diversity advantage. Traditionally, the diversity advantage is defined as the asymptotic rate of decay of the worst case pairwise probability of error as a function of SNR (e.g., [3], [11]). In [6], [12], the design criterion for maximizing the diversity advantage of noncoherent unitary constellations is derived. The choice of unitary constellations is based on an asymptotic information-theoretic optimality argument [4]. This optimality argument, however, is limited to the ergodic scenario and does not necessarily extend to the delay-limited scenario under consideration here. In addition, one can intuitively argue that the unitary constraint entails a loss of half the degrees of freedom (e.g., [13]). Here, we attempt to recover this loss by relaxing the unitary constraint on the matrix codebook $\{\boldsymbol{\bar{S}}_i\}$. As a consequence of this generalization, however, the probability of error analysis becomes intractable [14]. This intractability motivates the algebraic approach adopted in this section to characterize the diversity advantage in noncoherent channels.[1] This algebraic framework is inspired by an *identifiability* argument related to the performance of the GLRT receiver in noiseless environments. As a consistency check, we will recover the well-known diversity advantage of unitary signals [6], [12] as a special case of our algebraic framework. Based on this algebraic approach, we propose a novel *generalized* diversity advantage design criterion for arbitrary space–time signaling schemes. In a parallel work [15], the Kullback–Leibler (KL) distance between the conditional distributions was used to establish a design criterion for nonunitary signaling schemes. In the sequel, we will present simulation results that quantify the performance gains offered by the proposed designs as compared to those in [15]. More generally, we will show that the proposed designs offer significant performance gains and/or complexity reductions compared to all previously known codes in many relevant cases. Motivated by the fact that the overall diversity achieved by a space–time coding scheme is equal to the transmit diversity advantage scaled with the number of receive antennas (i.e., the total diversity advantage is equal to dL_r with d denoting the transmit diversity), we limit our discussion in this section to systems with $L_r = 1$ receive

---

[1]Note that the term (non)coherent channel suffers from a slight abuse of notation since the (non)coherence is more of a receiver assumption. We have resorted to this notation, however, to be consistent with the information-theoretic works [4], [5].

antenna. The channel matrix $\boldsymbol{H}$ will, therefore, reduce to the $1 \times ML_t$ vector $\boldsymbol{h}$.

The first step in our approach is to provide an algebraic interpretation of the diversity advantage for coherent block-fading channels [3], [11]. We will then generalize this interpretation to the noncoherent scenario. In coherent block-fading channels, the transmit diversity advantage of an STC, $\{\boldsymbol{S}_i\}$, is given by [3], [11], [16]

$$d = \min_{i,k:k\neq i} \operatorname{rank}(\boldsymbol{S}_i - \boldsymbol{S}_k)$$

$$= \min_{i,k:k\neq i} \sum_{m=1}^{M} \operatorname{rank}(\boldsymbol{S}_i[m] - \boldsymbol{S}_k[m]). \quad (7)$$

Now, let us consider a noiseless channel. Given the observation vector $\boldsymbol{y}$ and the channel realization $\boldsymbol{h}$, it can be shown that the receiver will correctly *identify* the transmitted matrix $\boldsymbol{S}_i$ if and only if $\|\boldsymbol{h}(\boldsymbol{S}_i - \boldsymbol{S}_k)\|^2 > 0$ for all $k \neq i$. In other words, given the noise-free observation and the channel realization, the subspace $\mathcal{H}_c(i,k)$ of channel vectors where the receiver will not be able to differentiate between $\boldsymbol{S}_i$ and $\boldsymbol{S}_k$ is given by

$$\mathcal{H}_c(i,k) = \left\{ \boldsymbol{h} \in \mathbb{C}^{ML_t} : \boldsymbol{h}\boldsymbol{S}_i = \boldsymbol{h}\boldsymbol{S}_k \right\} \quad (8)$$

where $\mathbb{C}$ denotes the field of complex numbers. Intuitively, one would like to design the code such that the maximum dimension of $\mathcal{H}_c(i,k)$ over all $i, k \neq i$ is minimized. Observe that the subspace $\mathcal{H}_c(i,k)$ is equal to the left null space of $(\boldsymbol{S}_i - \boldsymbol{S}_k)$, and hence the dimension of $\mathcal{H}_c(i,k)$ is given as

$$\dim(\mathcal{H}_c(i,k)) = ML_t - \operatorname{rank}(\boldsymbol{S}_i - \boldsymbol{S}_k). \quad (9)$$

Combining (7) and (9), we get

$$d = ML_t - \max_{i,k:k\neq i} \dim(\mathcal{H}_c(i,k)). \quad (10)$$

Now, we exploit this interpretation to develop an analogous algebraic formulation for the diversity advantage in the noncoherent scenario. The main difference in this scenario is the unknown channel realization which dictates the use of the GLRT detector for jointly estimating the channel and transmitted data matrix from the received data. Again, in the noiseless case (i.e., $\boldsymbol{n} = \boldsymbol{0}$), the receiver will be able to correctly identify the transmitted space–time signal if $\|\boldsymbol{h}\boldsymbol{S}_i - \hat{\boldsymbol{h}}\boldsymbol{S}_k\|^2 > 0$ for **every** arbitrary $\hat{\boldsymbol{h}} \in \mathbb{C}^{ML_t}$ and $\boldsymbol{S}_k$ with $k \neq i$. In other words, if channel realization belongs to the set in (11) at the bottom of the page, then the receiver will not be able to differentiate between $\boldsymbol{S}_i$ and $\boldsymbol{S}_k$. Using Lemma 1 in Part A of the Appendix, one can easily see that the set of channel vectors $\mathcal{H}_{nc}(i,k)$ is a subspace of $\mathbb{C}^{ML_t}$. This leads to the following definition for the space–time diversity advantage in noncoherent channels

Definition 3.1: (**Algebraic Diversity Advantage**) We say that an STC achieves d levels of transmit diversity advantage in

an $L_t \times L_r$ **noncoherent** block-fading channel with M blocks per codeword if

$$d = ML_t - \max_{i,k:k\neq i} \dim(\mathcal{H}_{nc}(i,k)) \quad (12)$$

where the subspace $\mathcal{H}_{nc}(i,k)$ is defined in (11).

Based on this definition, we characterize the diversity advantage of an arbitrary noncoherent STC in the following proposition.

Proposition 3.1: A noncoherent space–time coded system achieves an algebraic transmit diversity advantage of d in the sense of Definition 3.1 if

$$d = \min_{i,k:k\neq i} \sum_{m=1}^{M} \left(\operatorname{rank}\left([\boldsymbol{S}_i[m]^T \boldsymbol{S}_k[m]^T]\right) - \operatorname{rank}(\boldsymbol{S}_k[m]) \right)$$
$$\quad (13)$$

$$= \min_{i,k:k\neq i} \left(\sum_{m=1}^{M} \operatorname{rank}\left([\boldsymbol{S}_i[m]^T \boldsymbol{S}_k[m]^T]\right) \right.$$
$$\left. - \max\left(\sum_{m=1}^{M} \operatorname{rank}(\boldsymbol{S}_i[m]), \sum_{m=1}^{M} \operatorname{rank}(\boldsymbol{S}_k[m]) \right) \right). \quad (14)$$

Proof: See Part A in the Appendix.

Unfortunately, due to the intractability of the probability of error analysis, we do not have a mathematical proof for the *equivalence* between the algebraic and probabilistic diversity advantages, similar to the coherent case, for arbitrary codes in noncoherent channels. This equivalence, however, is established for the special case of unitary codes in the following proposition.

Proposition 3.2: Definition 3.1 coincides with the definition of transmit diversity advantage based on the pairwise error probability of the GLRT receiver for unitary signaling in noncoherent Rayleigh-fading channels.

Proof: See Part B of the Appendix.

Moreover, the simulation results in Section V support our claim that there is an agreement between the asymptotic slope of the probability of error curve and the algebraic diversity advantage. Finally, we conclude this section with the following upper bound on the achievable algebraic diversity advantage.

Proposition 3.3: The algebraic transmit diversity advantage of a noncoherent STC is upper-bounded by $\min\left(ML_t, \left\lfloor \frac{T_s}{2} \right\rfloor\right)$.

Proof: See Part C of the Appendix.

IV. TRAINING-BASED SIGNALING SCHEMES

One of the attractive approaches for noncoherent channels is training-based signaling. The main advantage of such signaling schemes is that they lend themselves to an efficient suboptimal decoding scheme that exhibits a *clean* separation between the channel estimation and data detection stages. In this section, we first characterize the tradeoff between the transmission rate and

$$\mathcal{H}_{nc}(i,k) = \left\{ \boldsymbol{h} \in \mathbb{C}^{ML_t} : \exists \hat{\boldsymbol{h}} \in \mathbb{C}^{ML_t} \text{ s.t. } \boldsymbol{h}\boldsymbol{S}_i = \hat{\boldsymbol{h}}\boldsymbol{S}_k \right\}$$
$$= \left\{ \boldsymbol{h} \in \mathbb{C}^{ML_t} : \exists \tilde{\boldsymbol{h}}^T \in \left(\operatorname{Range}\left(\boldsymbol{S}_i^T\right) \cap \operatorname{Range}\left(\boldsymbol{S}_k^T\right) \right) \text{ s.t. } \boldsymbol{S}_i^T \boldsymbol{h}^T = \tilde{\boldsymbol{h}}^T \right\} \quad (11)$$

achievable diversity[2] for training-based schemes in block-fading channels. This characterization allows for extracting guidelines for the proper choice of the number of active transmit antennas. This result complements the Zheng–Tse guideline for the choice of the number of active antennas in ergodic channels [5].

A. The Diversity-Versus-Rate Tradeoff

In the training based space–time signaling approach, the transmitted signal in the mth block has the form $\boldsymbol{S}_i[m] = [\boldsymbol{T}[m]\boldsymbol{D}_i[m]]$ where $\boldsymbol{T}[m]$ is the $L_t \times T_{t,m}$ training matrix, $T_{t,m}$ denotes the number of training symbols transmitted from each antenna in the mth block, and $\boldsymbol{D}_i[m]$ is the information matrix. We further denote

$$r_{ik}[m] = \mathrm{rank}\left(\left[\boldsymbol{S}_i[m]^T \boldsymbol{S}_k[m]^T\right]\right) \tag{15}$$

$$r_k[m] = \mathrm{rank}(\boldsymbol{S}_k[m]). \tag{16}$$

One can easily show that

$$r_{ik}[m] = \mathrm{rank}(\boldsymbol{T}[m]) + \mathrm{rank}(\boldsymbol{D}_i[m] - \boldsymbol{D}_k[m]). \tag{17}$$

Moreover, one has [17]

$$r_k[m] \geq \max(\mathrm{rank}(\boldsymbol{T}[m]), \mathrm{rank}(\boldsymbol{D}_k[m])) \tag{18}$$

with equality if

$$\mathrm{Range}(\boldsymbol{T}[m]) \cap \mathrm{Range}(\boldsymbol{D}_k[m])$$
$$= \begin{cases} \mathrm{Range}(\boldsymbol{T}[m]), & \text{if } \mathrm{rank}(\boldsymbol{T}[m]) \leq \mathrm{rank}(\boldsymbol{D}_k[m]) \\ \mathrm{Range}(\boldsymbol{D}_k[m]), & \text{otherwise.} \end{cases} \tag{19}$$

These relations lead to the conclusion that

$$r_{ik}[m] - r_k[m] \leq \mathrm{rank}(\boldsymbol{D}_i[m] - \boldsymbol{D}_k[m]) \tag{20}$$

with equality if

$$\mathrm{Range}(\boldsymbol{T}[m]) \supseteq \mathrm{Range}(\boldsymbol{D}_k[m]). \tag{21}$$

Consequently, we have the following upper bound on the diversity advantage:

$$d = \min_{i,k:k \neq i} \sum_{m=1}^{M} (r_{ik}[m] - r_k[m])$$
$$\leq \min_{i,k:k \neq i} \sum_{m=1}^{M} \mathrm{rank}(\boldsymbol{D}_i[m] - \boldsymbol{D}_k[m]) \tag{22}$$

with equality if

$$\mathrm{Range}(\boldsymbol{T}[m]) \supseteq \mathrm{Range}(\boldsymbol{D}_i[m]), \qquad \forall i \text{ and } m = 1, \dots, M.$$

One should, therefore, construct $\boldsymbol{T}[m]$ and $\boldsymbol{D}_i[m]$ such that

$$\mathrm{Range}(\boldsymbol{T}[m]) \supseteq \mathrm{Range}(\boldsymbol{D}_i[m]), \qquad \forall i \text{ and } m = 1, \dots, M. \tag{23}$$

In this case, equality is achieved in (22), i.e.,

$$d = \min_{i,k:k \neq i} \sum_{m=1}^{M} \mathrm{rank}(\boldsymbol{D}_i[m] - \boldsymbol{D}_k[m]) \tag{24}$$

which coincides with the diversity advantage of the *coherent* STC $\{\boldsymbol{D}_i\}$ [16]. It is easy to see that the constraint in (23) implies

$$\mathrm{rank}(\boldsymbol{T}[m]) \geq \max_i \mathrm{rank}(\boldsymbol{D}_i[m]). \tag{25}$$

Let $t_m = \mathrm{rank}(\boldsymbol{T}[m])$, then $t_m \leq T_{t,m}$ because $\boldsymbol{T}[m]$ is an $L_t \times T_{t,m}$ matrix. In order to maximize the transmission rate, one should minimize the training time by setting $T_{t,m} = t_m$. In addition, we know that

$$L_t \geq t_m \geq \max_i \mathrm{rank}(\boldsymbol{D}_i[m])$$

and

$$\mathrm{Range}(\boldsymbol{T}[m]) \supseteq \mathrm{Range}(\boldsymbol{D}_i[m]), \qquad \forall i.$$

As a result, one can see that the maximum *achievable* diversity will not be reduced if we only activate t_m transmit antennas in the mth block. This leads to the following canonical structure for the transmission matrix in the mth block:

$$\boldsymbol{S}_i[m] = \begin{bmatrix} \boldsymbol{T}[m]_{t_m \times t_m} & \boldsymbol{D}_i[m]_{t_m \times (T_c - t_m)} \\ \boldsymbol{0}_{(L_t - t_m) \times t_m} & \boldsymbol{0}_{(L_t - t_m) \times (T_c - t_m)} \end{bmatrix} \tag{26}$$

where $1 \leq t_m \leq L_t$, and $\boldsymbol{T}[m] = \boldsymbol{I}_{t_m}$, the $t_m \times t_m$ identity matrix. Since

$$\mathrm{rank}(\boldsymbol{D}_i[m] - \boldsymbol{D}_k[m]) \leq \min(t_m, T_c - t_m)$$

the following relation holds:

$$d \leq \sum_{m=1}^{M} t_m \leq M \min(L_t, \lfloor T_c/2 \rfloor). \tag{27}$$

It should be noted that using training-based schemes might entail a slight loss in maximum achievable diversity ($M\lfloor T_c/2 \rfloor$ versus $\lfloor (MT_c)/2 \rfloor$) if $L_t > \lfloor T_c/2 \rfloor$ and T_c is odd valued.

We are now ready to find the optimum number of active antennas for each block, and hence, characterize the diversity-versus-rate tradeoff. Using the Singleton bound,[3] one can see that, for a given t_m, $m = 1, \dots, M$, the maximum rate of a training-based scheme achieving d levels of transmit diversity is

$$\eta = \left(\frac{T_s - \sum_{m=1}^{M} t_m}{T_s} \right) \left(\frac{\sum_{m=1}^{M} t_m - d + 1}{M} \right) \log_2 |\mathcal{D}| \tag{28}$$

where $|\mathcal{D}|$ is the number of points in the transmitted constellation(s).[4] Differentiating this expression with respect to $\sum_{m=1}^{M} t_m$ and equating it to zero gives

$$\sum_{m=1}^{M} t_m = \frac{T_s + d - 1}{2}. \tag{29}$$

Since the upper bound is a concave function of $\sum_{m=1}^{M} t_m$, this extremum point is a maximum. Combining this result with the condition on $\sum_{m=1}^{M} t_m$ in (27) and the fact that t_m must be integer valued, one can see that the optimum choice of set of

---

[2]From this point on, unless otherwise stated, we will refer to the algebraic diversity advantage in Definition 3.1 as diversity for brevity.

[3]The Singleton bound is used to find the maximum transmission rate for the coherent code $\{\boldsymbol{D}_i\}$ subject to a constraint on the diversity advantage [16].

[4]Here, we impose the symmetry constraint that the size of the transmitted constellation does not change with the temporal and spatial dimensions.

t_m^* that maximizes the transmission rate for a given diversity advantage d satisfy

$$\sum_{m=1}^{M} t_m^* = \min\left(\left\lfloor \frac{T_s + d - 1}{2} \right\rfloor, ML_t, M\lfloor T_c/2 \rfloor\right). \quad (30)$$

It is worth noting that the optimality condition in (30) only applies to the *sum* of the number of active transmit antennas and any distribution of this sum among the individual t_m will *allow* for achieving the optimal diversity-versus-rate tradeoff. The final ingredient in achieving the optimal tradeoff is the construction of the STC, i.e., $\{\boldsymbol{D}_i\}$, that achieves the optimal diversity-versus-rate tradeoff in *coherent* channels. Guidelines for constructing these codes are available in [18].

B. Affine Space–Time Constellations

By restricting ourselves to training-based approaches, we have taken a first step toward an efficient receiver implementation. The main advantage of this approach is the clean separation between the channel estimation and decoding stages. While this separation leads to a significant reduction in the receiver complexity in most cases, the complexity of the decoding stage may still be prohibitive in certain applications. This is specially true in high-rate applications for certain classes of coherent STCs with exponentially growing decoding complexities (e.g., the trellis STCs in [18]). The most common strategy for controlling the complexity of the decoder is to impose additional constraints on the structure of the coherent STC. In the sequel, we follow this approach and choose the coherent code to be linear over the field of complex numbers, i.e.,

$$\text{vec}(\boldsymbol{D}_i) = \boldsymbol{G}\underline{u}_i \quad (31)$$

where $\text{vec}(\boldsymbol{D}_i)$ is a vector containing the stacked columns of $\boldsymbol{D}_i[m]$, $m = 1, \ldots, M$, $\boldsymbol{G}$ is the generator matrix of the code, and $\underline{u}_i$ is the $\nu \times 1$ information vector whose elements are drawn independently from the constellation $\mathcal{U}$. In order to further facilitate low average complexity decoding using the sphere decoder [19], [20], we further impose the constraint that

$$\nu \leq \left(T_s - \sum_{m=1}^{M} t_m\right) \min(L_t, L_r)$$

[21]. The latter constraint ensures that the number of variables is less than or equal to the number of equations in the system of linear equations to be solved by the sphere decoder. In the sequel, we refer to such a decoder as a well-conditioned sphere decoder.

By restricting ourselves to this class of linear mappings, the tradeoff based on the Singleton bound may not be achievable any more. One should, therefore, revisit the guideline for the proper choice of the number of active antennas in (30) in the new scenario. By taking into account the linearity of the code and the constraint on the size of the input information vector, we obtain the following result.

Proposition 4.1: For a given transmit diversity advantage d, the maximum rate of an affine space–time constellation that allows for a well-conditioned sphere decoder in the decoding stage is given by

$$\eta = \frac{1}{T_s} \sum_{m=1}^{M} (T_c - t_m^*) \min(t_m^*, L_r) \log_2 |\mathcal{U}| \quad (32)$$

where $|\mathcal{U}|$ is the size of the input constellation, and t_m^* is given by

$$t_m^* = \begin{cases} \min(L_t, L_r, \lfloor T_c/2 \rfloor), & \text{if } d \leq M \min(L_r, \lfloor T_c/2 \rfloor) \\ \lceil d/M \rceil, & \text{if } d > ML_r \text{ and } 1 \leq m \leq \mu \\ \lfloor d/M \rfloor, & \text{if } d > ML_r \text{ and } \mu < m \leq M \end{cases} \quad (33)$$

where $\mu = \text{rem}(d/M)$.

Proof: See Part D of the Appendix.

Two observations are now in order.

1) The use of linear mapping, over the field of complex numbers, entails an expansion of the input constellation. The rate in (32) is, however, given in terms of the size of the input constellation, and hence, does not account for this constellation expansion. We believe that reporting the result in this form gives more insights on the fundamental diversity-versus-rate tradeoff since it avoids the dependence on a particular design for the linear mapper.

2) It is interesting to compare now the optimal choice of set of t_m^* in (30) and (33) to the optimal choice for the number of active transmit antennas in ergodic channels. In [5], Zheng and Tse showed, based on an information-theoretic argument, that the optimum number of active transmit antennas in noncoherent ergodic channels is $\min(L_t, L_r, \lfloor T_c/2 \rfloor)$. One can recover this result by letting $M \to \infty$ in (33). In the general case, however, (33) shows that in a delay-limited channel, this choice for the number of active antennas may imply a price in the diversity advantage. For example, in the case of affine signaling schemes with $L_r < \min(L_t, \lfloor T_c/2 \rfloor)$, the transmission rate is maximized by activating only L_r antennas. This choice implies the following upper bound on the achievable transmit diversity advantage $d \leq ML_r$. Therefore, activating only L_r antennas as suggested by [5] implies a loss in diversity advantage. This loss may result in a performance degradation at high SNR as illustrated with a numerical example in Section V. If we were to be interested only in maximizing the diversity advantage, the optimal choice of active antennas as given in (33) is $\min(L_t, \lfloor T_c/2 \rfloor)$ which agrees with the observation in [10]. In general, the optimal choice for the number of active antennas in delay-limited channels depends on the operating SNR (and the targeted diversity advantage), the transmission rate (in bits per channel use), and the complexity of the receiver (as illustrated by the difference between (28) and (32)).

One of the main ingredients in achieving the tradeoff (32) is the use of efficient coherent STCs. A universal framework for constructing such codes (i.e., the TAST design framework) was reported in [9]. For the sake of completeness, we will summarize, in the following, the main elements of this framework.

Recall that a thread is an allocation of space–time resources that does not have spatial interference within itself (i.e., at each time instant only one transmit antenna is active), spans all the time instants (T_s), and over time all the transmit antennas are equally utilized. A TAST constellation supporting L input symbols per channel use is composed of L independent threads (i.e., every thread carries an independent information stream) where
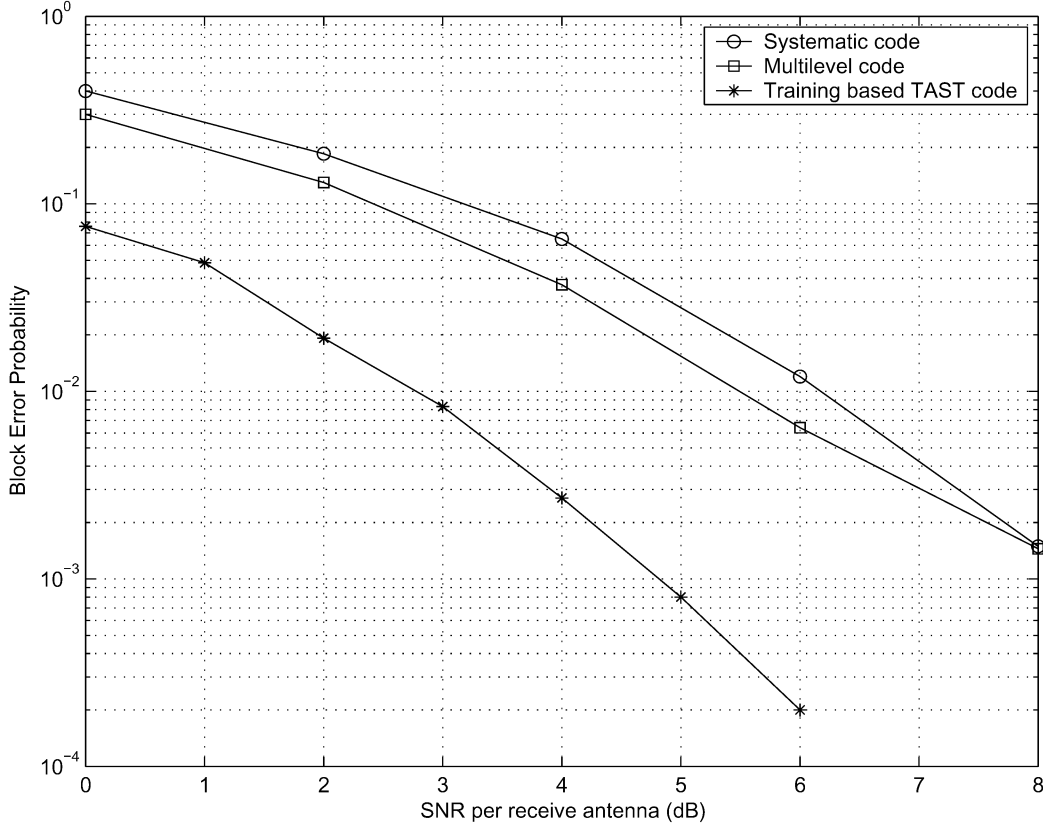
Fig. 1. Performance comparisons of noncoherent space–time constellations: $M = 1$, $L_t = 2$, $L_r = 10$, $T_c = 3$, and $\eta = 4/3$ bits per channel use.

a *scaled* single-input single-output (SISO) code (e.g., the diagonal algebraic space–time (DAST) signal [21], with rate one input symbol per channel use) is transmitted in each thread. The SISO codes are constructed such that each of them achieves full diversity in the absence of other threads. The scaling Diophantine factors are then chosen to ensure that the threads are *transparent* to each other. In [9], it was shown that full diversity is ensured if the scaling factors $\{1, \phi_1, \ldots, \phi_{L-1}\}$ are chosen to be simultaneously badly approximated by algebraic integers. Let $\mathcal{T}_{L_t, L, L}$ denote a full diversity coherent TAST constellation with L_t transmit antennas, L threads, and L input symbols per channel use. Note that choosing $L = \min(L_t, L_r)$ is intended to minimize the complexity of maximum-likelihood (ML) decoding for the coherent code. Here, we use the TAST constellations optimized for block-fading channels [9]. The affine constellation that achieves the optimal diversity-versus-rate tradeoff is now readily obtained by assigning the transmission matrix

$$\begin{bmatrix} \boldsymbol{I}_{t_m^*} & \mathcal{T}_{t_m^*, L_m^*, L_m^*}[m] \\ \boldsymbol{0}_{(L_t - t_m^*) \times t_m^*} & \boldsymbol{0}_{(L_t - t_m^*) \times (T_c - t_m^*)} \end{bmatrix} \quad (34)$$

to the mth block, where

$$L_m^* = \min(t_m^*, L_r),$$

$\boldsymbol{I}_{t_m^*}$ is the pilot matrix for training t_m^* active transmit antennas, and $\mathcal{T}_{t_m^*, L_m^*, L_m^*}[m]$ is the mth submatrix of the coherent TAST constellation $\mathcal{T}_{t_m^*, L_m^*, L_m^*}$ that achieves d levels of transmit diversity over the M blocks.

V. NUMERICAL RESULTS

In this section, we compare the block error performance of some exemplary training-based TAST constellations with a representative sample of the recently proposed noncoherent ST's in the literature.[5] The performances are plotted as a function of the average SNR per receive antenna. The TAST constellations utilize the full diversity SISO algebraic rotations [9]. Unless otherwise stated, the new TAST constellations are decoded via a two-stage process. In the first stage, the channel coefficients are estimated from the training (pilot) matrix based on a minimum mean-squared error (MMSE) criterion. These channel estimates are then used by the well-conditioned sphere decoder as the true channel coefficients when decoding the information stream. Throughout our simulation study, we used a Rayleigh-fading channel model.

Fig. 1 presents the performance of the training-based TAST constellation, the systematic unitary modulation in [22], and the multilevel constellation in [15]. In this figure, we assume a quasi-static fading channel $(M = 1)$ with $L_t = 2$, $L_r = 10$, $T_c = 3$, and $\eta = 4/3$ bits per channel use. It should be noted that the maximum achievable transmit diversity in this case is 1, and hence, we activate only one antenna in the proposed TAST constellation (i.e., the transmitted signals are in form $\boldsymbol{S}_i = [1 \, s_i(1) \, s_i(2)]$ where $s_i(k)$ is drawn from a quadrature amplitude modulation (QAM) constellation). Due to the simplicity of the proposed constellation and the low transmission rate in

---

[5]We restricted the reported results to this sample for brevity. In all the scenarios we considered, the TAST constellations either outperform or match the best known noncoherent codes.
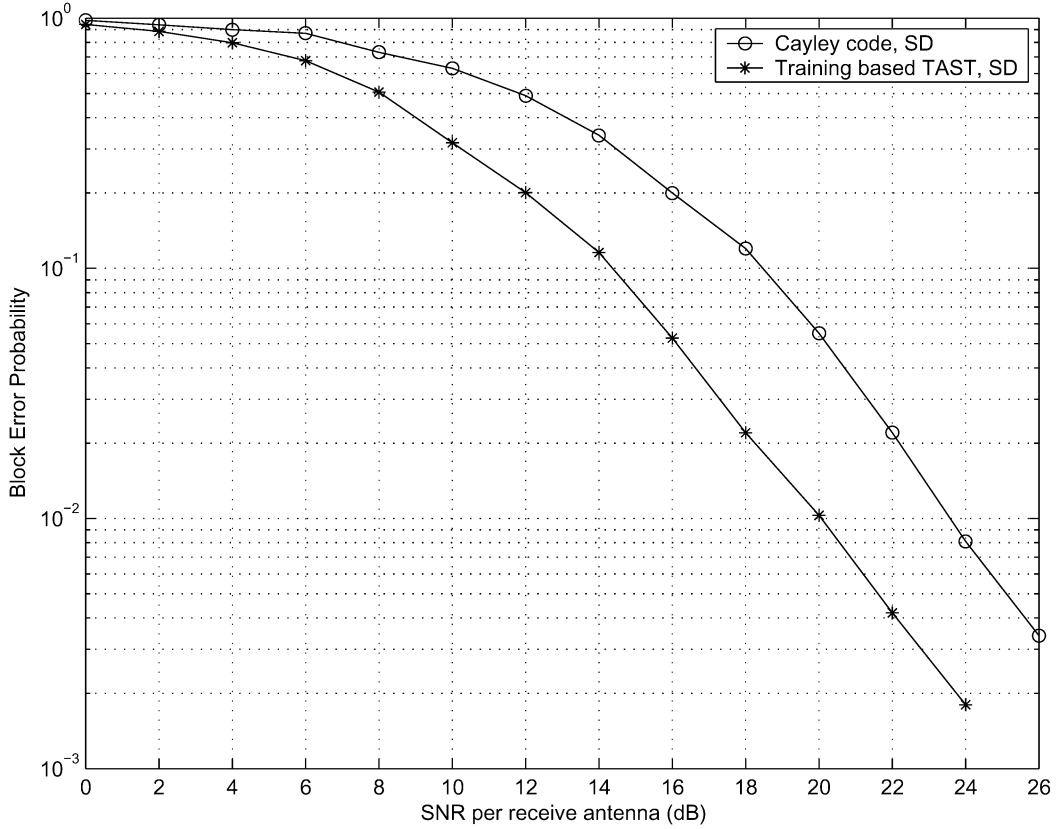
Fig. 2.   Performance comparisons of noncoherent space–time constellations: $M = 1$, $L_t = L_r = 2$, $T_c = 4$, and $\eta = 2$ bits per channel use.

this case, instead of using a sphere decoder for detecting the information stream, we employ an exhaustive search ML decoder after the channel estimation stage. From the figure, it is observed that the proposed scheme enjoys performance gains of 3 and 3.5 dB, at a block error rate of 0.01, over the multilevel constellation and the systematic unitary modulation, respectively.

In Fig. 2, the performance of the proposed approach is compared with that of the unitary signaling scheme based on the Cayley transform [13] in a quasi-static fading channel with $L_t = L_r = 2$, $T_c = 4$, and $\eta = 2$ bits per channel use. Both codes are decoded using the well-conditioned sphere decoder. From the figure, one can see that the training-based TAST scheme enjoys a performance advantage of about 4 dB over the Cayley transform constellation.

These numerical results demonstrate that the unitary signaling paradigm which has been accepted as a *de facto* standard for noncoherent STC design may entail a significant performance loss in delay-limited scenarios. One of the important lessons is that we can utilize the vast amount of research on STC design for coherent systems to construct *very good* noncoherent STCs that lend themselves to low-complexity decoding.

In Fig. 3, the performance of the training-based TAST constellation scheme for $M = 1$ (quasi-static) and $M = 2$ are compared for $L_t = L_r = 2$, $T_c = 4$, and $\eta = 4$ bits per channel use. To achieve this rate, both schemes utilize a 16-QAM input constellation. In both cases, the number of active transmit antennas is two. For the quasi-static scenario, a $d = 2$ TAST constellation with two threads and a 2×2 complex rotation matrix is used, whereas, for the $M = 2$ case, a $d = 4$ TAST constellation with

two threads and a 4×4 complex rotation matrix is used. From the figure, one can observe the significant performance gain resulting from the additional diversity.

Finally, we compare the performance of two training-based TAST constellations for $M = 1$, $L_t = 3$, $L_r = 2$, $T_c = 6$, and $\eta = 4$ bits per channel use in Fig. 4. In the training-based $\mathcal{T}_{3,2,2}$ scheme, all three transmit antennas are active and an optimized TAST constellation for $L_t = 3$, $L_r = 2$ is used. According to the information-theoretic results in [5], the optimum number of active transmit antennas in this case is $\min(L_t, L_r, \lfloor T_c/2 \rfloor) = 2$. In the training-based $\mathcal{T}_{2,2,2}$ scheme, only two transmit antennas are active and an optimized TAST constellation for $L_t = L_r = 2$ is used. In order to achieve the same rate, the training-based $\mathcal{T}_{3,2,2}$ scheme uses a 16-QAM input constellation, whereas the training-based $\mathcal{T}_{2,2,2}$ scheme uses an 8-QAM input constellation. As observed in the figure, the training-based $\mathcal{T}_{3,2,2}$ scheme achieves a larger diversity advantage, and hence, has a better performance at high SNR. This example illustrates that the proper choice of the number of active transmit antennas will depend, in general, on the SNR. Moreover, one can see that the optimal number of active transmit antennas in delay-limited channels may differ significantly from the optimal choice in ergodic channels [5].

VI. CONCLUDING REMARKS

In this paper, we have taken a fresh look at the problem of designing noncoherent STCs for delay-limited channels. Our approach is inspired by an algebraic perspective and, as a first step, allows for a generalized characterization of the diversity
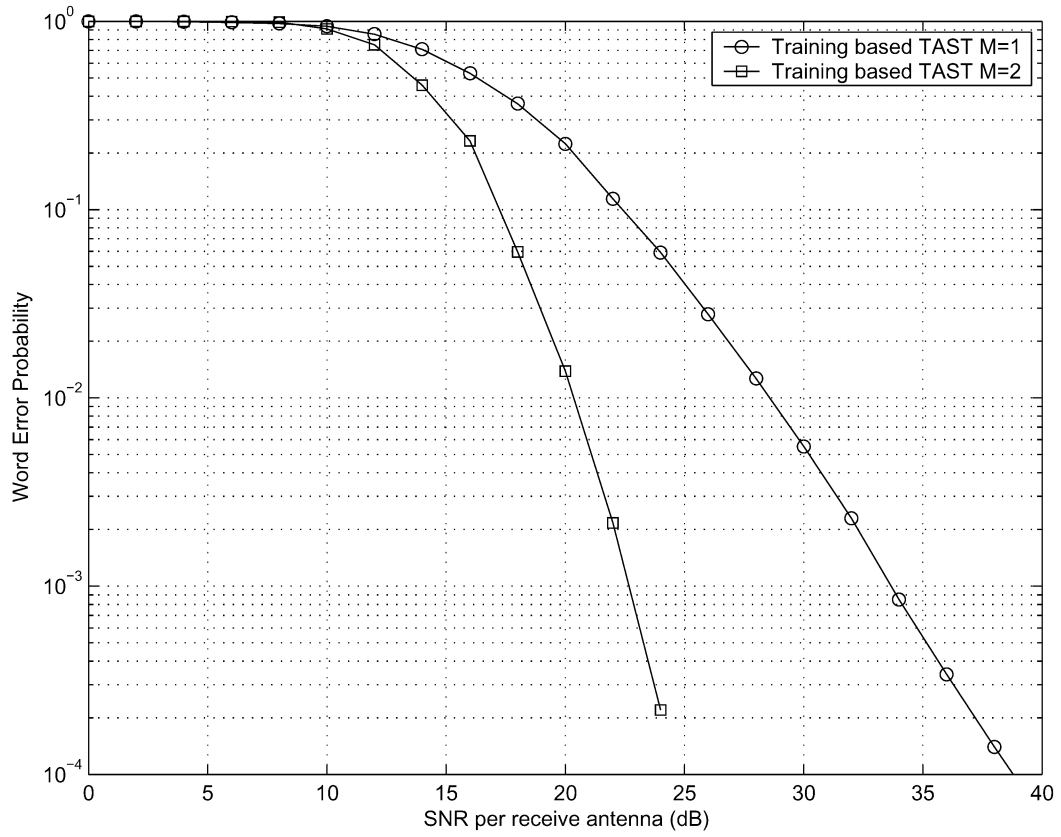
Fig. 3.   Performance comparisons of noncoherent space–time constellations: $L_t = 2$, $L_r = 2$, $T_c = 4$, and $\eta = 4$ bits per channel use over a block-fading channel with $M = 1$ and $M = 2$ blocks.
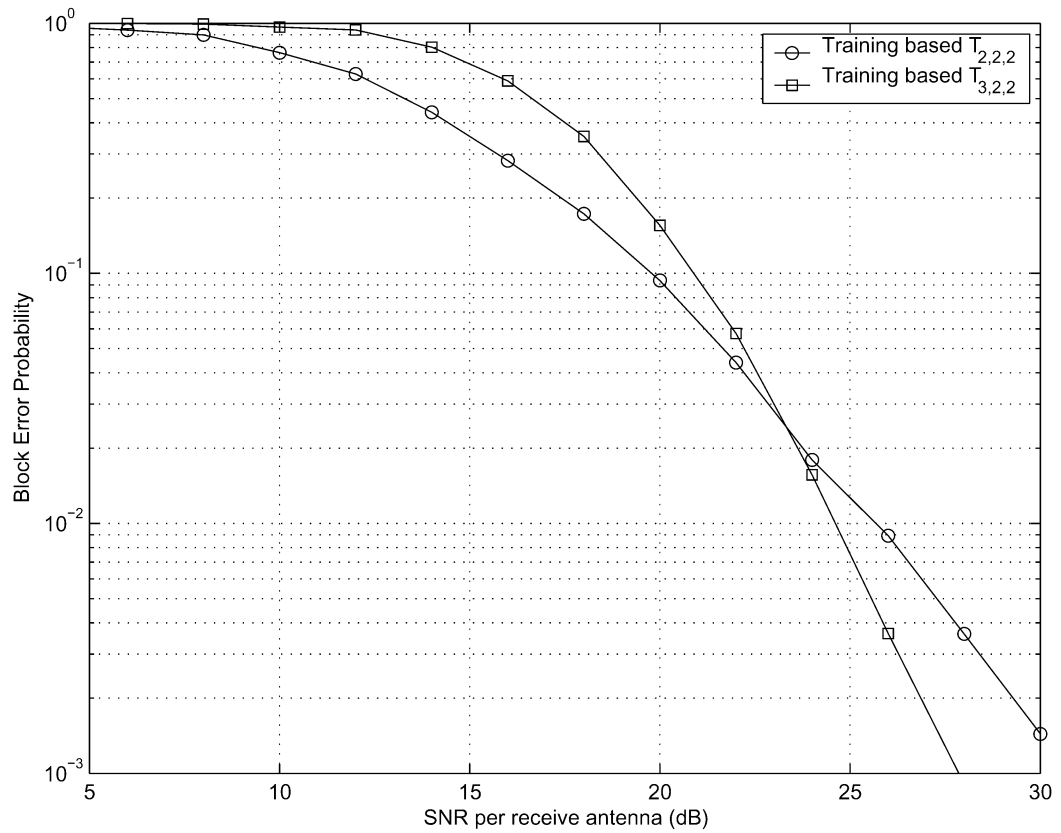

Fig. 4.   Performance comparisons of training-based TAST constellations: $M = 1$, $L_t = 3$, $L_r = 2$, $T_c = 6$, and $\eta = 4$ bits per channel use.

advantage of noncoherent STCs with arbitrary structure. We then shifted our attention to the class of training-based signaling schemes where we characterized the optimal diversity-versus-rate tradeoff using the Singleton bound. Within this class, we advocated the use of affine signals to exploit their amenability for low-complexity decoding algorithms. We presented exemplary designs for affine noncoherent constellations based on the recently proposed TAST coding framework. The new constellations were shown, through simulation results, to rival the best known noncoherent codes in all the considered scenarios.

One of the important conclusions that can be drawn from our work is that by relaxing the unitary constraint, significant performance and/or complexity reduction gains can be exploited in delay-limited noncoherent channels. Improvement over the schemes proposed here can be investigated in different research directions. For example, one can improve the suboptimal detector of affine signals (e.g., in relatively slow fading scenarios) by using iterative channel estimation and data detection. While one can see that the proposed constellations are only exemplary and limited to the class of training-based schemes, we believe that the algebraic machinery introduced here can be used to construct more sophisticated codes with, hopefully, better performance. In this light, one should view this paper as a first step toward a unified algebraic framework for constructing noncoherent STCs.

APPENDIX

In this appendix, the proofs of Propositions 3.1, 3.2, 3.3, and 4.1 are provided.

A. Proof of Proposition 3.1

The following lemma is a direct result from linear algebra [17] and is used to prove Proposition 3.1.

Lemma 1: Let A and B be two linear transformations from space $\mathcal{F}$ to space $\mathcal{G}$ where $\mathcal{F}$, $\mathcal{G}$ are vector spaces over $\mathbb{C}$. Define

$$\mathcal{V} = \text{Range}(A) \cap \text{Range}(B)$$

and

$$\mathcal{X} = \text{Preimage}(\mathcal{V}) = \{\boldsymbol{x} \in \mathcal{F} : \exists \boldsymbol{v} \in \mathcal{V} \text{ s.t. } A(\boldsymbol{x}) = \boldsymbol{v}\}.$$

Then, $\mathcal{X}$ is a subspace of $\mathcal{F}$ whose dimension is

$$\dim(\mathcal{X}) = \dim(\text{Null}(A)) + \dim(\mathcal{V}) \tag{35}$$
$$= \dim(\mathcal{F}) + \dim(\text{Range}(B))$$
$$- \dim(\text{Range}(A) + \text{Range}(B)). \tag{36}$$

Proof of Proposition 3.1: From Lemma 1

$$\dim(\mathcal{H}_{nc}(i,k)) = ML_t + \dim\left(\text{Range}\left(\boldsymbol{S}_k^T\right)\right)$$
$$- \dim\left(\text{Range}\left(\boldsymbol{S}_i^T\right) + \text{Range}\left(\boldsymbol{S}_k^T\right)\right)$$
$$= ML_t + \text{rank}(\boldsymbol{S}_k) - \text{rank}\left(\left[\boldsymbol{S}_i^T \boldsymbol{S}_k^T\right]\right) \tag{37}$$

where we used the identity [17]

$$\dim\left(\text{Range}\left(\boldsymbol{S}_i^T\right) + \text{Range}\left(\boldsymbol{S}_k^T\right)\right) = \text{rank}\left(\left[\boldsymbol{S}_i^T \boldsymbol{S}_k^T\right]\right).$$

From Definition 3.1

$$d = \min_{i,k:k\neq i}\left(\text{rank}\left(\left[\boldsymbol{S}_i^T \boldsymbol{S}_k^T\right]\right) - \text{rank}(\boldsymbol{S}_k)\right) \tag{38}$$
$$= \min_{i,k:k\neq i}\left(\text{rank}\left(\left[\boldsymbol{S}_i^T \boldsymbol{S}_k^T\right]\right)\right.$$
$$\left. - \max(\text{rank}(\boldsymbol{S}_i), \text{rank}(\boldsymbol{S}_k))\right) \tag{39}$$

where we used the observation

$$\min\left(\text{rank}\left(\left[\boldsymbol{S}_i^T \boldsymbol{S}_k^T\right]\right) - \text{rank}(\boldsymbol{S}_k),\right.$$
$$\left.\text{rank}\left(\left[\boldsymbol{S}_k^T \boldsymbol{S}_i^T\right]\right) - \text{rank}(\boldsymbol{S}_i)\right)$$
$$= \text{rank}\left(\left[\boldsymbol{S}_i^T \boldsymbol{S}_k^T\right]\right) - \max(\text{rank}(\boldsymbol{S}_i), \text{rank}(\boldsymbol{S}_k)). \tag{40}$$

From (39), one gets

$$d = \min_{i,k:k\neq i}\left(\sum_{m=1}^M \text{rank}\left(\left[\boldsymbol{S}_i[m]^T \boldsymbol{S}_k[m]^T\right]\right)\right.$$
$$\left. - \max\left(\sum_{m=1}^M \text{rank}(\boldsymbol{S}_i[m]), \sum_{m=1}^M \text{rank}(\boldsymbol{S}_k[m])\right)\right). \tag{41}$$

$\square$

B. Proof of Proposition 3.2

For unitary signaling, $\boldsymbol{S}_i \boldsymbol{S}_i^H = \boldsymbol{I}_{ML_t}$ with $\boldsymbol{I}_{ML_t}$ denoting the identity matrix of size ML_t. Therefore, all transmitted signals satisfy $\text{rank}(\boldsymbol{S}_i) = ML_t$. Let

$$r_{ik}[m] = \text{rank}([\boldsymbol{S}_i[m]^T \boldsymbol{S}_k[m]^T])$$
$$= \text{rank}([\boldsymbol{S}_i[m]^T \boldsymbol{S}_k[m]^T]^T)$$

then

$$r_{ik}[m] = \text{rank}\left(\left[\boldsymbol{S}_i[m]^T \boldsymbol{S}_k[m]^T\right]^T \left[\boldsymbol{S}_i[m]^H \boldsymbol{S}_k[m]^H\right]\right) \tag{42}$$
$$= \text{rank}\left(\begin{bmatrix} \boldsymbol{I}_{L_t} & \boldsymbol{A}[m] \\ \boldsymbol{A}[m]^H & \boldsymbol{I}_{L_t} \end{bmatrix}\right) \tag{43}$$

where $\boldsymbol{A}[m] = \boldsymbol{S}_i[m]\boldsymbol{S}_k[m]^H$. Let λ be an eigenvalue of $\left[\boldsymbol{S}_i[m]^T \boldsymbol{S}_k[m]^T\right]^T \left[\boldsymbol{S}_i[m]^H \boldsymbol{S}_k[m]^H\right]$, then one can show that $\lambda = 1 \pm \sqrt{\tilde{\lambda}}$ where $\tilde{\lambda}$ is the eigenvalue of $\boldsymbol{A}[m]^H \boldsymbol{A}[m]$. Let

$$q_{ik}[m] = \text{rank}(\boldsymbol{I}_{L_t} - \boldsymbol{A}[m]^H \boldsymbol{A}[m])$$

then $r_{ik}[m]$ is equal to the number of nonzero λ which is given by

$$r_{ik}[m] = 2 * \{\text{number of } \tilde{\lambda} \neq 1\}$$
$$+ \{\text{number of } \tilde{\lambda} = 1\} \tag{44}$$
$$= 2q_{ik}[m] + L_t - q_{ik}[m] \tag{45}$$
$$= L_t + \text{rank}\left(\boldsymbol{I}_{L_t} - \boldsymbol{S}_k[m]\boldsymbol{S}_i[m]^H \boldsymbol{S}_i[m]\boldsymbol{S}_k[m]^H\right). \tag{46}$$

From Proposition 3.1, the generalized transmit diversity advantage is

$$d = \min_{i,k:k\neq i} \sum_{m=1}^{M} r_{ik}[m] - ML_t \tag{47}$$

$$= \min_{i,k:k\neq i} \sum_{m=1}^{M} \mathrm{rank}\Big(\boldsymbol{I}_{L_t}$$
$$- \boldsymbol{S}_k[m]\boldsymbol{S}_i[m]^H \boldsymbol{S}_i[m]\boldsymbol{S}_k[m]^H\Big) \tag{48}$$

which is the well-known transmit diversity advantage definition of unitary signals in a noncoherent system based on the probability of error of the GLRT receiver [12]. □

C. Proof of Proposition 3.3

Let $\mathcal{S}$ be a noncoherent STC. The transmit diversity advantage of $\mathcal{S}$ in the sense of Definition 3.1 is

$$d_{\mathcal{S}} = \min_{\substack{\boldsymbol{S}_i,\boldsymbol{S}_k\in\mathcal{S} \\ \boldsymbol{S}_k\neq\boldsymbol{S}_i}} \Big(\mathrm{rank}\Big(\big[\boldsymbol{S}_i^T \boldsymbol{S}_k^T\big]\Big)$$
$$- \max(\mathrm{rank}(\boldsymbol{S}_i),\mathrm{rank}(\boldsymbol{S}_k))\Big). \tag{49}$$

One can easily show that

$$\mathrm{rank}\Big(\big[\boldsymbol{S}_i^T \boldsymbol{S}_k^T\big]\Big) \leq \mathrm{rank}(\boldsymbol{S}_i) + \mathrm{rank}(\boldsymbol{S}_k)$$

with equality if and only if

$$\mathrm{Range}(\boldsymbol{S}_i^T) \cap \mathrm{Range}(\boldsymbol{S}_k^T) = \{\boldsymbol{0}\}$$

i.e., $\boldsymbol{S}_i\boldsymbol{S}_k^H = \boldsymbol{0}$. As a result, for all $\boldsymbol{S}_i,\boldsymbol{S}_k \in \mathcal{S}$, the following holds:

$$\min(\mathrm{rank}(\boldsymbol{S}_i),\mathrm{rank}(\boldsymbol{S}_k))$$
$$\geq \mathrm{rank}\Big(\big[\boldsymbol{S}_i^T \boldsymbol{S}_k^T\big]\Big) - \max(\mathrm{rank}(\boldsymbol{S}_i),\mathrm{rank}(\boldsymbol{S}_k)) \tag{50}$$

with equality if and only if $\boldsymbol{S}_i\boldsymbol{S}_k^H = \boldsymbol{0}$. Therefore, one gets

$$d_{\mathcal{S}} \leq \min_{i,k:k\neq i}(\mathrm{rank}(\boldsymbol{S}_i),\mathrm{rank}(\boldsymbol{S}_k)). \tag{51}$$

Case i): $2ML_t \leq T_s$

In this case, $ML_t \leq T_s/2 < T_s$. For all $\boldsymbol{S}_i \in \mathcal{S}$

$$\mathrm{rank}(\boldsymbol{S}_i) \leq \min(ML_t, T_s) = ML_t.$$

From (51), $d_{\mathcal{S}} \leq ML_t$. Now we will show that there exists such an STC $\mathcal{S}$ that achieves this upper bound. Let $\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{T_c}\}$ be a basis for $\mathbb{C}^{T_c}$. Consider the code $\mathcal{S}$ with two codewords $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ such that

$$\boldsymbol{S}_1[1] = \cdots = \boldsymbol{S}_1[M] = [\boldsymbol{v}_1\cdots\boldsymbol{v}_{L_t}]^T$$

and

$$\boldsymbol{S}_2[1] = \cdots = \boldsymbol{S}_2[M] = [\boldsymbol{v}_{L_t+1}\ldots\boldsymbol{v}_{2L_t}]^T.$$

In this case

$$\mathrm{rank}(\boldsymbol{S}_1) = \mathrm{rank}(\boldsymbol{S}_2) = ML_t$$

since $\{\boldsymbol{v}_1,\ldots\boldsymbol{v}_{L_t}\}$ and $\{\boldsymbol{v}_{L_t+1},\ldots\boldsymbol{v}_{2L_t}\}$ are linearly independent and

$$\mathrm{rank}([\boldsymbol{S}_1^T \boldsymbol{S}_2^T]) = 2ML_t$$

since $\{\boldsymbol{v}_1,\ldots\boldsymbol{v}_{L_t},\boldsymbol{v}_{L_t+1},\ldots\boldsymbol{v}_{2L_t}\}$ is a linearly independent set. Therefore, $d_{\mathcal{S}} = ML_t$.

Case ii): $2ML_t > T_s$

Let $\boldsymbol{S}_i,\boldsymbol{S}_k \in \mathcal{S}$, then

$$\mathrm{rank}\Big(\big[\boldsymbol{S}_i^T \boldsymbol{S}_k^T\big]\Big) \leq \mathrm{rank}\Big(\big[\tilde{\boldsymbol{S}}_i^T \tilde{\boldsymbol{S}}_k^T\big]\Big)$$

for some $\tilde{\boldsymbol{S}}_i,\tilde{\boldsymbol{S}}_k \in \mathbb{C}^{ML_t\times T_s}$, such that $\mathrm{rank}(\tilde{\boldsymbol{S}}_i) = \mathrm{rank}(\boldsymbol{S}_i)$, $\mathrm{rank}(\tilde{\boldsymbol{S}}_k) = \mathrm{rank}(\boldsymbol{S}_k)$, and $\tilde{\boldsymbol{S}}_i\tilde{\boldsymbol{S}}_k^H = \boldsymbol{0}$. This implies

$$\min(\mathrm{rank}(\tilde{\boldsymbol{S}}_i),\mathrm{rank}(\tilde{\boldsymbol{S}}_k))$$
$$\geq \mathrm{rank}\Big(\big[\boldsymbol{S}_i^T \boldsymbol{S}_k^T\big]\Big) - \max(\mathrm{rank}(\boldsymbol{S}_i),\mathrm{rank}(\boldsymbol{S}_k)). \tag{52}$$

However,

$$\mathrm{rank}\Big(\big[\tilde{\boldsymbol{S}}_i^T \tilde{\boldsymbol{S}}_k^T\big]\Big) = \mathrm{rank}(\tilde{\boldsymbol{S}}_i) + \mathrm{rank}(\tilde{\boldsymbol{S}}_k)$$
$$\leq \min(2ML_t, T_s) = T_s.$$

Since

$$2\min(\mathrm{rank}(\tilde{\boldsymbol{S}}_i),\mathrm{rank}(\tilde{\boldsymbol{S}}_k)) \leq \mathrm{rank}(\tilde{\boldsymbol{S}}_i) + \mathrm{rank}(\tilde{\boldsymbol{S}}_k) \tag{53}$$

it follows that

$$\min(\mathrm{rank}(\tilde{\boldsymbol{S}}_i),\mathrm{rank}(\tilde{\boldsymbol{S}}_k)) \leq T_s/2.$$

From (52), one gets $d_{\mathcal{S}} \leq \lfloor T_s/2 \rfloor$. Now we will show that this bound is achievable. Again, we let $\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{T_c}\}$ be a basis for $\mathbb{C}^{T_c}$. Then we consider the code $\mathcal{S}$ with two codewords $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ such that

$$\boldsymbol{S}_1[2(\ell-1)+1] = \boldsymbol{S}_2[2\ell]$$
$$= \big[\boldsymbol{v}_1\cdots\boldsymbol{v}_{\lfloor T_c/2\rfloor}\boldsymbol{0}_{T_c\times(L_t-\lfloor T_c/2\rfloor)}\big]^T$$

and

$$\boldsymbol{S}_1[2\ell] = \boldsymbol{S}_2[2(\ell-1)+1]$$
$$= \big[\boldsymbol{v}_{\lfloor T_c/2\rfloor+1}\ldots\boldsymbol{v}_{T_c}\,\boldsymbol{0}_{T_c\times(L_t-\lceil T_c/2\rceil)}\big]^T$$

$\ell = 1,\ldots,\lceil M/2 \rceil$. It can be shown that $\mathrm{rank}([\boldsymbol{S}_1^T \boldsymbol{S}_2^T]) = T_s$. Furthermore, $\mathrm{rank}(\boldsymbol{S}_1) = \mathrm{rank}(\boldsymbol{S}_2) = T_s/2$ if M is even, and $\mathrm{rank}(\boldsymbol{S}_1) = (M-1)T_c/2 + \lfloor T_c/2 \rfloor$ and $\mathrm{rank}(\boldsymbol{S}_2) = (M-1)T_c/2 + \lceil T_c/2 \rceil$ if M is odd. In this case, if T_c is even, $\mathrm{rank}(\boldsymbol{S}_1) = \mathrm{rank}(\boldsymbol{S}_2) = T_s/2$ and if T_c is odd, $\mathrm{rank}(\boldsymbol{S}_1) = \lfloor T_s/2 \rfloor$ and $\mathrm{rank}(\boldsymbol{S}_2) = \lceil T_s/2 \rceil$. Therefore, $d_{\mathcal{S}} = T_s - \lceil T_s/2 \rceil = \lfloor T_s/2 \rfloor$.

Combining Cases i) and ii), one gets

$$d_{\mathcal{S}} \leq \min(ML_t, \lfloor T_s/2 \rfloor). \quad \square$$

D. Proof of Proposition 4.1

From the constraint on the size of the input information vector, the maximum rate of an affine space–time constellation with t_m active transmit antennas in each block is given by

$$\frac{1}{M} \sum_{m=1}^{M} \left(\frac{T_c - t_m}{T_c}\right) \min(t_m, L_r) \log_2 |\mathcal{U}| \tag{54}$$

bits per channel use, which is the average of the information rate at each block. Now we need to find the optimum choice of set of t_m^* that maximizes this value. So the optimization problem at our hands is

$$\max_{t_m} \sum_{m=1}^{M} (T_c - t_m) \min(t_m, L_r) \tag{55}$$

subject to

$$1 \leq t_m \leq L_t, \text{ with } t_m \text{ integer} \tag{56}$$

and

$$d \leq \sum_{m=1}^{M} t_m \leq M \min(L_t, \lfloor T_c/2 \rfloor). \tag{57}$$

Note that a linear coherent STC that achieves d levels of diversity with t_m active transmit antennas in mth block is used. In Section V, we construct examples of such codes.

Case i): $L_r \leq \lfloor T_c/2 \rfloor$ and $d \leq ML_r$

When $L_r \leq \lfloor T_c/2 \rfloor$, $(T_c - t_m) \min(t_m, L_r)$ is maximized when $t_m = L_r$. Since with this choice, the sum of t_m is larger than or equal to d, one can see that this is the optimum choice (i.e., $t_m^* = \min(L_r, L_t)$).

Case ii): $L_r > \lfloor T_c/2 \rfloor$ and $d \leq M \lfloor T_c/2 \rfloor$

When $L_r > \lfloor T_c/2 \rfloor$, $(T_c - t_m) \min(t_m, L_r)$ is maximum at $t_m = \lfloor T_c/2 \rfloor$. Since with this choice, the sum of t_m is larger than or equal to d, it follows that this is the optimum choice (i.e., $t_m^* = \min(\lfloor T_c/2 \rfloor, L_t)$).

Case iii): $M \lfloor T_c/2 \rfloor \geq d > ML_r$

From Case i), it follows that the optimal choice satisfies $t_m^* \geq L_r$. Since the diversity achieved in any fading block is upperbounded by $\min(t_m, \lfloor T_c/2 \rfloor)$, the optimal choice satisfies $t_m \leq \lfloor T_c/2 \rfloor$. With simple manipulations, it can be shown that all the choices of set of t_m satisfying $L_r \leq t_m \leq \lfloor T_c/2 \rfloor$ and $\sum_{m=1}^{M} t_m = d$ will result in the same *maximum* transmission rate. One particular choice that satisfies these constraints and enjoys symmetry is given in the proposition.

From Cases i)–iii), one gets the optimum choice of set of t_m^* as in (33). $\quad\square$

REFERENCES

[1] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecomm.*, vol. 10, pp. 585–595, Nov.–Dec. 1999.

[2] G. J. Foschini and M. Gans, "On the limits of wireless communication in a fading environment," *Wireless Personal Commun.*, vol. 6, pp. 311–355, Mar. 1998.

[3] V. Tarokh, N. Seshadri, and A. R. 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.

[4] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–158, Jan. 1999.

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

[6] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 543–563, Mar. 2000.

[7] V. Tarokh and I.-M. Kim, "Existence and construction of noncoherent unitary space-time codes," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3112–3117, Dec. 2002.

[8] M. Brehler and M. K. Varanasi, "Training-codes for the noncoherent multi-antenna block-Rayleigh-fading channel," in *Proc. Conf. Information Sciences and Systems (CISS)*, Baltimore, MD, Mar. 2003.

[9] H. El Gamal and M. O. Damen, "Universal space–time coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1097–1119, May 2003.

[10] P. Dayal, M. Brehler, and M. K. Varanasi, "Leveraging coherent space–time codes for noncoherent communication via training," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 2058–2080, Sep. 2004.

[11] 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," in *Proc. IEEE Vehicular Technology Conf.*, Atlanta, GA, Apr. 1996, pp. 136–140.

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

[13] Y. Jing and B. Hassibi, "Unitary space-time modulation via Cayley transform," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2891–2904, Nov. 2003.

[14] M. Brehler and M. K. Varanasi, "Asymptotic error probability analysis of quadratic receivers in Rayleigh-fading channels with applications to a unified analysis of coherent and noncoherent space-time receivers," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2383–2399, Sep. 2001.

[15] 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.

[16] H. El Gamal and A. R. Hammons Jr, "On the design of algebraic space-time codes for block fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 151–163, Jan. 2003.

[17] G. Strang, *Linear Algebra and Its Applications*, 3rd ed. San Diego, CA: Harcourt Brace Jovanavich, 1988.

[18] H. El Gamal and A. R. Hammons Jr, "On the theory of space-time codes for PSK modulation," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 524–542, Mar. 2000.

[19] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Math..Comput.*, vol. 44, pp. 463–471, Apr. 1985.

[20] M. O. Damen, A. Chkeif, and J.-C. Belfiore, "Lattice codes decoder for space-time codes," *IEEE Commun. Lett.*, vol. 4, no. 5, pp. 161–163, May 2000.

[21] M. O. Damen, K. Abed-Meraim, and J.-C. Belfiore, "Diagonal algebraic space–time block codes," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 628–636, Mar. 2002.

[22] B. M. Hochwald, T. L. Marzetta, T. R. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 1962–1973, Sep. 2000.