Space-Time Coding: an Overview

Giuseppe Caire, Petros Elia, and K. Raj Kumar　.

Abstract— **This work provides an overview of the fundamental aspects and of some recent advances in space-time coding (STC). Basic information theoretic results on Multiple-Input Multiple-Output (MIMO) fading channels, pertaining to capacity, diversity, and to the optimal Diversity-Multiplexing Tradeoff (DMT), are reviewed. The code design for the quasi-static, outage limited, fading channel is recognized as the most challenging and innovative with respect to traditional "Gaussian" coding. Then, a survey of STC constructions is presented. This culminates with the description of families of codes that are optimal with respect to the DMT criterion and have error performance that is very close to the information theoretic limits. The paper concludes with some important recent topics, including open problems in STC design.**

Keywords: Space-Time Coding, MIMO systems, fading channels, Diversity-Multiplexing Tradeoff.

I. INTRODUCTION

Since the seminal work of Telatar [1], Foschini and Gans [2], Tarokh *et al.* [3], and Guey *et al.* [4], multiple antenna transmission/reception has emerged as a key tool to achieving high spectral and power efficiency in wireless communications. Loosely speaking, signaling schemes that exploit both the classical Shannon degrees of freedom (time-frequency) and the additional spatial degrees of freedom (antennas) are referred to as *Space-Time Codes* (STC), a term popularized by [3]. The literature on STC is extremely vast (an incomplete sample is provided, for example, by [5], [6], [7], [8] and references therein).

Several settings have been developed on the basis of different physical channel models and, for each setting, information theoretic results and associated coding schemes have been addressed.

The general complex baseband channel for space-time signaling with M transmit and N receiving antennas is given by the Multi-Input Multi-Output (MIMO) Gaussian linear channel model[1]

$$\mathbf{y}_t = \sqrt{\rho}\mathbf{H}_t\mathbf{x}_t + \mathbf{w}_t, \quad t = 1, \ldots, T \qquad (1)$$

where T denotes the code block length (expressed in "channel uses"), $\{\mathbf{x}_t \in \mathbb{C}^M : t = 1, \ldots, T\}$ are the transmitted vectors, $\{\mathbf{y}_t \in \mathbb{C}^N : t = 1, \ldots, T\}$ are the received vectors, $\{\mathbf{w}_t \in \mathbb{C}^N : t = 1, \ldots, T\}$ are the channel Gaussian noise vectors,

G. Caire, P. Elia, and K. R. Kumar are with the Department of Electrical Engineering, University of Southern California, Los Angeles, California (e-mail: {caire, elia, rkkrishn}@usc.edu)

[1]**Notation:** $^{\mathsf{T}}$ denotes transpose and $^{\mathsf{H}}$ denotes Hermitian transpose. The notation $\mathbf{v} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ indicates that $\mathbf{v}$ is a proper [9] complex Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. For real Gaussian random vector we use the notation $\mathbf{v} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

assumed temporally and spatially white with i.i.d. entries $\sim \mathcal{CN}(0, 1)$, and $\{\mathbf{H}_t : t = 1, \ldots, T\}$ is the $N \times M$ channel matrix process with the elements $h_{i,j,t}$ representing the fading coefficients between the j-th transmit and the i-th receive antenna at time t. Without loss of generality, we assume a normalized channel such that $\frac{1}{TNM} \sum_{t=1}^{T} \mathbb{E}[\mathrm{tr}(\mathbf{H}_t^{\mathsf{H}}\mathbf{H}_t)] = 1$. We hasten to say that the above model encompasses also the case of a general frequency and time-selective fading channel, where the index t denotes an "abstract" index corresponding to some predetermined sequence of time-frequency bins, obtained for example by using OFDM. In this case, $\mathbf{H}_t$ denotes the MIMO channel response at time-frequency bin t. The following average transmit power constraint is enforced:

$$\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[|\mathbf{x}_t|^2\right] \leq 1. \qquad (2)$$

With the above normalizations, ρ represents the average transmit SNR (total transmit power divided by the noise power spectral density), that corresponds to the average received SNR per receiving antenna.

With respect to code construction, it turns out that the optimal code design varies quite significantly depending on the basic model assumptions. Section II provides an overview of the main settings and the related information theory and coding results. Section III focuses on the case of quasi-static fading, originally considered in [3], for which the asymptotic performance in high-SNR is characterized by the *diversity-multiplexing tradeoff* (DMT) studied in [10]. This case offers perhaps the richest and most original coding design problem, that has been the subject of very active research in the recent past. Section IV reviews a number of space-time coding constructions some of which achieve the optimal DMT of [10]. Finally, Section V presents some recent research thrusts. Since most of this work is still on-going, this section is intended more as a motivation for future work than as a comprehensive overview of a well settled set of results. Conclusions are pointed out in Section VI. Due to space limitation and to the tutorial nature of this paper, we omit the proofs of the results and we provide pointers to the corresponding literature.

II. RELIABLE COMMUNICATIONS WITH MULTIPLE ANTENNAS

The early information theoretic analysis of the MIMO channel (1) focused on the *ergodic* capacity. The assumption underlying the ergodic setting is that the random matrix process $\{\mathbf{H}_t\}$ is stationary and ergodic, and that the block length T is arbitrarily large. As a matter of fact, as of today the capacity of such channels is generally still unknown and only the low-SNR (see for example [11]) and high-SNR (see

for example [12], [13], [14]) have been characterized through asymptotically tight bounds.

At the other extreme of the channel statistics assumptions treated in the literature, we find the quasi-static block-fading model, where $\mathbf{H}_t = \mathbf{H}$, a random but constant matrix, for any (arbitrarily long) block length T. The information theoretic treatment of this channel model requires a more general capacity formula than the standard *information-stable* (in brief, ergodic) setting mentioned above. For channel statistics practically relevant in wireless communications, it turns out that the Shannon capacity is identically zero and the ϵ-capacity (also known as "outage capacity") provides the limiting performance behavior of coding. Interestingly, the literature on STC design has mainly focused on the latter model (quasi-static block-fading), while information theory has mostly investigated the capacity of the former model (ergodic). In the next two subsections we give an account of the some relevant results, and point out why STC design for the quasi-static block-fading channel presents some unique and novel problems in coding theory.

A. Ergodic fading

Given the fact that the general ergodic channel model appears to be very hard to analyze, research has focused and (almost) completely solved a much simpler problem that nevertheless yields very relevant results in practical settings, as it shall be argued in the following. The simplified model considers the augmented channel with output $\{\mathbf{y}_t, \mathbf{H}_t\}$, that is, it assumes perfect *Channel State Information* at the Receiver (CSIR). This is representative of systems based on "coherent" detection under the simplifying assumption that the channel matrix can be estimated very accurately.

For high SNR, if the channel is a strictly band-limited Doppler process [15] with single-side Doppler bandwidth $B/2$, the capacity is asymptotically given by the perfect CSIR capacity multiplied by the factor $(1 - M/L)$, where $L = \lfloor W/B \rfloor$ is the minimum pilot insertion period in order to sample the channel process at the Nyquist rate, and where $W/2$ denotes the single-side bandwidth of the signal [16], [13], [17]. In practice, $B \approx 100$ Hz and $W \approx 1$ MHz (see the discussion on these order of magnitude provided in [18]), therefore $L \approx 10^4$. This implies that for most cases that are relevant in practice, in the high-SNR high-spectral efficiency region where the application of MIMO systems is typically advocated, the cost of explicitly estimating the channel is very small. We conclude that the approximation of perfect CSIR yields indeed very relevant results in practice and has the advantage of simplicity for information theoretic analysis and code design. Therefore, this paper focuses on this case, with the caveat that the results presented here might not be relevant for cases where the channel matrix process is not a band-limited Doppler process (e.g., $\{\mathbf{H}_t\}$ evolves according to a Gauss-Markov AR-1 process), and for the region of low SNR and low spectral efficiency (e.g., in the Ultra-Wide-Band case [11], [19]).

From standard information theoretic arguments it follows

that the ergodic capacity under the perfect CSIR assumption[2] is given by [1]

$$C_{\mathrm{erg}}(\rho) = \max_{\boldsymbol{\Sigma}_x \geq 0, \, \mathrm{tr}(\boldsymbol{\Sigma}_x) \leq 1} \mathbb{E}\left[\log\left(1 + \rho \mathbf{H} \boldsymbol{\Sigma}_x \mathbf{H}^{\mathsf{H}}\right)\right] \quad (3)$$

where $\mathbf{H}$ is a random matrix with pdf $p_H(\cdot)$, the first-order distribution of the process $\{\mathbf{H}_t\}$ that does not depend on t by stationarity. Notice that the ergodic capacity with perfect CSIR depends only on the first-order statistics of the channel process. As long as the channel process is stationary and ergodic, its correlation w.r.t. t is irrelevant to capacity. For the class of isotropically right-invariant channels, i.e., channels such that $\mathbf{HU}$ and $\mathbf{H}$ are identically distributed for any fixed unitary matrix $\mathbf{U}$ (independent of $\mathbf{H}$), [1] shows that $\boldsymbol{\Sigma}_x = \frac{1}{M}\mathbf{I}$ is the optimal input covariance matrix. In particular, this is the case for the Rayleigh i.i.d. fading model, where

$$p_H(\mathbf{H}) = \frac{1}{\pi^{MN}} \exp\left(-\mathrm{tr}\left(\mathbf{H}^{\mathsf{H}}\mathbf{H}\right)\right). \quad (4)$$

In general, the mutual information

$$I_{\mathrm{erg}}(\rho) \triangleq \mathbb{E}\left[\log \det\left(\mathbf{I} + \frac{\rho}{M}\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)\right] \quad (5)$$

is the supremum of rates achievable by the ensemble of random Gaussian codes that send uncorrelated symbols with uniform power allocation over the antennas.

A different point of view on the optimality of the uniform power allocation is provided in [20], in the context of a MIMO *compound* channel where the channel matrix is deterministic and fixed but unknown to the transmitter, and can take any arbitrary value in some constraint set $\mathcal{H}$ satisfying the property that if $\mathbf{H} \in \mathcal{H}$ then also $\mathbf{HU} \in \mathcal{H}$, for any unitary matrix $\mathbf{U}$. It turns out that the compound MIMO channel capacity in this case is also achieved by the input covariance matrix $\boldsymbol{\Sigma}_x = \frac{1}{M}\mathbf{I}$. Furthermore, using the fact that $\log \det(\cdot)$ is concave and non-decreasing on the cone of positive semidefinite matrices, for any channel distribution $p_H(\cdot)$ it follows that

$$I_{\mathrm{erg}}(\rho) \leq C_{\mathrm{erg}}(\rho) \leq I_{\mathrm{erg}}(M\rho). \quad (6)$$

Hence, the penalty of uniform power allocation is at most $10 \log_{10} M$ dB in SNR. This represents a vanishing fraction of capacity in the region of large SNR, where capacity grows like $C_{\mathrm{erg}}(\rho) \approx \min\{M, N\} \log \rho$.

Considerable effort (see for example [21], [22], [6], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32]) has been devoted to determine the optimal input covariance matrix $\boldsymbol{\Sigma}_x$ in (3) for various channel first-order statistics $f_H(\cdot)$ despite the fact that, practically, it is not very relevant for the system design. The following channel statistics have been considered: 1) the correlated "Rayleigh" fading channel, where $\mathbf{H}$ has zero-mean Gaussian elements jointly distributed according to some covariance matrix; 2) the Ricean fading channel where $\mathbf{H} = \mathbf{ab}^{\mathsf{H}} + \mathbf{S}$, with $\mathbf{a}$ and $\mathbf{b}$ deterministic vectors (line-of-sight component), and $\mathbf{S}$ with zero-mean Gaussian i.i.d. elements (scattering component); 3) the Ricean correlated channel model where $\mathbf{S}$ has correlated elements.

---

[2]**Notation:** we use natural logarithms and measure information rates in nats per channel use, unless said otherwise.

From the code design viewpoint, achieving ergodic capacity does not present any conceptual difficulty. It can be shown that $C_{\mathrm{erg}}(\rho)$ is achieved by the random Gaussian code ensemble such that codewords $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ are generated independently with i.i.d. components, with $\mathbf{x}_t \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_x)$ for the capacity-achieving input covariance $\boldsymbol{\Sigma}_x$. Instances in this ensemble can be generated by the following conceptually simple "spatial-shaping" scheme: 1) generate codes in the Gaussian random coding ensemble that achieves the capacity of the classical unfaded single-input single-output AWGN channel, with i.i.d. components $\sim \mathcal{CN}(0,1)$ and block length MT; 2) format the MT code symbols into a space-time code array $\mathbf{X}'$ of dimension $M \times T$; 3) obtain the actual space-time codeword $\mathbf{X}$ by letting $\mathbf{X} = \boldsymbol{\Sigma}_x^{1/2} \mathbf{X}'$, where $\boldsymbol{\Sigma}_x^{1/2}$ is a square-root factor of $\boldsymbol{\Sigma}_x$ (e.g., obtained by Karhunen-Loeve or Cholesky decomposition).

The above conceptual construction suggests that good codes for the *ergodic* MIMO channel can be easily obtained from standard good codes designed for the AWGN channel via simple signal processing. In fact, the best known schemes for the ergodic setting have been obtained by using powerful codes (e.g., turbo-codes and LDPC and the like) mapped over standard QAM/PSK modulation constellations by standard bit-interleaved coded modulation [33] and sent directly over the M transmit antennas possibly through a linear precoder (the number of references dealing with this construction is overwhelming, see for example [34], [35], [36], [37], [38] and references therein for a very partial list). Here, research has focused mostly on the efficient signal-processing at the receiver rather than on the code design itself. In [34], [36], [39], [37], [38] the use of a soft-in soft-out Sphere-Decoder for the MIMO channel is advocated, in order to implement an iterative space-time decoder analogous to the turbo-equalization scheme proposed for ISI channels [40]. In order to avoid the heavy computational complexity of the soft-in soft-out MIMO detectors, a layered structure with *successive interference cancellation* can be exploited. The mutual information chain-rule [41]

$$I(\mathbf{x}; \mathbf{y}|\mathbf{H}) = \sum_{j=1}^{M} I(x_j; \mathbf{y}|\mathbf{H}, x_1, \ldots, x_{j-1})$$

applied to (3) yields the decomposition

$$\mathbb{E}\left[\log \det\left(\mathbf{I} + \frac{\rho}{M}\mathbf{H}\mathbf{H}^{\mathsf{H}}\right)\right] = \sum_{j=1}^{M} \mathbb{E}[\log(1 + \beta_j)] \quad (7)$$

where β_j is the Signal to Interference plus Noise Ratio (SINR) at the output of a Minimum Mean-Square Error (MMSE) detector that estimates the j-th antenna signal x_j assuming $x_1, \ldots, x_{j-1}$ perfectly known and treating $x_{j+1}, \ldots, x_M$ as Gaussian interference. The above mutual information decomposition suggests the horizontal layered scheme represented in Fig. 1 (encoding) and Fig. 2 (decoding). The MMSE decision-feedback receiver structure of Fig. 2, also known as "nulling and canceling" approach, is usually referred to as the "V-BLAST" receiver [42]. With the V-BLAST architecture, the MIMO soft-output detection problem is reduced to a sequence

of soft-output detectors for the resulting single-input single-output channels obtained at each cancellation stage, along with the additional (small) complexity of MMSE filtering and interference cancellation. If we choose the coding rates of the component codes in the layered scheme of Fig. 1 such that layer j has rate R_j slightly less than the corresponding ergodic rate $\mathbb{E}[\log(1 + \beta_j)]$, then the probability of decoding error at each component decoder is very small and the effect of error propagation is negligible. This rate allocation is a common trait of layered coding architectures with multistage decoding [43].
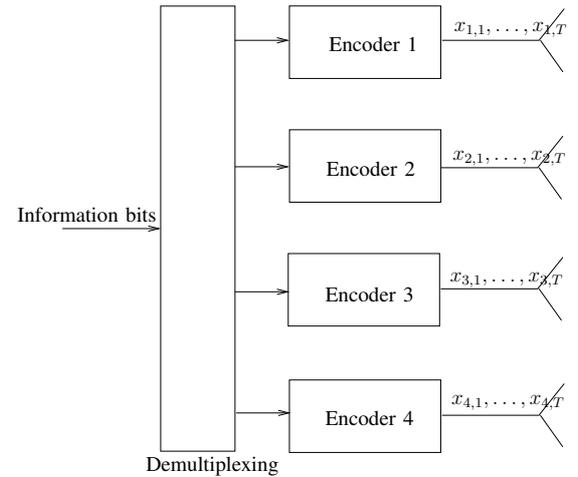

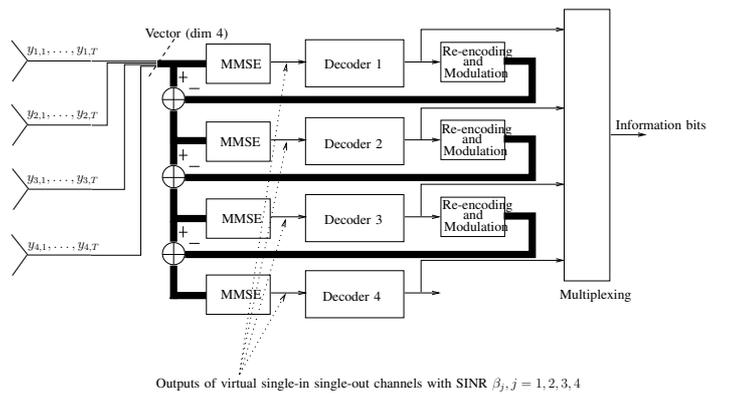Fig. 1.   Transmitter block diagram for a layered V-BLAST scheme with $M = 4$.

Fig. 2.   Receiver block diagram for a layered V-BLAST scheme with $N = M = 4$.

B. Quasi-static block-fading

We have seen that the design of codes for the ergodic MIMO fading channel does not present particular challenges. On the contrary, the STC literature has focused mainly on the quasi-static block-fading channel model. As mentioned before, in this case the STC design presents some unique challenges that make STC a distinct and new field in coding theory, for which new tools have been developed. The quasi-static block-fading model is obtained from (1) by assuming that $\mathbf{H}_t = \mathbf{H}$ for all $t = 1, \ldots, T$, where $\mathbf{H}$ is a random channel matrix, distributed

according to some pdf $p_H(\mathbf{H})$, that stays constant over the whole duration of a codeword. This assumption is justified by considering very slowly-varying fading and coding with some practical latency constraint, that prevents from interleaving codewords over a very large number of independent channel realizations.

From the information theoretic viewpoint, this channel is non-ergodic (or "non-information stable" [44]). For such class of channels, a coding theorem can be stated in terms of the general Verdú and Han capacity formula [44] rather than in terms of an average mutual information, as for the ergodic case. Intuitively, since $\mathbf{H}$ is random but constant over the duration of a codeword and the transmitter does not know its realization, the randomness due to the fading cannot be "averaged out" over the codeword block length. Therefore, for any chosen transmission rate $R > 0$ there might be a non-zero probability that the channel matrix $\mathbf{H}$ does not support rate R. Hence, the error probability may not vanish with $T \to \infty$ no matter how small R is chosen. For such class of channels, a more meaningful performance measure is ϵ-capacity [44], defined as the supremum of all ϵ-achievable rates.[3] From the general theory developed in [44], applied to the quasi-static block-fading MIMO channel, we obtain

$$C_\epsilon(\rho) = \sup\{R : P_{\text{out}}(\rho, R) \le \epsilon\} \tag{8}$$

where $P_{\text{out}}(R)$, referred to as the *information outage probability* [45], [15], is defined by

$$P_{\text{out}}(\rho, R) = \inf_{\boldsymbol{\Sigma}_x \ge 0, \text{tr}(\boldsymbol{\Sigma}_x) \le 1} P\left(\log \det\left(\mathbf{I} + \rho \mathbf{H}\boldsymbol{\Sigma}_x\mathbf{H}^{\mathsf{H}}\right) \le R\right) \tag{9}$$

Simply put, the outage probability $P_{\text{out}}(\rho, R)$ is the probability that the mutual information for a given (fixed) channel realization $\mathbf{H}$ falls below the transmitted rate R. Since $\mathbf{H}$ is random, the mutual information for fixed $\mathbf{H}$ is a random variable. This probability is minimized with respect to the input covariance matrix $\boldsymbol{\Sigma}_x$. It can be easily shown (see for example [10]) that $P_{\text{out}}(\rho, R)$ is the best possible block error probability of any code in the limit of $T \to \infty$. Notice that this does not mean that some specific code with small block length T and rate R cannot achieve a block error probability smaller than $P_{\text{out}}(\rho, R)$ for some ρ. However, when we communicate at rate R we consider the transmission of an arbitrarily large number of information bits. Then, when we look at the probability of message error for arbitrarily large message size, no code can do better than $P_{\text{out}}(R)$.

The computation of $P_{\text{out}}(\rho, R)$ for given channel statistics $p_H(\cdot)$ is generally a hard problem (see for example [46]). Also, the determination of the outage-minimizing input covariance $\boldsymbol{\Sigma}_x$, that is generally different from the ergodic capacity achieving input covariance, seems to be quite hard in general (see for example [21], [24]). Nevertheless, the simple bound

$$\log \det\left(\mathbf{I} + \rho \mathbf{H}\boldsymbol{\Sigma}_x\mathbf{H}^{\mathsf{H}}\right) \le \log \det\left(\mathbf{I} + \rho \mathbf{H}\mathbf{H}^{\mathsf{H}}\right)$$

that holds for all $\boldsymbol{\Sigma}_x$ such that $\text{tr}(\boldsymbol{\Sigma}_x) \le 1$, yields that

$$\widetilde{P}_{\text{out}}(\rho, R) \ge P_{\text{out}}(\rho, R) \ge \widetilde{P}_{\text{out}}(M\rho, R) \tag{10}$$

[3] A rate R is said to be ϵ-achievable if there exists a sequence of codes of size $\exp(TR)$, block length T and error probability $P_e(T)$ such that $\limsup_{T \to \infty} P_e(T) \le \epsilon$ [41].

where

$$\widetilde{P}_{\text{out}}(\rho, R) = P\left(\log \det\left(\mathbf{I} + \frac{\rho}{M}\mathbf{H}\mathbf{H}^{\mathsf{H}}\right) \le R\right) \tag{11}$$

is the information outage probability when we impose uncorrelated inputs with uniform power allocation over all antennas. As a matter of fact, the optimization of $\boldsymbol{\Sigma}_x$ yields negligible outage probability improvement in the region of high SNR. In fact, the DMT characterization of the quasi-static block-fading MIMO channel (see Section III) is obtained by studying the asymptotics of $\widetilde{P}_{\text{out}}(\rho, R)$ for large ρ.

Let's now consider the error performance of codes over the quasi-static block-fading MIMO channel. Consider an STC $\mathcal{C}$, defined by a set of $|\mathcal{C}|$ codeword matrices $\mathbf{X}$. The maximum-likelihood decoder operates according to the minimum distance rule

$$\widehat{\mathbf{X}} = \arg\min_{\mathbf{X} \in \mathcal{C}} \sum_{t=1}^{T} |\mathbf{y}_t - \sqrt{\rho}\mathbf{H}\mathbf{x}_t|^2$$

A simple union bound on the average block error probability, conditioned on the channel realization $\mathbf{H}$, is given by

$$P_e(\rho, \mathbf{H}) \le \frac{1}{|\mathcal{C}|} \sum_{\substack{\mathbf{X}, \mathbf{X}' \in \mathcal{C} \\ \mathbf{X} \ne \mathbf{X}'}} P(\mathbf{X} \to \mathbf{X}'|\mathbf{H}) \tag{12}$$

where the conditional pairwise error probability is given by

$$
\begin{aligned}
P(\mathbf{X} \to \mathbf{X}'|\mathbf{H}) &= Q\left(\sqrt{\frac{\rho}{2}\text{tr}(\mathbf{H}^{\mathsf{H}}\mathbf{H}\boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathsf{H}})}\right) \\
&\le \exp\left(-\frac{\rho}{4}\text{tr}(\mathbf{H}^{\mathsf{H}}\mathbf{H}\boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathsf{H}})\right)
\end{aligned}
\tag{13}
$$

where $Q(\cdot)$ denotes the Gaussian tail function, where we define the *codeword difference matrix* $\boldsymbol{\Delta} = \mathbf{X} - \mathbf{X}'$ and where the last line follows from the Chernoff bound. Unfortunately, by averaging (12) with respect to $\mathbf{H}$ the resulting union bound is generally very loose. A much more precise upper bound is obtained by [47], [37]

$$P_e(\rho) \le \mathbb{E}\left[\min\left\{1, \frac{1}{|\mathcal{C}|} \sum_{\substack{\mathbf{X}, \mathbf{X}' \in \mathcal{C} \\ \mathbf{X} \ne \mathbf{X}'}} \exp\left(-\frac{\rho}{4}\text{tr}(\mathbf{H}^{\mathsf{H}}\mathbf{H}\boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathsf{H}})\right)\right\}\right] \tag{14}$$

The drawback of (14) is that the expectation with respect to $\mathbf{H}$ must be performed by Monte Carlo simulation. However, by looking at the individual average pairwise error probability, some important guidelines for the design of STCs for the quasi-static block-fading channel can be drawn [3]. From the theory of quadratic forms in complex Gaussian random variables, we obtain[4]

$$
\begin{aligned}
\mathbb{E}[P(\mathbf{X} \to \mathbf{X}'|\mathbf{H})] &\le \mathbb{E}\left[\exp\left(-\frac{\rho}{4}\text{tr}(\mathbf{H}^{\mathsf{H}}\mathbf{H}\boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathsf{H}})\right)\right] \\
&= \left(\det\left(\mathbf{I} + \frac{\rho}{4}\boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathsf{H}}\right)\right)^{-N} \\
&\approx \rho^{-\nu N}\left(\frac{1}{4}\left[\prod_{j=1}^{\nu}\lambda_j\right]^{1/\nu}\right)^{-\nu N}
\end{aligned}
\tag{15}
$$

[4] From now on, implicitly we assume the i.i.d. Rayleigh fading channel model, with pdf given by (4).

where $\{\lambda_j : j = 1, \ldots, \nu\}$ are the non-zero eigenvalues of $\mathbf{\Delta}\mathbf{\Delta}^{\mathsf{H}}$, and $\nu = \mathrm{rank}(\mathbf{\Delta})$. The last line reveals that the average pairwise error probability decreases as an inverse power of ρ, that is, in the classical log-log chart ($\log_{10} P_e$ versus SNR in dB), for large ρ it is a straight line with slope $-\nu N$. The slope of error probability is referred to as "diversity". The above bound yields that the code diversity is given by $\nu_{\min} N$, where $\nu_{\min}$ is the minimum rank over all non-zero codeword matrix differences. In the high-SNR region, where typically the performance of a code becomes meaningful, the code diversity becomes the dominating factor for good performance. Hence, maximizing the minimum codeword difference rank can be considered as the first distinctive and most important criterion for STC design. For $T \geq M$, it is clear that we seek "full-rank" codes, that is, codes with $\nu_{\min} = M$. In this case, the code diversity is MN, that corresponds to the maximum diversity order of the channel itself (notice that the number of fading degrees of freedom, equal to the number of independent fading coefficients, is indeed MN). The second criterion that was proposed in the STC literature, also based on (15), is to maximize the harmonic mean of the non-zero eigenvalues, $\left[\prod_{j=1}^{\nu} \lambda_j\right]^{1/\nu}$ for the codeword pairs with minimum rank $\nu_{\min}$. For "full-rank" codes, we have $\prod_{j=1}^{M} \lambda_j = \det(\mathbf{\Delta}\mathbf{\Delta}^{\mathsf{H}})$. Therefore, the relevant design criterion becomes the maximization of the minimum determinant $\det(\mathcal{C}) = \min_{\mathbf{\Delta} \neq 0} \det(\mathbf{\Delta}\mathbf{\Delta}^{\mathsf{H}})$.

We shall see in the following sections that the "rank and determinant" criteria, that were proposed in early STC literature based on the pairwise error probability analysis, indeed plays a fundamental role in the definition of a family of STCs that achieve the optimal DMT. However, these criteria have to be re-interpreted in light of the theory that shall be developed in the next section.

III. DIVERSITY-MULTIPLEXING TRADEOFF OF QUASI-STATIC CHANNELS

The capacity of the ergodic MIMO channel for high-SNR is given by

$$C_{\mathrm{erg}}(\rho) = \min\{M, N\} \log \rho + o(1)$$

On the other hand, the block error probability of codes on the quasi-static block-fading MIMO channel is, at best, $P_e(\rho) \doteq \rho^{-MN}$. [5] The early literature on STC focused on the design of "full-rate" and "full-diversity" codes, where the former means that the number of transmitted independent information symbols per channel use is equal to $\min\{M, N\}$ and the latter means that the error probability slope (in the sense said before) is equal to MN. More recently it was recognized that these two goals can be accomplished simultaneously, so that there is no actual "tradeoff" between full-rate and full-diversity. Constructions of full-rate full-diversity STCs are by now well-known (see for example [48]).

Zheng and Tse in [10] formulated a more meaningful tradeoff in terms of the channel "SNR exponent". Let's fix

the block length T and let $\rho \to \infty$. Let's consider a family of codes $\{\mathcal{C}_\rho\}$, where code $\mathcal{C}_\rho$ operates at SNR ρ, with rate $R(\rho)$ and error probability $P_e(\rho)$. We define the *multiplexing gain* r of the family as the pre-log factor of $R(\rho)$, that is, we let $R(\rho) \doteq r \log \rho$. Also, we define the *diversity gain* d of the family as the slope of $P_e(\rho)$ in the log-log error probability chart, that is, $P_e(\rho) \doteq \rho^{-d}$. Then, the channel SNR exponent is defined as the largest achievable diversity gain $d^\star(r)$ for every $r \geq 0$. This represents the best possible diversity-multiplexing tradeoff (DMT) achievable by any family of STCs

In [10], $d^\star(r)$ is determined for the case of i.i.d. Rayleigh fading. In particular, $d^\star(r)$ is equal to zero for $r \geq \min\{M, N\}$ and for $0 \leq r < \min\{M, N\}$ it is given by the piecewise linear function interpolating the points

$$(r = k, \ d^\star(r) = (M - k)(N - k)), \qquad (16)$$

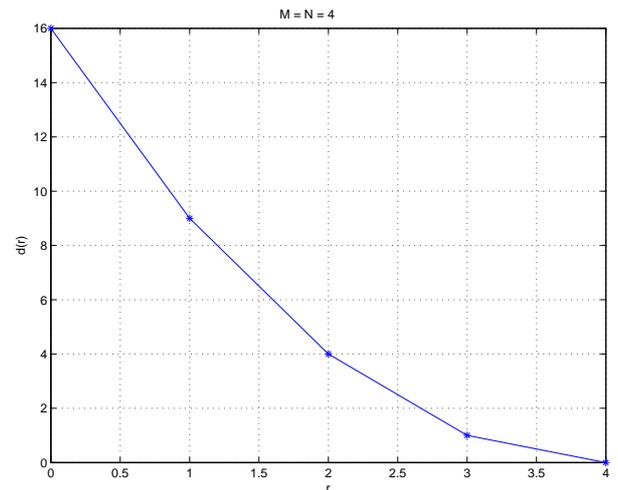for $k = 0, 1, \ldots, \min\{M, N\}$. Fig. 3 shows $d^\star(r)$ versus r for $M = N = 4$.

Fig. 3. DMT $d^\star(r)$ in the case of $M = N = 4$.

We notice here the following interesting parallel: in the classical theory of error exponents, we are interested in the exponential rate of decay of P_e versus the code block length $T \to \infty$ for fixed channel parameter (the SNR ρ in our case) and for codes of fixed rate (bit per symbol, i.e., unit of block length). In the quasi-static block-fading channel, we already know that the classical error exponent is zero (since the Shannon capacity itself is zero). The DMT formulation exchanges T for $\log \rho$, and investigates the rate of decay of P_e versus $\log \rho \to \infty$ for fixed T and multiplexing gain r (bit per unit of $\log \rho$, proportional to "bit per dB"). It turns out that, while for classical error exponents only upper (sphere-packing) and lower (random coding) bounds are generally known[6], in the case of DMT, $d^\star(r)$ is fully determined. Zheng and Tse proved the converse by analysing the behavior of $\widetilde{P}_{\mathrm{out}}(\rho, r \log \rho)$ for large ρ and by using Fano inequality,

---

[5]**Notation:** $\doteq$ denotes exponential equality, i.e., $f(z) \doteq z^b$ means that $\lim_{z \to \infty} \frac{\log f(z)}{\log z} = b$, $\dot{\geq}$ and $\dot{\leq}$ are used similarly.

[6]They may coincide for certain ranges of rate.

and the achievability part by considering a random Gaussian coding ensemble. Their upper and lower bound coincide for all r only under the additional block length condition $T \geq M + N - 1$. This condition is needed because of the weakness of the random coding approach. More recently [49] provided a general explicit construction of codes achieving the optimal DMT for all block length $T \geq M$.

The above results show that the performance of any "good" code over the quasi-static block-fading MIMO channel in the regime of high SNR is essentially determined by the *outage set*, that is, the set of channel matrices

$$\mathcal{O} = \left\{ \mathbf{H} \ : \ \log\left(1 + \frac{\rho}{M}\mathbf{H}\mathbf{H}^{\mathsf{H}}\right) \leq R \right\}. \quad (17)$$

The particular form of $d^\star(r)$ obtained in (16) is obtained for i.i.d. Rayleigh fading. [7] Zheng and Tse posed the problem of finding coding constructions that achieve the optimal DMT and admit algorithms for efficient encoding and decoding other than exhaustive codebook lookup table and brute-force minimum distance calculation, as in the case of random codes. At that time, the only known construction was the famous Alamouti code for the $M = 2, N = 1$ channel [51] (see Section IV-A). Zheng and Tse also showed that Orthogonal Space-Time Block Codes (OSTBC) [52] in general, achieve maximum diversity but not maximum multiplexing gain whereas the simple "uncoded" V-BLAST (see Section IV-B) achieves maximum multiplexing gain but not maximum diversity.

The code construction problem posed in [10] sparked considerable research effort. The first STC ensemble that achieves the optimal DMT for all M, N while admitting efficient encoding and decoding is the LAST code construction provided in [53] (see Section IV-E). The first explicit (non-random ensemble) construction of STCs achieving the optimal DMT for all M, N was then provided in [49], and it is based on Cyclic Division Algebra (CDA) codes (see Section IV-D).

In general, the outage probability can be defined as $\widetilde{P}_{\text{out}}(\rho, R) = P(\mathcal{O})$ and the corresponding optimal DMT $d^\star(r)$ depends on the channel statistics $p_H(\cdot)$. A code family $\{\mathcal{C}_\rho\}$ may achieve the optimal DMT for some channel statistics but may not be optimal for some others. A stronger version of the DMT formulation is given in [54], [18] in terms of the *approximate universality* condition. Intuitively, the operational significance of the information outage probability relies on the existence of "universal codes" of rate R and sufficiently large block length such that, for every $\epsilon > 0$, their probability of error is smaller than ϵ for all channel realizations $\mathbf{H} \notin \mathcal{O}$ while it can be larger than $1 - \epsilon$ for all channel realizations $\mathbf{H} \in \mathcal{O}$.

While universal codes can be found only in the limit of large block length T (e.g., in the ensemble of random Gaussian codes), for high-SNR and fixed T it makes sense to consider the laxer approximate universality condition given as follows. A code family $\{\mathcal{C}_\rho\}$ is approximately universal if there exists $\alpha > 0$ independent of ρ such that $P_e(\rho) \ \dot{\leq} \ \exp(-\alpha\rho)$ for all $\mathbf{H} \notin \mathcal{O}$.

[7] For Gaussian $\mathbf{H}$ with arbitrary mean and arbirary full-rank correlation, it can be shown that $d^\star(r)$ takes on the same form [50].

In [54] (see also [18]), necessary and sufficient conditions for approximate universality of STCs are given in terms of the eigenvalues of the codeword matrix differences $\Delta\Delta^{\mathsf{H}}$. Remarkably, the CDA codes constructed in [49] are also shown to satisfy the non-vanishing determinant (NVD) property (see Section IV-D). This implies the conditions given in [54] for approximate universality. It follows that the CDA codes of [49] are also approximately universal.

IV. LINEAR-DISPERSION AND LATTICE SPACE-TIME CODES

A general STC setting is that of the Linear-Dispersion (LD) codes, as introduced in [55], in which each STC codeword matrix $\mathbf{X}$ is defined to be of the form

$$\mathbf{X} = \sum_{q=1}^{Q}(\alpha_q\mathbf{A}_q + \imath\beta_q\mathbf{B}_q), \quad (18)$$

where $\{\mathbf{A}_q, \mathbf{B}_q\}$ are arbitrary space-time "spreading" matrices in $\mathbb{C}^{M \times T}$ and the $\{\alpha_q, \beta_q\}$ are real information symbols such that the corresponding complex symbol $s_q = \alpha_q + \imath\beta_q$ belongs to some discrete and finite complex set $\mathcal{A}$ referred to as the "information symbol constellation". The coding rate of a LD code defined as above is given by $R = \frac{Q}{T}\log|\mathcal{A}|$, since Q independently chosen symbols from $\mathcal{A}$ are transmitted in T channel uses.

A more convenient description is obtained by considering the vector equivalent $\mathbf{x}$ of $\mathbf{X}$, where $\mathbf{x}$ is the image of the real information vector $\mathbf{s} \in \mathbb{R}^{2Q}$ with components given by the α_q's and β_q's symbols. The corresponding mapping $\mathbf{s} \mapsto \mathbf{x}$ can be described by a generator matrix $\mathbf{G} \in \mathbb{R}^{2MT \times Q}$ such that

$$\mathbf{x} = \mathbf{G}\mathbf{s}. \quad (19)$$

Furthermore, the channel model (1) can be rewritten in the real-vectorized form,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \quad (20)$$

by column stacking of the real and imaginary components of the respective matrices. In (20) we have $\mathbf{y} \in \mathbb{R}^{2NT}$, $\mathbf{x} \in \mathbb{R}^{2MT}$ and with some abuse of notation, we redefine $\mathbf{H}$ to be the $2NT \times 2MT$ real equivalent channel matrix (including the SNR factor $\sqrt{\rho}$ as part of the channel for ease of notation). For

$$\mathbf{x} = \left[\text{Re}\{\mathbf{x}_1\}^{\mathsf{T}}, \text{Im}\{\mathbf{x}_1\}^{\mathsf{T}}, \ldots, \text{Re}\{\mathbf{x}_T\}^{\mathsf{T}}, \text{Im}\{\mathbf{x}_T\}^{\mathsf{T}}\right]^{\mathsf{T}},$$

and with $\mathbf{y}, \mathbf{w}$ defined similarly, then $\mathbf{H}$ takes on the block-diagonal form [53]

$$\mathbf{H} = \sqrt{\rho}\,\text{diag}\left(\begin{bmatrix} \text{Re}\{\mathbf{H}_1\} & -\text{Im}\{\mathbf{H}_1\} \\ \text{Im}\{\mathbf{H}_1\} & \text{Re}\{\mathbf{H}_1\} \end{bmatrix}, \ldots, \right.$$
$$\left. \ldots, \begin{bmatrix} \text{Re}\{\mathbf{H}_T\} & -\text{Im}\{\mathbf{H}_T\} \\ \text{Im}\{\mathbf{H}_T\} & \text{Re}\{\mathbf{H}_T\} \end{bmatrix}\right).$$

Typically, $\mathcal{A}$ is a QAM constellation, that is, the *real* information vectors $\mathbf{s}$ are given by the intersection of a scaled and translated version of the integer lattice $\mathbb{Z}^{2Q}$ with a hypercube *boundary* region. We denote by $\mathcal{U}$ the set of all information vectors $\mathbf{s}$. It follows that the image $\mathbf{G}\mathcal{U}$ of the encoding

mapping (19) is a subset of an integer lattice in $\mathbb{R}^{2MT}$ with generator matrix $\mathbf{G}$. The relation between LD codes and lattices shall be fully exploited by the LAST code construction defined later on. At this point, it is worthwhile to notice that the inherent lattice structure can be exploited both for efficient encoding via (19) and for efficient ML decoding. In fact, the ML decoder operates according to the decision rule

$$\widehat{\mathbf{s}} = \arg\min_{\mathbf{s}\in\mathcal{U}} \; |\mathbf{y} - \mathbf{HGs}|^2 \qquad (21)$$

Provided that $\mathcal{U}$ is defined by a simple boundary region, as in the case of the hypercube (QAM information symbol constellation), then the minimum distance search in (21) can be implemented efficiently by using ideas from the literature on *closest lattice point search*, also known as "Sphere-Decoding" (SD) [56], [57], [58], [59]. This significantly decreases the average decoding complexity with respect to an exhaustive search. Furthermore, SD has been recently recognized to be an instance of the more general *branch and bound* tree search technique [60]. As such, better tree search algorithms such as sequential decoding (stack or Fano algorithms) can be used with even more computational savings [60].

The remainder of this Section discusses some notable instances of LD codes, the related class of LAST codes and finally an approach that to combines the best of both constructions.

A. The Alamouti code: achieving full diversity

The Alamouti code [51] consists of a set of mutually orthogonal matrices, taking the form:

$$\mathcal{C}_{\text{Alam}} = \left\{ \mathbf{X} = \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix} \;,\; s_1, s_2 \in \mathcal{A} \right\}. \qquad (22)$$

In the real vectorized notation (19), letting $\mathbf{s} = (\alpha_1, \beta_1, \alpha_2, \beta_2)^{\mathsf{T}}$ the matrix $\mathbf{G}$ corresponding to (22) is given by

$$\mathbf{G}_{\text{Alam}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 \end{bmatrix}.$$

For QAM information symbol constellation of size $|\mathcal{A}| = \rho^r$, with $0 \le r \le 1$, the Alamouti code has rate $R = r \log \rho$ and probability of error $P_e(\rho) \doteq \rho^{-2(1-r)}$ over the $M = 2, N = 1$ channel [10]. Hence, the Alamouti code is DMT optimal for this case. Unfortunately, for any $N > 1$ it fails to achieve the optimal DMT. The intuition behind this shortcoming is that the Alamouti code transmits *only* one QAM symbol per channel use, while the channel with $N > 1$ supports up to $\min\{2, N\} = 2$ degrees of freedom, that is, up to 2 independent symbols can be transmitted on each channel use.

The Alamouti code is the simplest and best known instance of the family of OSTBC discovered in [52] (see also [61], [62] for additional constructions and properties) that are easily

shown to support non-zero diversity gain only for multiplexing gains $r \le 1$. The main advantage of OSTBCs is that they allow for very simple decoding, where the MIMO channel is turned into a set of non-interfering parallel channels by simple signal combining operations, so that ML decoding reduces to the independent detection of the information symbols and not even the SD is needed in order to implement (21), as explained in [51], [52].

B. Uncoded V-BLAST: achieving full rate

To improve upon the rate limitation of OSTBCs, full-rate codes were constructed. Perhaps the simplest example is uncoded V-BLAST, that corresponds to the encoder of Fig. 1 with trivial components codes with block length $T = 1$. In this case, the encoder matrix $\mathbf{G}$ is equal to the $M \times M$ identity. Assuming $M \le N$, we use independently chosen QAM symbols from a constellation $\mathcal{A}$ of size $|\mathcal{A}| = \frac{r}{M}\log\rho$ to drive the M antennas. At the receiver, the ML (minimum distance) decoder (21) is implemented (e.g., by SD).[8] It follows that for transmission rate equal to $R = r \log \rho$ the resulting error probability can be shown to decay like $P_e \doteq \rho^{-N(1-r/M)}$. Notice that for the squared channel case $M = N$ this SNR exponent coincides with the optimal DMT $d^\star(r)$ in the segment $M - 1 \le r \le M$ of high multiplexing gain. On the contrary, for $N > M$ uncoded V-BLAST is strictly suboptimal. It is interesting to notice that the exponent $N(1 - r/M)$ for $N \ge M$ is the best possible exponent achieved by "space-only" codes, that is, by coding across the transmit antennas without exploiting the time dimension (block length $T = 1$). Furthermore, it can also be shown that by coding across time without exploiting the antenna dimension (e.g., the scheme of Fig. 1 with non-trivial component codes), the above SNR exponent of error probability cannot be improved. This shows somehow intuitively that coding across antennas (space) and time is necessary in order to achieve the whole DMT curve.

C. TAST codes: achieving full rate and full diversity

Before the DMT optimality problem was posed, most literature on LD codes focused on the careful design of the encoding matrix $\mathbf{G}$ in (19) in order to obtain full-rate full-diversity codes. Recall that full-rate here means codes that send $\min\{M, N\}$ independently chosen symbols from the underlying information set $\mathcal{A}$ per channel use, and full-diversity indicates codes with error probability slope MN, that is, with $\det(\mathcal{C}) \ne 0$. Among the first systematic constructions of such codes, the Threaded Algebraic Space-Time (TAST) code construction of [48] is particularly noteworthy for its generality, simplicity and performances.

The TAST construction makes use of codes designed for the Single-Input Single-Output (SISO) channel as component codes. Each SISO component codeword is assigned to a *thread* in the ST codeword, the concept of which is formalized as follows. A *layer* is defined as a section of the ST codeword

---

[8]Interestingly, uncoded V-BLAST has been proposed in the literature in conjunction with the nulling and canceling strategy of Fig. 2 [42]. This strategy, that can achieve the ergodic capacity, is unfortunately very suboptimal under quasi-static block-fading as shown for example in [10], [18].

matrix having the property that each symbol interval within the section is allocated to at most one antenna [63]. This property ensures that all spatial interference experienced by the layer comes from outside the layer. For example, in the coded V-BLAST scheme of Fig. 1 each layer is associated to one particular antenna and spans all time slots $t = 1, \ldots, T$. In another variant called D-BLAST [64], [10], [18], the layers are stacked along the diagonals in the ST codeword matrix. A *thread* is defined as a layer which has full temporal and spatial span, i.e., we are guaranteed to find a component of the thread along any particular antenna $j = 1, \ldots, M$ or time slot $t = 1, \ldots, T$. The work in [48] considers the design of $M \times T$ STCs that are composed of $L \le M$ threads. The design problem now reduces to the design of SISO component codes for each thread such that the resultant STC is endowed with full diversity, that is, its minimum determinant is non-zero.

This goal is achieved in two distinct design steps. First, each thread is guaranteed to achieve full diversity when it is transmitted in isolation, i.e., by zeroing all other threads. This is obtained by generating the vector of symbols in each thread by rotating an M-dimensional information symbol vector by an $M \times M$ rotation matrix $\mathbf{R}$, with elements from a rational number field extension, specifically designed to provide non-zero product distance [65], [66]. Then, in order to guarantee that the set of matrices formed by stacking all threads together forms a code $\mathcal{C}$ with $\det(\mathcal{C}) \ne 0$, the threads are scaled by a set of diophantine numbers (properly chosen algebraic numbers), or transcendental numbers. This provides the required "algebraic separation" of the threads such that the non-zero determinant is enforced.

However, it turns out that for the choice of rotation matrices and diophantine scaling coefficients chosen in [48] (and in many other subsequent works), when the size of the underlying information symbol constellation $\mathcal{A}$ increases the minimum determinant approaches zero. This is due to the fact that the approximation of certain real numbers by rationals obtained by the diophantine numbers progressively gets better and better. This fact prevents the original TAST construction from achieving the optimal DMT, and also shows that, generally speaking, constructing a DMT-optimal code family is more difficult than constructing a full-rate full-diversity code operating at given rate R.

D. STCs from Cyclic Division Algebras

The first construction of space-time codes from Cyclic Division Algebras (CDAs) was proposed by Sethuraman and Sundar Rajan [67] and independently, shortly after, by Belfiore and Rekaya [68]. The multiplicative invertibility of all the elements in a division algebra together with the additive closure, renders the algebra as a natural source of full-diversity STCs. The construction is based on the matrix representation of the elements of the division algebra. The particular class of CDAs has a simple structure, (see [69]). Codes derived for CDAs include the 2×2 Golden code. The 2×2 Yao-Wornell and Dayal-Varanasi codes are also intimately related to the family of CDA codes.

Besides their reduced signaling and decoding complexity, CDA codes meet several information theoretic criteria. For example as previously mentioned, it was shown in [49], [70] that CDA-based ST codes that also possess the non-vanishing determinant (NVD-[71]) property, achieve the optimal DMT for all (M, N). Based on existing constructions for restricted values of M, (see [71], [68], [72]), the authors in [49], [70] provide constructions of such DMT optimal codes for all M, all $T \ge M$. We now present an overview of this construction and limit our exposition to the minimum delay case of $T = M = n$ for the sake of space limitation.

The starting point for the construction of such an $n \times n$ STC is a cyclic division algebra D, with center $\mathbb{F}$ and maximal field $\mathbb{L}$, where $\mathbb{L}/\mathbb{F}$ is an n-degree cyclic Galois extension. Let σ be the generator of the Galois group $\mathrm{Gal}(\mathbb{L}/\mathbb{F})$, and let $\{1, z, \ldots, z^{n-1}\}$ be the basis of D, as a right vector-space over $\mathbb{L}$. z is an indeterminate satisfying

$$\ell z = z\sigma(\ell) \quad \forall \; \ell \in \mathbb{L} \quad \text{and} \quad z^n = \gamma,$$

for some non-norm element $\gamma \in \mathbb{F} \backslash \{0\}$. More specifically, γ has the property that the smallest positive integer t for which γ^t is the relative norm $N_{\mathbb{L}/\mathbb{F}}(u)$ of some element u in $\mathbb{L} \backslash \{0\}$, is n. The corresponding CDA $D(\mathbb{L}/\mathbb{F}, \sigma, \gamma)$ consists of all elements of the following form

$$u = \sum_{i=0}^{n-1} z^i \ell_i, \quad \ell_i \in \mathbb{L}, \tag{23}$$

whose *left-regular* matrix representation is given as

$$\begin{bmatrix} \ell_0 & \gamma\sigma(\ell_{n-1}) & \gamma\sigma^2(\ell_{n-2}) & \ldots & \gamma\sigma^{n-1}(\ell_1) \\ \ell_1 & \sigma(\ell_0) & \gamma\sigma^2(\ell_{n-1}) & \ldots & \gamma\sigma^{n-1}(\ell_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \ell_{n-1} & \sigma(\ell_{n-2}) & \sigma^2(\ell_{n-3}) & \ldots & \sigma^{n-1}(\ell_0) \end{bmatrix}. \tag{24}$$

Despite the fact that the elements of the above matrix belong to the maximal subfield $\mathbb{L}$, the CDA structure ensures that the determinant of the matrix is an element of the center $\mathbb{F}$.

In order to construct an STC $\mathcal{C}$ with full rate and the NVD property, one can choose [68]

- $\gamma \in \mathcal{O}_{\mathbb{F}}^*$,
- $\{\beta_1, \ldots, \beta_n\}$ to be an integral basis for $\mathbb{L}/\mathbb{F}$,
- restrict the ℓ_i in (24) to be of the form

$$\left\{ \sum_i a_i \beta_i \mid a_i \in \mathcal{A} \right\}.$$

where $\mathcal{A}$ is a QAM constellation.

The above restrictions result in the determinant being an element of $\mathbb{Z}[\imath]$. The NVD property of the resulting STC follows from the closure under subtraction of the division algebra and from the fact that $\mathbb{Z}[\imath]$ is discrete. Therefore, $\det(\mathcal{C})$ is bounded away from zero by some constant.

CDA codes achieve optimality in the high-SNR regime, but they may not perform well in the range of moderate to low SNR. A DMT optimal constructions that also performs well in low SNR will be explored later.

E. Lattice Space-Time Codes

We have already pointed out that the encoding mapping (19) yields a $2Q$ dimensional integer lattice in $\mathbb{R}^{2MT}$ when the information vector $\mathbf{s}$ takes on values in the whole $\mathbb{Z}^{2Q}$.[9] In particular, we are interested in *full-dimensional* LAST codes [53], where $Q = 2MT$ and $\mathbf{G}$ has rank $2MT$.

This approach consists of finding a lattice $\Lambda = \{\boldsymbol{\lambda} = \mathbf{Gs} : \mathbf{s} \in \mathbb{Z}^{2MT}\}$, a suitable Jordan-measurable shaping region $\mathcal{R} \subset \mathbb{R}^{2MT}$ and a translation vector $\mathbf{u}_0$, such that the space-time vectorized codewords are given as the set of points $\mathcal{C} = \{\Lambda + \mathbf{u}_0\} \cup \mathcal{R}$. The translation vector $\mathbf{u}_0$ is chosen such that the resulting code is centered around the origin, i.e., $\frac{1}{|\mathcal{C}|} \sum_{\mathbf{x} \in \mathcal{C}} \mathbf{x} = \mathbf{0}$. The shaping region $\mathcal{R}$ is designed in order to obtain the desired rate (number of points in $\mathcal{C}$) and in order to achieve a large *shaping gain*. Notice that this is not possible if we insists on the linear encoding mapping given by (19). In fact, with the general LD encoding, the code $\mathcal{C}$ is given by $\mathbf{G}\mathcal{U}$, the image of the information set $\mathcal{U}$ under the linear map $\mathbf{G}$. If this is contained in a hypercube, then $\mathcal{G}$ is contained in some diamond-shaped paralleletope that may be far from the best (spherical) shaping.

Beyond the above general construction, we focus here on the class of *nested* LAST codes [53], where $\mathcal{R}$ is the fundamental Voronoi cell of a sublattice $\Lambda_s \subset \Lambda$, referred to as the "shaping lattice", and $\mathbf{u}_0$ is a pseudo-random *dithering* vector uniformly distributed over $\mathcal{R}$ whose realization is known to the receiver.[10] The resulting encoding scheme is represented in Fig. 4.
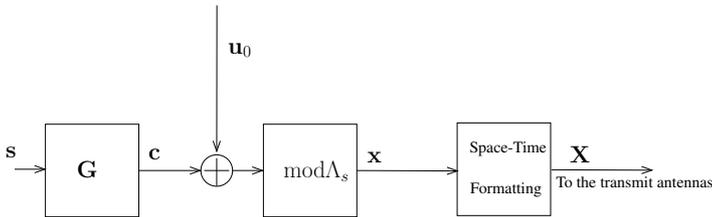

Fig. 4.   Transmitter block diagram of the modulo-Λ scheme.

We define the lattice minimum distance quantization function

$$Q_\Lambda(\mathbf{z}) = \arg \min_{\boldsymbol{\lambda} \in \Lambda} |\mathbf{z} - \boldsymbol{\lambda}|^2 \qquad (25)$$

and the modulo-Λ function

$$\mathbf{z} \mod \Lambda = \mathbf{z} - Q_\Lambda(\mathbf{z}) \qquad (26)$$

The LAST codeword in the scheme of Fig. 4 is given by $\mathbf{x} = [\mathbf{Gs} + \mathbf{u}_0] \mod \Lambda_s$. By definition, $\mathbf{x}$ is uniformly distributed on $\mathcal{R}$. Furthermore, $\mathbf{x}$ and $\mathbf{s}$ are statistically independent if not conditioned on $\mathbf{u}_0$. It follows also that the transmitted energy per channel use is related to the second-moment of the shaping

---

[9]In this section, without loss of generality, we assume no translation and scaling of the integer grid.

[10]Pseudo-random dithering known at the receiver is a common procedure in digital communications. For example, pseudo-random direct-sequence spreading is currently used in CDMA and information bit scrambling using pseudo-random sequences is typically used to enforce the correct signal power spectral density and/or for privacy purposes. Hence, the assumption made here that $\mathbf{u}_0$ is known to the receiver is not particularly restrictive in practice.

---

lattice Λ_s by

$$\rho = \frac{1}{T} \int_{\mathcal{R}} |\mathbf{z}|^2 d\mathbf{z}$$

It is clear that any two points $\boldsymbol{\lambda} \in \Lambda$ in the same coset of the partition Λ / Λ_s are completely equivalent for the above encoder, in the sense that they are mapped in the same transmitted point $\mathbf{x}$. Hence, the modulo-Λ encoder of Fig. 4 maps effectively the information vectors into the *cosets* of Λ_s in Λ.

Enumerating the points in $\mathcal{C}$ (even in the absence of the random dithering) is usually difficult, given the complicated polyhedral shape of $\mathcal{R}$. A much simpler but suboptimal decoder is provided by the so-called Lattice Decoding rule, that consists of finding the point $\mathbf{H}\widehat{\boldsymbol{\lambda}}$ of the infinite lattice $\mathbf{H}\Lambda$ with minimum distance from the received point $\mathbf{y}$. Then, since the information message is encoded in the cosets of the partition Λ / Λ_s, the index of the coset that contains $\widehat{\boldsymbol{\lambda}}$ is output as the decoded message. Unfortunately, such a direct but "naive" application of Lattice Decoding is shown in [53] to fall short of achieving the DMT optimality.

Erez and Zamir [73] showed that nested lattice codes, constructed according to the modulo-Λ scheme of Fig. 4, can indeed achieve the channel capacity of the single-input single-output *unfaded* AWGN channel under lattice decoding provided that a linear MMSE estimation stage is used at the receiver, and the lattice decoder is applied to the output of the MMSE estimator. In [53] this idea is extended to the fading MIMO channel. It turns out that the linear MMSE estimation stage that works for the SISO channel must be replaced by a MMSE decision feedback equalizer (MMSE-DFE) in the MIMO case. The resulting modulo-Λ MMSE-DFE scheme is shown in Fig. 5.
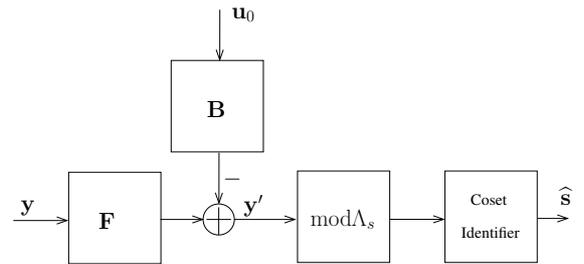

Fig. 5.   Transmitter block diagram of the modulo-Λ scheme..

The receivers makes use of an MMSE-DFE pre-processing stage on the received signal. Let $\mathbf{F}$ and $\mathbf{B}$ denote the forward and feedback MMSE-DFE matrix filters, given by [53]. The output of the MMSE-DFE preprocessing stage is given by

$$\mathbf{y}' = \mathbf{Fy} - \mathbf{Bu}_0 \qquad (27)$$

Notice that the dithering vector is removed through the MMSE-DFE feedback metrix $\mathbf{B}$. Due to the mod-Λ encoding at the transmitter, it can be shown that $\mathbf{y}'$ can be rewritten in the form [53]

$$\mathbf{y}' = \mathbf{Bc} + \mathbf{e}, \qquad (28)$$

where $\mathbf{c} = \mathbf{HGs} + \boldsymbol{\lambda}$ for some $\boldsymbol{\lambda} \in \Lambda_s$ and where $\mathbf{e}$ is a noise plus residual interference that is generally not Gaussian but

is uncorrelated and white, i.e., $\mathbb{E}[\mathbf{e}\mathbf{e}^{\mathsf{T}}]$ is proportional to the identity matrix. Since information was encoded into cosets of Λ_s in Λ, the transformation of $\mathbf{G}\mathbf{s}$ into $\mathbf{c}'$ involves no loss of information. The identification of the coset $\mathbf{c} + \Lambda_s$ is done as follows. The decoder first finds

$$\widehat{\mathbf{z}} = \arg \min_{\mathbf{z} \in \mathbb{Z}^{2MT}} \left| \mathbf{y}' - \mathbf{B}\mathbf{G}\mathbf{z} \right|^2, \tag{29}$$

Since the above search is applied to the whole (infinite) lattice, efficient closest lattice-point search methods (SD, sequential decoding, [60]) can be applied without any complicated "boundary control".[11] Finally, the decoded information message $\widehat{\mathbf{s}}$ is retrieved by finding the coset of Λ_s in Λ that contains $\mathbf{G}\widehat{\mathbf{z}}$. This can be obtained by computing $\mathbf{G}\widehat{\mathbf{z}} \mod \Lambda_s$.

The main result of [53] is the proof that there exist families of nested LAST codes with fixed block length $T \geq M+N-1$ that achieve the optimal DMT for all $0 \leq r \leq \min\{M, N\}$ under the above modulo-Λ scheme. These codes are based on the ensemble of random modulo-prime lattices studied in [74] and more recently in [75], that are very easy to generate. In [53] it is shown that lattices generated at random from the ensemble yield good codes with high probability with little effort to construct them. Furthermore, this good performance is achieved with greatly reduced decoding complexity thanks to the modulo-Λ scheme of Figs. 4 and 5. Nevertheless, finding the *best* LAST codes for a given N, M and T requires some carefully crafted construction or some criterion for searching the lattice ensemble. One of such constructions is examined in the following.

F. Merging of two different methodologies

There exist three important ideas underlying the construction of good STCs, viz., approximate universality, energy efficient signaling and, at the decoder side, low-complexity Lattice Decoding. A synergy of the mathematical structure of CDA codes and the LAST framework is expected to yield schemes that are both approximately universal (high SNR), energy efficient (lower SNR) and that can be decoded efficiently using MMSE-DFE Lattice Decoding. Moreover, these approaches are complimentary and do not compromise on each other, as we will see in the sequel. Notice that a scheme that is both energy efficient and endowed with NVD, essentially packs more codewords in a given volume without compromising on the minimum determinant of the scheme. The following discussion presents an overview of two recent schemes that have strived to exploit this synergy to construct STCs that have excellent performances.

1) Perfect space time codes: Perfect space-time codes were first introduced in [76] as space-time codes that have full rate, full diversity, non-vanishing determinant, uniform average transmitted energy per antenna and cubic shaping, i.e., the points $\mathbf{x} \in \mathcal{C}$ form a regular finite grid contained in an hypercube. The above traits endow perfect codes with near-optimal error performance. The ability of the codes to have a uniform average transmitted energy per antenna and across

time slots and their cubic shaping follow from the fact that the generator matrix in (19) is unitary. Therefore, by choosing the information symbol constellation $\mathcal{A}$ to be a QAM constellation, the code $\mathcal{C}$ is contained in a rotated hypercube in $2M^2$ real dimensions.[12] Such a unitary transformation preserves the energy efficiency of the code.

Perfect codes were constructed in [76] for $n = 2, 3, 4, 6$ and in [77] for all integer values of n. In practice, codes that achieve the defining conditions can be derived from CDA codes with a proper unit-magnitude non-norm element and with an information set carved out of an orthogonal lattice. The construction for odd n is given by the following recipe, while the generalization for all n follows easily [77] and is omitted for brevity.

Perfect code construction: Let $\mathbf{G}$ be a unitary, circulant $n \times n$ lattice generator matrix with first row elements given by

$$G(0,j) = \frac{1}{p} w^\lambda \prod_{k=0}^{\frac{p-1}{2}-1} \left((1-w^{r^k}) \sum_{k=1}^{\frac{p-1}{n}} (-1)^{kn+j} (1-w^{r^{kn+j}}) \right)$$

where $j \in [0, n-1]$, $p \equiv 1 \pmod{n}$ is a prime, $w = e^{\frac{2\pi i}{p}}$, r is a primitive element of the multiplicative group $\mathbb{Z}_p^*$ of integers modulo p without zero, and $\lambda(r-1) \equiv 1 \pmod{p}$. Set

$$\boldsymbol{\Gamma} = \begin{bmatrix} 0 & 0 & \cdots & 0 & \frac{\pi_1}{\pi_1^*} \\ 1 & 0 & \cdots & 0 & 0 \\ & & \vdots & & \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \tag{30}$$

where $\pi_1 \pi_1^* = q \equiv 1 \pmod 4$, q a prime that generates $\mathbb{Z}_p^*$. For $\{\underline{f}_j\}_{j=0}^{n-1}$ being n independent QAM n-tuples, it follows that the $n \times n$ matrices given by

$$\mathbf{X} = \sum_{j=0}^{n-1} \boldsymbol{\Gamma}^j \left(\mathrm{diag}\left(\underline{f}_j \mathbf{G}\right) \right)$$

yield perfect codes.

2) Structured LAST codes: The choice of the lattices Λ and Λ_s greatly influences the performance (in terms of error probability) of the nested LAST codes. In a recent work, the authors in [78] constructed structured-LAST (S-LAST) codes using rotated versions of dense lattices in $\mathbb{R}^{2n^2}$ space as the coding lattice Λ. Specifically, they choose

$$\mathbf{G} = \mathbf{G}_p \mathbf{G}_{\Lambda'},$$

where $\mathbf{G}_{\Lambda'}$ and $\mathbf{G}_p$ are respectively the lattice generator matrices of the $n \times n$ perfect code, and an integral lattice in $2n^2$-dimensions that is good for packing [79]. Since $\mathbf{G}$ is not necessarily unitary, carving a set of points from Λ using a conventional linear map is not energy efficient [78]. This leads naturally to the use of a modulo-Λ scheme. A simple way to obtain the shaping lattice consists of using a self-similar nested sublattice of Λ, i.e., letting $\Lambda_s = Q\Lambda$, with $Q \in \mathbb{Z}^+$. This ensures that the resulting code is a rotated and scaled version of a dense lattice packing, resulting in excellent performance at finite SNRs. Due to the fact that Λ' is an integral lattice, the S-LAST codes also possess the CDA structure and therefore are

---

[11]Following [58], [60], we refer to the boundary control as the problem of checking that a vector in the infinite lattice belongs to the code $\mathcal{C}$. This may be easy or very complicated, depending on the shaping region.

[12]As for CDA codes, we restrict the treatment here to the case of $M = T = n$.

approximately universal (under minimum distance ML decoding). As a matter of fact, they provide very good performance also under the much simpler MMSE-DFE Lattice Decoding scheme of Fig. 5. Fig. 6 shows a comparison between the Golden code and the Gosset S-LAST code (obtained by taking Λ' to be the Gosset E_8 lattice in 8 dimensions) under ML and MMSE-DFE Lattice Decoding. Notice that the performance of MMSE-DFE Lattice Decoding is very close to that of ML decoding, while achieving significant reductions in complexity [60]. Gains due to shaping become more pronounced at higher data rates. A corresponding plot comparing 3×3 perfect codes with an S-LAST code constructed taking Λ' equal to the Λ_{18} lattice under MMSE-DFE lattice decoding is shown in Fig. 7.
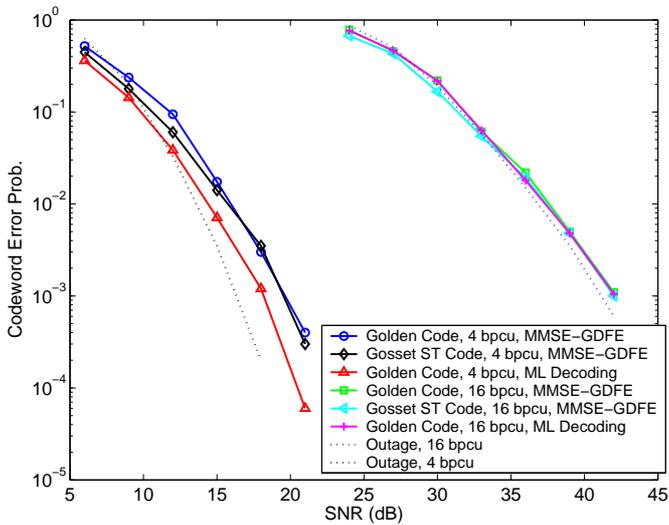

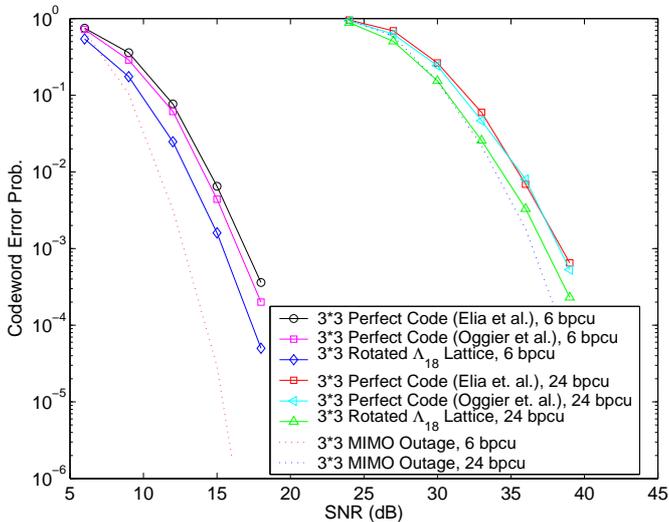Fig. 6.   Golden and Gosset S-LAST codes under ML and MMSE-DFE decoding.

Fig. 7.   Performance of 3×3 Perfect and Λ_{18} S-LAST codes.

V. RECENT DEVELOPMENTS

A. Space-Time Coding for ARQ systems

El Gamal, Caire and Damen in [80] considered signaling across the same quasi-static Rayleigh fading channel using an L-round Automatic Retransmission reQuest (ARQ) protocol. The protocol assumes the presence of a noiseless feedback channel, which is capable of conveying one bit of information (ACK/NACK message) per T uses of the feedback channel. The channel may change independently at every new transmission (this is representative of a system where the ARQ blocks are transmitted at different subcarriers or well separated in time by some time-frequency hopping pattern), or stay fixed over all L protocol rounds. Two schemes are considered. In the basic scheme, the input power constraint is imposed on a per-block basis: the transmitter cannot modulate its transmit power as a function of the the received ACK/NACK messages. In the "power control" scheme, the input power constraint is imposed on average, over a long sequence of protocol rounds. In this case, the transmitter is allowed to boost the transmit energy in order to decrease the probability of error in the rare events where the channel misbehaves.

The optimal SNR exponent in the various cases discussed above is determined in [80] and it is termed the Diversity-Multiplexing-Delay Tradeoff (DMDT) since the ARQ protocol introduces an additional dimension to the problem, namely, the maximum delay L up to which a decision on the current information message must be made.

For channels with independent Rayleigh fading, where the channel remains constant throughout all the ARQ rounds (the long-term static fading case), the SNR exponent of the MIMO-ARQ channel is given by

$$d_{1s}^*(r) = d^\star\left(\frac{r}{L}\right), \tag{31}$$

for $0 \le r < \min\{M, N\}$ and zero elsewhere. An example for $M = N = 4$ and different values of L is shown in Fig. 8.
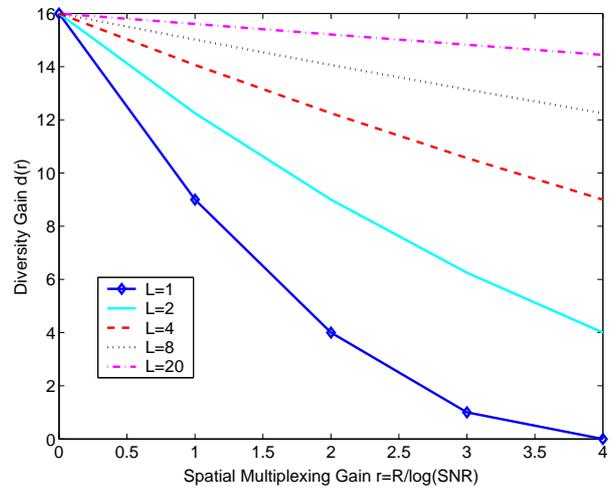

Fig. 8.   Long-term static DMDT versus the multiplexing gain r for $M = N = 4$.

For the case where the channel changes independently at each block (short-term static fading case), the corresponding expression is given by:

$$d_{\text{ss}}^*(r) = L d^\star \left(\frac{r}{L} \right). \tag{32}$$

In both cases, we see that especially in the region of high multiplexing gain, an optimal ARQ scheme allows for considerably larger diversity gains than a system with no ARQ ($L = 1$).

The DMDT for the case of power control yields extremely large values of the SNR exponent. This shows that, if the transmitter is allowed to boost power in order to compensate for channel outages, very significant improvement in error probability is possible.

The intuitive explanation of why ARQ is able to greatly improve the SNR exponent (diversity) can be provided as follows. At high SNR, errors are essentially due to the channel misbehavior, i.e., when the channel matrix is in the outage set. This is a rare event, in the sense that it decreases as ρ^{-d_1} for some $d_1 > 0$, the SNR exponent of errors in the first ARQ round. The transmission rate R is essentially determined by the rate of the first round R_1, since $R \approx R_1/(1 - \rho^{d_1}) \approx R_1$ when $\rho \to \infty$. On the contrary, if the decoder is able to *detect* errors with sufficiently high probability, the probability of error is essentially determined by the last round, where decoding is based on the received signal over all L blocks. The effective rate at the last round is R_1/L. Hence, it is reasonable to expect that the SNR exponent of the ARQ system with rate $R \doteq r \log \rho$ is given by the exponent of the corresponding MIMO channel $d^\star(\cdot)$ calculated at multiplexing gain r/L. The factor L in (32) is due to the additional *time diversity* that arises when each protocol round transmits a coding block over an independent channel realization. The rationale behind the above argument applies also to the case of power control, even though the expression of the SNR exponent is much more complicated and can be computed only via a recursive formula (see [80]) for details.

On the code construction side, [80] also introduced the *incremental redundancy lattice space-time (IR-LAST)* scheme. This is the ARQ version of the LAST codes treated in [53]. It is proved through a random coding argument that IR-LAST achieve the optimal DMDT. The scheme uses a list decoder, based on the MMSE-DFE Lattice Decoder modified to be a bounded distance decoder. The search radius is chosen in order to strike the optimal balance between probability of error (accepting a wrong message) and the probability of declaring a decoding error.

Following these results, explicit constructions of codes for the MIMO-ARQ channel are provided in [81]. This work provides sufficiency criteria for DMDT optimality based on a combination of the NVD criterion and another criterion that involves the Frobenius norm of the codeword matrices of the first round of transmission. Then, [81] explicitly constructs STCs based on structural modifications of previous square and rectangular CDA codes, that meet these criteria and are therefore DMDT-optimal. These codes are constructed for the cases where L is a multiple of M or L is a multiple of M, and incur minimum block length.

B. Space-Time Trellis Coded Modulation Revisited

In typical terrestrial wireless links, the coherence time is several orders of magnitude greater than the symbol duration. This immediately opens up the possibility of designing STCs with long block lengths T in order to achieve larger coding gain. This scenario is also relevant in a packet data communication system where the metric of performance is the packet error rate. In this case it is of interest to design codes with RT information bits of the order of one packet size (typically, between 100 and 1000 bits, depending on the applications). An additional design objective is to ensure that the STCs so constructed can be decoded with reasonable complexity. For example, for block length T of the order of 100 channel uses and $M = 2, 4$ antennas, even the most efficient implementation of the SD or of the MMSE-DFE (sequential) Lattice Decoder yields prohibitive computational complexity.

One option is to use the horizontal-stacking or row-deletion constructions of [49], in which rectangular DMT optimal STCs are constructed from their square counterparts by either concatenating several small block-length components or deleting rows from a large square code, obtained as described in Section IV-D. Although it is proven in [49] that these codes are DMT optimal, it turns out that the rectangular codes obtained by row deletion require very high decoding complexity and the horizontal stacking construction yields a large performance gap from optimal (information outage probability) for large block length.

In a recent work [82], the authors propose the use of a Space-Time Trellis-Coded Modulation (ST-TCM) scheme to improve the coding gain for large T. They focus on the $M = 2$ case and construct ST-TCM schemes based on partitions of the Golden code (these correspond to constructing partitions of the rotated $\mathbb{Z}^8$ lattice owing to the Golden code being a perfect code). The resulting codes are termed Golden ST-TCM (GST-TCM). Each branch on the trellis is labeled with these partitions in such a way that a lower bound on the minimum determinant of the resulting GST-TCM is maximized. Decoding of these codes is performed using a Viterbi decoder operating on the trellis, in conjunction with a SD that computes the branch metrics. While this type of trellis STC schemes is still at its early stage, we may say that designing STC that are provably DMT optimal, that performs well also at low to moderate SNR and that can handle information block length of the order to 100 to 1000 bits is an interesting open problem that is also extremely relevant for applications such as future wireless LANs.

C. Cooperative wireless networks

Space time schemes can be applied in the context of *cooperative diversity* protocols in wireless networks, where several users cooperate with each other to jointly achieve better error performance.

Several cooperative diversity protocols were recently presented (e.g., [83],[84],[85],[86]). These can be grouped into two main categories: 1) *amplify-and-forward*, where the assisting nodes perform a linear operation on the signal vector they

receive from the information source, before forwarding it; 2) *decode-and-forward*, where the assisting nodes try to decode the received signal, and eventually re-encode it before sending it again. The works in [83], [84], [85] investigate the DMT (high-SNR performance) of different cooperation protocols.

In this setting, outage based asymptotic optimality was thought to require infinite block length and infinite decoding complexity, and in some cases to require full knowledge of the channel at all receiver nodes. Novel space-time code constructions (see [87],[88]) have played an important role in improving these existing performance optimality bounds, and then meeting these bounds with finite and minimum block length, small encoding an decoding complexity, reduced channel knowledge, and for general network topologies and statistical characterizations [88].

D. The multiple access space-time channel

Consider a multiple-access quasi-static block-fading MIMO channel with independent Rayleigh fading with K users, each transmitting with M transmit antennas. Differently from the cooperative setting mentioned before, where the terminals can either transmit or receive and therefore can cooperate in order to deliver their message to the destination, in the multiple-access setting the K user terminals can only transmit and independently encode independently generated information messages. Hence, no explicit cooperation is possible [41]. The receiver is equipped with N antennas. Each user i has rate $R_i = r_i \log \rho$, i.e., its multiplexing gain is r_i. We say that the system is in error if one or more users' messages are erroneously decoded. For the overall system error probability to decay as fast as ρ^{-d}, there exists a limit on the multiplexing gain K-tuple $(r_1, r_2, \cdots, r_K)$ that the channel can support. In terms of the DMT, [89] shows that for block length $T \geq KM + N - 1$, the multiplexing gain region $\mathcal{R}(d)$ of K-tuples $(r_1, r_2, \cdots, r_K)$ such that the overall system error probability decays as fast as ρ^{-d}, is given by:

$$
\mathcal{R}(d) = \left\{ (r_1, r_2, \cdots, r_K) \in \mathbb{R}_+^K : \right.
$$
$$
\left. \sum_{s \in \mathcal{S}} r_s \leq r^*_{|\mathcal{S}|M,N}(d), \ \forall \mathcal{S} \subseteq \{1, \cdots, K\} \right\}
$$
(33)

where $r^*_{|\mathcal{S}|M,N}(d)$ is the solution of

$$
d^\star_{|\mathcal{S}|M,N}(r) = d
$$

where in this section we denote by $d^\star_{m,n}(r)$ the SNR exponent of a single-user MIMO channel with m transmit and n receive antennas, as given in (16). In words, $r^*_{|\mathcal{S}|M,N}(d)$ denotes the maximum allowable multiplexing gain for a given diversity d for a single user channel with $|\mathcal{S}|M$ transmit and N receiving antennas. The above results says that the system is not in outage if all single user channels corresponding to some selection of the users are not in outage.

For the symmetric case where all users have the same multiplexing gain $r_1 = \cdots = r_K = r$ and diversity gain d, equation (33) yields

$$
r = \min_{k=1,\cdots,K} \frac{1}{k} r^*_{kM,N}(d)
$$
(34)

and as a result the Multiple Access Channel (MAC) SNR exponent $d^\star_{\mathrm{mac}}(r)$ becomes

$$
d^\star_{\mathrm{mac}}(r) = \min_{k=1,\cdots,K} d^\star_{kM,N}(kr).
$$
(35)

Theorem 3 of [89], provides some insight by pointing out that there exist two regions of multiplexing gain where the performance of the multiple-access channel distinctly varies. In the first region, corresponding to a 'light' information load of the channel, the diversity-multiplexing performance of the system is as though there was only one user present. In the other region corresponding to a heavily loaded channel, the diversity-multiplexing performance is as if all the users are (implicitly) cooperating. In these two regions the optimum MAC MDT is given in [89] by the following SNR exponent:

$$
d^\star_{\mathrm{mac}}(r) = \begin{cases} d^\star_{M,N}(r) \\ \text{for } 0 \leq r \leq \min\left\{ M, \frac{N}{K+1} \right\} \\ d^\star_{KM,N}(Kr) \\ \text{for } \min\left\{ M, \frac{N}{K+1} \right\} \leq r \leq \min\left\{ M, \frac{N}{K} \right\} \end{cases}
$$
(36)

given that $T \geq KM + N - 1$. For $r > N/K$, the SNR exponent is zero. The explicit construction of codes for the MIMO MAC that achieve the above MAC DMT is an interesting open problem. In particular, the standard approach of successive decoding that achieves the capacity region of the non-fading Gaussian MAC falls short of achieving the optimal MAC DMT. Truly multiuser joint decoding is called for here, and its instrumentation with low complexity appears to be very challenging. The LAST framework with low-complexity MMSE-DFE Lattice Decoding (and related sequential decoding algorithms) might prove to be useful in this context.

VI. CONCLUSIONS

In this paper we have motivated the study of space-time codes for the quasi-static block-fading channel model starting from information theoretic arguments. We have identified the DMT optimality, and the stronger approximate universal DMT optimality, as the most desirable properties that a STC family should enjoy. A sequence of code constructions was presented, starting from the general linear-dispersion framework, and moving progressively upward to constructions that provably achieve the optimal DMT. Among those, the most interesting families of codes are CDA and LAST codes. Then, we have briefly discussed some approaches to combine the good features of these two families. The objective is to construct good codes in a deterministic manner (non-random lattices) while retaining the good features of LAST codes, and in particular the low decoding complexity of the MMSE-DFE Lattice Decoder in the modulo-Λ scheme.

Finally, we have listed a number of recent topics of research, such as the code design for MIMO ARQ schemes, the design of Trellis Coded STCs re-interpreted in light

of the CDA/LAST framework, the design of codes for the cooperative channel and for the MAC channel.

These, and several more subjects that have been left out for lack of space, are on-going research and offer a rich set of open problems. One of the purposes of this tutorial paper is that of introducing interested researchers to the state-of-the-art and motivating them to the study of this fascinating areas of coding and information theory.

REFERENCES

[1] I.E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[2] GJ Foschini and MJ Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[3] V. Tarokh, N. Seshadri, and AR Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *Information Theory, IEEE Transactions on*, vol. 44, no. 2, pp. 744–765, 1998.

[4] J.C. Guey, MP Fitz, MR Bell, and W.Y. Kuo, "Signal design for transmitter diversity wireless communicationsystems over Rayleigh fading channels," *Communications, IEEE Transactions on*, vol. 47, no. 4, pp. 527–537, 1999.

[5] B. M. Hochwald, G. Caire, B. Hassibi, and ed. T. L. Marzetta, "Special Issue on Space-Time Transmission, Reception, Coding, and Signal Processing," *IEEE Trans. Inform. Theory*, Oct. 2003.

[6] E. Biglieri and G. Taricco, *Transmission And Reception With Multiple Antennas: Theoretical Foundations*, Now Publishers Inc, 2004.

[7] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge University Press, 2003.

[8] E.G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*, Cambridge University Press, 2003.

[9] FD Neeser and JL Massey, "Proper complex random processes with applications to informationtheory," *Information Theory, IEEE Transactions on*, vol. 39, no. 4, pp. 1293–1302, 1993.

[10] L. Zheng and DNC Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *Information Theory, IEEE Transactions on*, vol. 49, no. 5, pp. 1073–1096, 2003.

[11] S. Verdu, "Spectral efficiency in the wideband regime," *Information Theory, IEEE Transactions on*, vol. 48, no. 6, pp. 1319–1343, 2002.

[12] A. Lapidoth and S.M. Moser, "Capacity Bounds Via Duality With Applications to Multiple-Antenna Systems on Flat-Fading Channels," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, 2003.

[13] A. Lapidoth, "On the Asymptotic Capacity of Stationary Gaussian Fading Channels," *IEEE TRANSACTIONS ON INFORMATION THEORY*, vol. 51, no. 2, pp. 437, 2005.

[14] T. Koch and A. Lapidoth, "The fading number and degrees of freedom in non-coherent MIMO fading channels: a peace pipe," in *Intern. Symp. on Inform. Theory, ISIT 2005*, Adelaide, Aus., Sept. 2005, pp. 661–665.

[15] E. Biglieri, J. Proakis, and S. Shamai, "Fading Channels: Information-Theoretic and Communications Aspects," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2619–2692, October 1998.

[16] J. Baltersee, G. Fock, and H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *Selected Areas in Communications, IEEE Journal on*, vol. 19, no. 12, pp. 2358–2368, December 2001.

[17] RH Etkin and DNC Tse, "Degrees of Freedom in Some Underspread MIMO Fading Channels," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1576–1608, 2006.

[18] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.

[19] A. Lozano, AM Tulino, and S. Verdu, "Multiple-antenna capacity in the low-power regime," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, pp. 2527–2544, 2003.

[20] DP Palomar, JM Cioffi, and MA Lagunas, "Uniform power allocation in MIMO channels: a game-theoretic approach," *Information Theory, IEEE Transactions on*, vol. 49, no. 7, pp. 1707–1727, 2003.

[21] SH Simon and AL Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *Selected Areas in Communications, IEEE Journal on*, vol. 21, no. 3, pp. 406–417, 2003.

[22] AL Moustakas and SH Simon, "Optimizing multiple-input single-output (MISO) communication systems with general Gaussian channels: nontrivial covariance and nonzero mean," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, pp. 2770–2780, 2003.

[23] L.W. Hanlen and A.J. Grant, "Optimal Transmit Covariance for Ergodic MIMO Channels," *Arxiv preprint cs.IT/0510060*, 2005.

[24] I. Bjelakovic and H. Boche, "Structure of Optimal Input Covariance Matrices for MIMO Systems with Covariance Feedback under General Correlated Fading," *Arxiv preprint cs.IT/0601107*, 2006.

[25] SA Jafar, S. Vishwanath, and A. Goldsmith, "Channel capacity and beamforming for multiple transmit and receiveantennas with covariance feedback," *Communications, 2001. ICC 2001. IEEE International Conference on*, vol. 7, 2001.

[26] W. Rhee and G. Taricco, "On the Ergodic Capacity-Achieving Covariance Matrix of Certain Classes of MIMO Channels," *Information Theory, IEEE Transactions on*, vol. 52, no. 8, pp. 3810–3817, August 2006.

[27] M. Debbah and RR Muller, "MIMO channel modeling and the principle of maximum entropy," *Information Theory, IEEE Transactions on*, vol. 51, no. 5, pp. 1667–1690, 2005.

[28] H. Shin and J.H. Lee, "Capacity of multiple-antenna fading channels: spatial fading correlation, double scattering, and keyhole," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, pp. 2636–2647, 2003.

[29] T. Ratnarajah, R. Vaillancourt, and M. Alvo, "Complex random matrices and Rayleigh channel capacity," *Communications in Information and Systems*, vol. 3, no. 2, pp. 119–138, 2003.

[30] M. Kiessling and J. Speidel, "Exact ergodic capacity of MIMO channels in correlated Rayleigh fading environments," *Proc. Intern. Zurich Seminar on Communications (IZS)*.

[31] M. Chiani, MZ Win, and A. Zanella, "On the capacity of spatially correlated MIMO Rayleigh-fading channels," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, pp. 2363–2371, 2003.

[32] A.M. Tulino, A. Lozano, and S. Verdú, "Capacity-achieving input covariance for single-user multi-antenna channels," *IEEE Trans. Wireless Commun*, 2006.

[33] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *Information Theory, IEEE Transactions on*, vol. 44, no. 3, pp. 927–946, 1998.

[34] Hochwald, B.M. and ten Brink, S., "Achieving near-capacity on a multiple-antenna channel," *Communications, IEEE Transactions on*, vol. 51, no. 3, pp. 389–399, March 2003.

[35] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *Communications, IEEE Transactions on*, vol. 52, pp. 670–678, April 2004.

[36] H. Vikalo and B. Hassibi, "Low-complexity iterative decoding over multiple antenna channels via a modified sphere decoder," in *Proc. 39th Allerton Conf. on Communication, Control and Computing*, Allerton, IL, October 2001, pp. 805–814.

[37] A.G. Fabregas and G. Caire, "Coded modulation in the block-fading channel: coding theorems and code construction," *Information Theory, IEEE Transactions on*, vol. 52, no. 1, pp. 91–114, 2006.

[38] AG i Fabregas and G. Caire, "Impact of signal constellation expansion on the achievable diversity of pragmatic bitinterleaved space-time codes," *IEEE Trans. Wireless Commun*, 2004.

[39] J. Boutros, N. Gresset, L. Brunel, and M. Fossorier, "Soft-input soft-output lattice sphere decoder for linear channels," *Global Telecommunications Conference, 2003. GLOBECOM'03. IEEE*, vol. 3, 2003.

[40] C. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: Turbo-equalization," *European Transactions on Telecommunications*, vol. 6, no. 5, pp. 507–511, 1995.

[41] T.M. Cover and J.A. Thomas, *Elements of information theory*, Wiley New York, 1991.

[42] PW Wolniansky, GJ Foschini, GD Golden, and RA Valenzuela, "V-BLAST: an architecture for realizing very high data rates overthe rich-scattering wireless channel," *Signals, Systems, and Electronics, 1998. ISSSE 98. 1998 URSI International Symposium on*, pp. 295–300.

[43] U. Wachsmann, R. Fischer, and J. Huber, "Multilevel Codes: Theoretical Concepts and Practical Design Rules," *Information Theory, IEEE Transactions on*, vol. 45, no. 4, pp. 1361–1391, July 1999.

[44] S. Verdu and T.S. Han, "A general formula for channel capacity," *Information Theory, IEEE Transactions on*, vol. 40, no. 4, pp. 1147–1157, 1994.

[45] LH Ozarow, S. Shamai, and AD Wyner, "Information theoretic considerations for cellular mobile radio," *Vehicular Technology, IEEE Transactions on*, vol. 43, no. 2, pp. 359–378, 1994.

[46] Zhengdao Wang and G.B. Giannakis, "Outage mutual information of space-time MIMO channels," *Information Theory, IEEE Transactions on*, vol. 50, no. 4, pp. 657–662, April 2004.

[47] E. Malkamaki and H. Leib, "Coded diversity on block-fading channels," *Information Theory, IEEE Transactions on*, vol. 45, no. 2, pp. 771–781, 1999.

[48] H. El Gamal and MO Damen, "Universal space-time coding," *Information Theory, IEEE Transactions on*, vol. 49, no. 5, pp. 1097–1119, 2003.

[49] P. Elia, K.R. Kumar, S.A. Pawar, P.V. Kumar, and H.F. Lu, "Explicit, minimum-delay space-time codes achieving the diversity-multiplexing gain tradeoff," *To appear, Information Theory, IEEE Transactions on,*, vol. 52, no. 9, 2006.

[50] K.R. Kumar and G. Caire, "Capacity Analysis and Code Design for MIMO Fading Channels," *USC CSI Technical Report: CSI-06-07-01*, 2006.

[51] SM Alamouti, "A simple transmit diversity technique for wireless communications," *Selected Areas in Communications, IEEE Journal on*, vol. 16, no. 8, pp. 1451–1458, 1998.

[52] V. Tarokh, H. Jafarkhani, and AR Calderbank, "Space-time block codes from orthogonal designs," *Information Theory, IEEE Transactions on*, vol. 45, no. 5, pp. 1456–1467, 1999.

[53] HE Gamal, G. Caire, and MO Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels," *Information Theory, IEEE Transactions on*, vol. 50, no. 6, pp. 968–985, 2004.

[54] P. Tavildar, S.; Viswanath, "Approximately universal codes over slow-fading channels," *Information Theory, IEEE Transactions on*, vol. 52, no. 7, pp. 3233 – 3258, 2006.

[55] B. Hassibi and BM Hochwald, "High-rate codes that are linear in space and time," *Information Theory, IEEE Transactions on*, vol. 48, no. 7, pp. 1804–1824, 2002.

[56] E. Viterbo and J. Bouros, "A universal lattice code decoder for fading channels," *Information Theory, IEEE Transactions on*, vol. 45, no. 5, pp. 1639–1642, 1999.

[57] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *Information Theory, IEEE Transactions on*, vol. 48, no. 8, pp. 2201–2214, 2002.

[58] MO Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, pp. 2389–2402, 2003.

[59] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. Expected complexity," *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, vol. 53, no. 8, pp. 2806–2818, 2005.

[60] AD Murugan, HE Gamal, MO Damen, and G. Caire, "A unified framework for tree search decoding: rediscovering the sequential decoder," *Information Theory, IEEE Transactions on*, vol. 52, no. 3, pp. 933–953, 2006.

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

[62] X.B. Liang and X.G. Xia, "On the nonexistence of rate-one generalized complex orthogonal designs," *Information Theory, IEEE Transactions on*, vol. 49, no. 11, pp. 2984–2988, 2003.

[63] HE Gamal and AR Hammons Jr, "A new approach to layered space-time coding and signal processing," *Information Theory, IEEE Transactions on*, vol. 47, no. 6, pp. 2321–2334, 2001.

[64] G.J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[65] J. Boutros, E. Viterbo, C. Rastello, and J.C. Belfiore, "Good lattice constellations for both Rayleigh fading and Gaussian channels," *Information Theory, IEEE Transactions on*, vol. 42, no. 2, pp. 502–518, 1996.

[66] J. Boutros and E. Viterbo, "Signal space diversity: a power-and bandwidth-efficient diversitytechnique for the Rayleigh fading channel," *Information Theory, IEEE Transactions on*, vol. 44, no. 4, pp. 1453–1467, 1998.

[67] B.A. Sethuraman and B.S. Rajan, "Full-rank, full-rate STBCs from division algebras," *Information Theory Workshop, 2002. Proceedings of the 2002 IEEE*, pp. 69–72, 2002.

[68] J.C. Belfiore and G. Rekaya, "Quaternionic lattices for space-time coding," *Information Theory Workshop, 2003. Proceedings. 2003 IEEE*, pp. 267–270.

[69] BA Sethuraman, BS Rajan, and V. Shashidhar, "Full-diversity, high-rate space-time block codes from division algebras," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, pp. 2596–2616, 2003.

[70] P. Elia, KR Kumar, SA Pawar, and PV Kumar, "Explicit space-time codes that achieve the diversity-multiplexing gain tradeoff," *Proc. International Symposium on Information Theory*, pp. 896–900, 2005.

[71] J.C. Belfiore, G. Rekaya, and E. Viterbo, "The golden code: a 2 × 2 full-rate space-time code with nonvanishing determinants," *Information Theory, IEEE Transactions on*, vol. 51, no. 4, pp. 1432–1436, 2005.

[72] T. Kiran and B.S. Rajan, "STBC-Schemes with Nonvanishing Determinant for Certain Number of Transmit Antennas," *IEEE Transactions on Information Theory*, vol. 51, no. 8, pp. 2984–2992, 2005.

[73] U. Erez and R. Zamir, "Achieving 1/2 log (1+ SNR) on the AWGN channel with lattice encoding and decoding," *Information Theory, IEEE Transactions on*, vol. 50, no. 10, pp. 2293–2314, 2004.

[74] H.A. Loeliger, "Averaging bounds for lattices and linear codes," *Information Theory, IEEE Transactions on*, vol. 43, no. 6, pp. 1767–1773, 1997.

[75] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *Information Theory, IEEE Transactions on*, vol. 51, no. 10, pp. 3401–3416, 2005.

[76] F. Oggier, G.R.B. Othman, J.C. Belfiore, and E. Viterbo, "Perfect Space Time Block Codes," *Arxiv preprint cs.IT/0604093*, 2006.

[77] P. Elia, BA Sethuraman, and PV Kumar, "Perfect Space-Time Codes with Minimum and Non-Minimum Delay for Any Number of Antennas," *Submitted to IEEE Transactions on Information Theory*, 2006.

[78] K. Raj Kumar and Giuseppe Caire, "Construction of Structured LaST Codes," *Proc. International Symposium on Information Theory, Seattle*, July 2006.

[79] J.H. Conway and NJ Sloane, *Sphere Packings, Lattices and Groups*, Springer, 1998.

[80] H. El Gamal, G. Caire, and M.O. Damen, "The MIMO ARQ Channel: Diversity-Multiplexing-Delay Tradeoff," *Information Theory, IEEE Transactions on*, vol. 52, no. 8, pp. 3601–3621, 2006.

[81] S.A. Pawar, K.R. Kumar, P. Elia, BA Sethuraman, and P.V. Kumar, "Minimum-Delay Space-Time Codes Achieving the DMD Tradeoff of the MIMO-ARQ Channel," *submitted to IEEE Trans. Inform. Theory, Oct*, 2005.

[82] Y. Hong, E. Viterbo, and J.C. Belfiore, "Golden Space-Time Trellis Coded Modulation," *Arxiv preprint cs.IT/0604063*, 2006.

[83] JN Laneman and GW Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, pp. 2415–2425, 2003.

[84] JN Laneman, DNC Tse, and GW Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *Information Theory, IEEE Transactions on*, vol. 50, no. 12, pp. 3062–3080, 2004.

[85] K. Azarian, HE Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *Information Theory, IEEE Transactions on*, vol. 51, no. 12, pp. 4152–4172, 2005.

[86] H. Bolcskei, R.U. Nabar, O. Oyman, and A.J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Commun*, 2006.

[87] S. Yang and J.C. Belfiore, "Optimal space-time codes for the MIMO amplify-and-forward cooperative channel," *Arxiv preprint cs.IT/0509006*, 2005.

[88] Petros Elia, "Asymptotic Universal Optimality in Wireless Multi-Antenna Delay-Limited Communications and Wireless Networks," *Ph.D. Thesis, University of Southern California*, 2006.

[89] D.N.C. Tse, P. Viswanath, and L. Zheng, "Diversity–Multiplexing Tradeoff in Multiple-Access Channels," *Information Theory, IEEE Transactions on*, vol. 50, no. 9, pp. 1859–1874, 2004.

Giuseppe Caire was born in Torino, Italy, in 1965. He received the B.Sc. in Electrical Engineering from Politecnico di Torino (Italy), in 1990, the M.Sc. in Electrical Engineering from Princeton University in 1992 and the Ph.D. from Politecnico di Torino in 1994. He was a recipient of the AEI G.Someda Scholarship in 1991, has been with the European Space Agency (ESTEC, Noordwijk, The Netherlands) from May 1994 to February 1995, was a recipient of the COTRAO Scholarship in 1996 and of a CNR Scholarship in 1997.

He has been visiting Princeton University in summer 1997 and Sydney University in summer 2000. He has been Assistant Professor in Telecommunications at the Politecnico di Torino, Associate Professor at the University of Parma, Italy, Professor with the Department of Mobile Communications at the Eurecom Institute, Sophia-Antipolis, France, and he is now professor with the EE Department of the Viterbi School of Engineering, University of Southern California, Los Angeles, CA.

He served as Associate Editor for the IEEE Transactions on Communications in 1998-2001 and as Associate Editor for the IEEE Transactions on Information Theory in 2001-2003. He received the Jack Neubauer Best System Paper Award from the IEEE Vehicular Technology Society in 2003, and the Joint IT/Comsoc Best Paper Award in 2004. Since November 2004 he is member of the Board of Governors of the IEEE Information Theory Society and is Fellow of IEEE.

His current interests are in the field of communications theory, information theory and coding theory with particular focus on wireless applications.

Petros Elia received the B.Sc. degree in electrical engineering from the Illinois Institute of Technology, Chicago, in 1997. During 1997-99 he worked as a consultant for MIOD Technologies, Chicago. He received the M.Sc. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles, in 2001 and 2006, respectively. His current research interests include communication and cooperation in wireless networks, MIMO systems and information theory.

K. Raj Kumar received the B.E. degree from the University of Madras, Madras, India in 2003 and the M.Sc.(Engg.) degree from the Indian Institute of Science, Bangalore, India in 2005. Since August 2005, he has been pursuing the Ph.D. degree at the Department of Electrical EngineeringSystems, University of Southern California, Los Angeles. His current research interests include spacetime coding, coded modulation, and information theory.