Diversity Embedded Space-Time Codes

S N. Diggavi[*] A R. Calderbank S. Dusad N. Al-Dhahir

Abstract

Rate and diversity impose a fundamental trade-off in wireless communication. High-rate space-time codes come at a cost of lower reliability (diversity), and high reliability (diversity) implies a lower rate. However, wireless networks need to support applications with very different Quality-of-Service (QoS) requirements, and it is natural to ask what characteristics should be built into the physical layer link in order to accomodate them. In this paper we design high-rate space-time codes that have a high-diversity code embedded within them. This allows a form of communication where the high-rate code opportunistically takes advantage of good channel realizations while the embedded high-diversity code provides guarantees that at least part of the information is received reliably. We provide constructions of linear and non-linear codes for a *fixed transmit alphabet* constraint. The non-linear constructions are a natural generalization to wireless channels of multilevel codes developed for the AWGN channel that are matched to binary partitions of QAM and PSK constellations. The importance of set-partitioning to code design for the wireless channel is that it provides a mechanism for translating constraints in the binary domain into lower bounds on diversity protection in the complex domain. We investigate the systems implications of embedded diversity codes by examining value to unequal error protection, rate opportunism and packet delay optimization. These applications demonstrate that diversity-embedded codes have the potential to outperform traditional single-layer codes in moderate SNR regimes.

1 Introduction

Multiple antennas are a means of enabling reliable high-data rate wireless communications. High rates are enabled by spatial multiplexing as shown in [37, 22]. High reliability is enabled by correlation of signals across the multiple antennas as shown in [39].

These two points-of-view can be combined by observing that there is a trade-off between rate and reliability (diversity). This was explored in the context of finite-rate (fixed alphabet size) codes in [39] and in the context of information-theoretic rate growth in [43]. Both these results show that to achieve a high transmission rate, one might need to sacrifice reliability (diversity) and vice-versa. Therefore, the main question addressed by most researchers has been how to design transmission techniques that achieve a particular point on rate-reliability trade-off. For example, the emphasis

[*]S N. Diggavi is the contact author.

in the coding literature has been the design of maximal-diversity codes, *i.e.* one extremal point in this trade-off. In [37, 22], the focus is on obtaining the maximal spatial multiplexing rate for ergodic channels. There are also codes which achieve particular points in the rate-reliability trade-off (see for example, [29]).

Wireless networks need to support a variety of applications with different quality-of-service (QoS) requirements. For example, real-time applications need lower delay and therefore higher reliability (diversity) as compared to non-real time applications. The delay requirement for real-time applications could be stringent enough to require decoding over a *single coherence time*[1]. Therefore, we would need ensure a high diversity order for real-time applications in order to ensure reliability over a quasi-static channel. However, for non-real time applications, the delay requirement could span several channel coherence times, and therefore reliability can be obtained by coding over these different channel fades, for example through ARQ protocols. Therefore, a natural question to ask is what characteristics do we need to build into the physical layer link in order to accomodate these diverse applications with heterogenous requirements. In particular, if we design the overall system for a fixed rate-diversity operating point, we might be over-provisioning a resource which could be flexibly allocated to different applications. This thought process leads to viewing diversity (reliability) as a systems resource that can be allocated judiciously to satisfy the QoS requirements for the different applications. It motivates the design of a physical layer that can *simultaneously* provide multiple rate-reliability points which can be allocated to the applications appropriately. The main contribution of this paper is the design of codes that achieve a high-rate, but have embedded within them a high-diversity (lower-rate) code. Clearly the overall code will still be governed by the rate-diversity trade-off, but the idea is to ensure the high reliability (diversity) for one stream with perhaps higher rate and lower reliability for the other. By providing flexibility in the diversity allocation, we can simultaneously accommodate multiple applications with very different rate-reliability requirements. A simple way to achieve this would be to time division multiplex (TDM) between codes that attain different rate-diversity points. In this paper, we construct space-time codes that guarantee rate-reliability tuples that cannot be achieved by such simple switching strategies.

If data streams come from the same application (information source), we might interpret the different reliabilities as giving unequal error protection (UEP) to different constituents of the same data stream. The idea of UEP has a long and rich history (see for example [34, 11] and references therein). UEP codes have been designed for the binary Hamming distance metric (see for example [30]) and for the Euclidean distance metric encountered in the Gaussian channel (see for example [6]). When streams come from the same application, we may interpret our framework for code-design as enabling UEP with respect to the diversity metric suitable for fading channels [39].

Hybrid ARQ strategies using incremental redundancy codes have been proposed for transmission over highly variable wireless channels [25, 32]. Here, a low redundancy code is sent first and if an error is detected then the parity symbols are sent upon feedback of an error message. Initial implementations of the hybrid ARQ made use of incremental redundancy codes designed for binary or Gaussian channels. More recently, hybrid ARQ schemes with codes suitable for the fading

[1]A coherence time of the channel is the period over which the fading channel remains approximately constant. This varies from a few tens of milliseconds to hundreds of milliseconds, depending on mobility and carrier frequency [40].

channels have also been proposed [21]. Diversity-embedded codes provide a means of provisioning queues with different priorities (reliabilities). We propose a variant of hybrid ARQ that uses feedback to schedule information over these prioritized queues.

Another way to view the embedded diversity is from the point of view of compound channels [12]. Here, information is transmitted over a channel chosen by nature from a certain class of channels. The particular choice is unknown to the transmitter (and sometimes the receiver as well). Clearly, the quasi-static fading wireless channel, where the coding is over one coherence time, is a non-ergodic channel that fits the compound channel problem. A robust strategy would be to transmit at a rate such that the information can be decoded reliably for *every* choice of the channel. However, for the wireless channel, this would mean a rate of zero, since nature can always choose with positive probability measure a fading channel that cannot support the given rate [33]. However, one can give up on a subset of the channels and declare an *outage* and transmit reliably at a positive rate for the rest of the channels. [33, 37, 22]. Therefore, most transmission schemes are designed for a particular rate at a given outage probability (reliability). In [9], Cover proposed a broadcast strategy for transmission over compound channels, where the idea was to superpose codewords such that the achievable rate could be strictly better than the above conservative strategy if a more benign channel were available. The idea was to make the average performance better at the cost of the worst case performance. In light of this, one can interpret the provisioning of multiple rate-reliability points in terms of Cover's broadcast strategy for compound channels. This gives an interpretation of diversity embedded codes as an opportunistic coding strategy that takes advantage of benign channels while giving some guarantees when deep fades occur.

In fact, such a broadcast strategy was proposed for single-antenna fading channels in [35], where superposition coding was used to maximize the average transmission rate. The multiple-antenna channel provides additional degrees of freedom that can be exploited to increase the rate and/or diversity [37, 22, 39, 43]. In this paper we utilize these additional degrees of freedom in order to design diversity embedded codes. Diversity embedded codes were introduced in [15, 16], a class of non-linear constructions was presented in [7] and analysis of the diversity embedding properties was given in [17]. Independently, Shamai [36] also presented some extensions of the broadcast approach of single-antenna channels of [35] to multiple antenna channels, again to maximize the average information rate. An information-theoretic analysis of diversity embedded codes has been done in [18, 19, 20].

The main contributions of this paper are as follows. We introduce the idea of space-time codes which have multiple levels of reliability for fading channels, which we call *diversity-embedded codes*. We give the objectives and design criteria for such constructions in Section 3. This parallels the background on "single-layer" space-time codes given in Section 2. We give some linear constructions of diversity-embedded codes and analyze their properties in Section 4. We give a class of multi-level constructions for diversity embedded block codes in Section 5 and analyze their diversity embedding properties in the Appendix. We extend these constructions to diversity embedded trellis codes in Section 6. The multi-level constructions given in Section 5 were first presented in [7]. Independently, a similar idea for space-time code design for achieving a *particular* rate-diversity point has also been presented in [29]. Our constructions are for diversity embedded codes and *not* for a fixed rate-diversity point, and were developed independently[2] [7]. The system implications of

[2]We thank P V. Kumar for bringing [29] to our attention.

these codes is explored in Section 7. We conclude the paper with a brief discussion of several open issues.

2 Preliminaries

Our focus in this paper is on the quasi-static flat–fading channel where we transmit information coded over M_t transmit antennas and have M_r antennas at the receiver. Furthermore, the assumption is that the transmitter has no channel state information, whereas the receiver is able to perfectly track the channel (a common assumption, see for example [3, 39]). The code is designed over T transmission symbols and the received signal after demodulation and sampling can be written as

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z}, \tag{2.1}$$

where $\mathbf{Y} = [\mathbf{y}(0), \ldots, \mathbf{y}(T-1)] \in \mathbb{C}^{M_r \times T}$ is the received sequence, $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$ is the quasi-static channel fading matrix, $\mathbf{X} = [\mathbf{x}(0), \ldots, \mathbf{x}(T-1)] \in \mathbb{C}^{M_t \times T}$ is the space-time coded transmission sequence with transmit power constraint P, and $\mathbf{Z} = [\mathbf{z}(0), \ldots, \mathbf{z}(T-1)] \in \mathbb{C}^{M_r \times T}$ is assumed to be additive white Gaussian noise with variance σ^2.

The coding scheme is limited to one quasi-static transmission block. Similar arguments can be made if we are allowed to code across only a *finite* number of quasi-static transmission blocks [33]. This allows us to view the channel in (2.1) as a non-ergodic channel since the performance is determined by single randomly chosen channel fading matrix $\mathbf{H}$. In this context we define the notion of diversity order [39] as follows.

Definition 2.1 *A coding scheme which has an average error probability $\bar{P}_e(SNR)$ as a function of SNR that behaves as*

$$\lim_{SNR \to \infty} \frac{\log(\bar{P}_e(SNR))}{\log(SNR)} = -d \tag{2.2}$$

is said to have a diversity order of d.

In words, a scheme with diversity order d has an error probability at high SNR behaving as $\bar{P}_e(\mathrm{SNR}) \approx \mathrm{SNR}^{-d}$.

2.1 Code Design Criteria

For codes designed for a finite (and fixed) rate, one can bound the error probability by using *pairwise error probability* (PEP) between two candidate codewords. This leads to the now well-known rank criterion for determining the diversity order of a space-time code [39, 24]. Consider a codeword sequence $\mathbf{X} = [\mathbf{x}(0), \ldots, \mathbf{x}(T-1)]$ as defined in (2.1), where $\mathbf{x}(k) = [\mathbf{x}_1(k), \ldots, \mathbf{x}_{M_t}(k)]^T$. The PEP between two codewords $\mathbf{x}$ and $\mathbf{e}$ can be determined by the codeword difference matrix $\mathbf{B}(\mathbf{x}, \mathbf{e})$ [39, 24], where

$$\mathbf{B}(\mathbf{x}, \mathbf{e}) = \begin{pmatrix} \mathbf{x}_1(0) - \mathbf{e}_1(0) & \ldots & \mathbf{x}_1(T-1) - \mathbf{e}_1(T-1) \\ \vdots & \vdots & \vdots \\ \mathbf{x}_{M_t}(0) - \mathbf{e}_{M_t}(0) & \ldots & \mathbf{x}_{M_t}(T-1) - \mathbf{e}_{M_t}(T-1) \end{pmatrix}.$$

For fixed rate codebook $\mathcal{C}$, the pairwise error probability between two distinct codewords $\mathbf{x}$ and $\mathbf{e}$ is given by[3] [39]

$$P_e(\mathrm{SNR}, \mathbf{x} \longrightarrow \mathbf{e}) \doteq \mathrm{SNR}^{-M_r \mathrm{rank}(\mathbf{B}(\mathbf{x},\mathbf{e}))}. \tag{2.3}$$

Since we are dealing with a fixed rate codebook, by using the simple union bound argument, it can be shown that the diversity order is given by [39]

$$d = M_r \min_{\mathbf{x} \neq \mathbf{e} \in \mathcal{C}} \mathrm{rank}(\mathbf{B}(\mathbf{x}, \mathbf{e}))\ . \tag{2.4}$$

The error probability is determined by *both* the coding gain and the diversity order. Hence, the code design criterion prescribed in [39] is to design the codebook $\mathcal{C}$ so that the minimal rank of the codeword difference matrix corresponds to the required diversity order and the minimal determinant gives the corresponding coding gain. Note that the actual error performance is also determined by the number of nearest neighbors in the code. In this paper the focus is on the diversity order only, and we define corresponding criteria for embedded codebooks.

2.2 Rate and Diversity

For a given diversity order, it is natural to ask for upper bounds on achievable rate. For a flat Rayleigh fading channel, this has been examined in [39] where the following result was obtained.

Theorem 2.2 *([39, 28]) If we use a constellation of size 2^b and the diversity order of the system is qM_r, then the rate R that can be achieved is bounded as*

$$R \leq (M_t - q + 1) \log_2 |\mathcal{S}_T| \tag{2.5}$$

in bits per transmission, where $\mathcal{S}_T$ is the transmit alphabet or constellation.

The discussion in Section 2.1 and Theorem 2.2 is suitable for fixed-rate (finite-alphabet) codebooks. Another point-of-view explored in [43], allows the rate of the codebook to increase with SNR, *i.e.,* defining a multiplexing rate. This viewpoint is motivated because the capacity of the multiple-antenna channel grows with SNR behaving as $\min(M_r, M_t) \log(\mathrm{SNR})$ [22, 43, 14], at high SNR even for finite M_r, M_t. Therefore, just as Theorem 2.2 shows the tension between achieving high-rate and high-diversity for fixed transmit alphabet constraint, there exists a tension between multiplexing rate and diversity as well [43]. The question addressed in this paper is whether we can design codes with high-rate which contain within them a code which has high-diversity.

---

[3] We use the notation $\doteq$ to denote exponential equality [10], *i.e.,* $g(\mathrm{SNR}) \doteq \mathrm{SNR}^a$ means that $\lim_{\mathrm{SNR}\to\infty} \frac{\log g(\mathrm{SNR})}{\log \mathrm{SNR}} = a$. Moreover, if $g(\mathrm{SNR}) \doteq f(\mathrm{SNR})$, it means that, $\lim_{\mathrm{SNR}\to\infty} \frac{\log g(\mathrm{SNR})}{\log \mathrm{SNR}} = \lim_{\mathrm{SNR}\to\infty} \frac{\log f(\mathrm{SNR})}{\log \mathrm{SNR}}$.

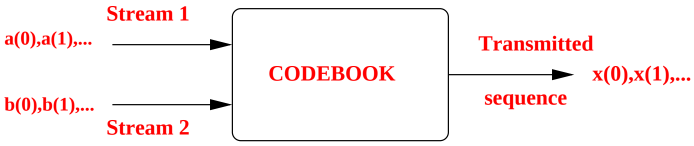


Figure 1: Embedded codebook

3 Diversity Embedding

The codebook structure proposed in this paper takes two information streams (see Figure 1) and outputs the transmitted sequence $\{\mathbf{x}(k)\}$. The objective is to ensure that each information stream gets the designed rate and diversity levels. In this section, we describe this approach and criteria in further detail, with specific code constructions given in Sections 4, 5 and 6.

3.1 Opportunistic Communication

The basic idea is to design a codebook which allows many levels of diversity for information streams. In particular we focus on two levels of diversity, but this procedure can be easily generalized to more levels. We design the transmitted information stream $\{\mathbf{x}(k)\}$ to depend on two information streams $\{\mathbf{a}(k)\}$ and $\{\mathbf{b}(k)\}$. Let $\mathcal{A}$ denote the message set from the first information stream and $\mathcal{B}$ denote that from the second information stream. The rates for the two message sets are, respectively, $R(\mathcal{A})$ and $R(\mathcal{B})$. The decoder jointly decodes the two message sets and we can define two error probabilities, $\bar{P}_e(\mathcal{A})$ and $\bar{P}_e(\mathcal{B})$ which denote the average error probabilities for message sets $\mathcal{A}$ and $\mathcal{B}$, respectively. We design the code such that a certain tuple (R_a, D_a, R_b, D_b) of rates and diversities are achievable, where $R_a = R(\mathcal{A}) = \frac{\log(|\mathcal{A}|)}{T}$, $R_b = R(\mathcal{B}) = \frac{\log(|\mathcal{B}|)}{T}$ and analogous to Definition 2.1,

$$D_a = \lim_{\mathrm{SNR}\to\infty} \frac{\log \bar{P}_e(\mathcal{A})}{\log(\mathrm{SNR})}, \quad D_b = \lim_{\mathrm{SNR}\to\infty} \frac{\log \bar{P}_e(\mathcal{B})}{\log(\mathrm{SNR})}. \tag{3.1}$$

Hence, if we examine the joint codebook, $\mathcal{C} = \{\mathcal{A}, \mathcal{B}\}$, the total rate is $R = R_a + R_b$ and the diversity is $D = \min(D_a, D_b)$. Therefore, in a baseline comparison with a single-layer codebook, the rate-diversity operating point is $(R_a + R_b, \min(D_a, D_b))$. A trivial outer-bound for the rate tuple is given from the single-layer bound in Theorem 2.2 as follows.

$$\begin{aligned} D_a &\leq \lfloor M_t - R_a / \log(|\mathcal{S}_T|) + 1 \rfloor \\ D_b &\leq \lfloor M_t - R_b / \log(|\mathcal{S}_T|) + 1 \rfloor \\ \min(D_a, D_b) &\leq \lfloor M_t - (R_a + R_b) / \log(|\mathcal{S}_T|) + 1 \rfloor. \end{aligned} \tag{3.2}$$

Note that we could also have examined information-theoretic bounds on the achievable tuple (R_a, D_a, R_b, D_b), and we explore this question in [19].

If $(R_a + R_b, D^*)$ is the optimal single-layer rate-diversity point predicted by Theorem 2.2, then we would like $D_a > D^*$ implying that $D_b \leq D^*$. This means that a rate $(R_a + R_b, \min(D_a, D_b))$, has a high-diversity message set $\mathcal{A}$ embedded within it. The main focus of this paper is to construct codes with this property and also to explore some characteristics of such codes. As mentioned earlier, a simple switching (TDM) strategy between codebooks designed for different rate-diversity levels could produce such a construction. However, in Section 4 we construct codes that cannot be constructed by simple switching strategies.

The reason why embedded diversity codes allow a form of opportunistic communication is that we can design the total rate $R_a + R_b$ to be large, and design D_a to be large as well. The opportunism arises because if the channel realization $\mathbf{H}$ is good, in that the channel matrix is far from singular, then the high total rate allows us to send a large amount of information through such a clear channel without knowing a-priori the channel realization at the transmitter. However, if the channel realization is not good, the embedded high-diversity message set $\mathcal{A}$ might still be decodable.

3.2 Code Design Criteria

We examine code design criteria for finite-alphabet (fixed-rate) codes. The transmitted sequence $\{\mathbf{x}(k)\}$ can be divided into codeword clusters as shown in Figure 2, where the codeword clusters shown correspond to all $\{\mathbf{x}(k)\}$ arising due to a particular message $\mathbf{a} \in \mathcal{A}$. The joint decoding of the two message sets yields candidate clusters and elements within the clusters. Given a desired tuple (R_a, D_a, R_b, D_b), we need to design message sets $\mathcal{A}, \mathcal{B}$ with size $2^{R_a T}$ and $2^{R_b T}$, respectively, to have the following properties. For message set $\mathcal{A}$, we want to design the code such that the pairwise error probability behaves as SNR^{-D_a}. Similarly for message set $\mathcal{B}$, we want to achieve diversity order of D_b. The code design criterion ensures that both these constraints are satisfied.

Let us examine a maximum-likelihood decoder jointly on the message sets $\mathcal{A}, \mathcal{B}$. Denote transmitted sequences as $\{\mathbf{x}_{\mathbf{a},\mathbf{b}}(k)\}$, with the subscripts denoting that the sequence arose due to messages $\mathbf{a} \in \mathcal{A}, \mathbf{b} \in \mathcal{B}$ as shown in Figure 1. Using (2.1) the maximum likelihood decoding metric when the channel state information is available at the receiver is given by,

$$\left(\hat{\mathbf{a}}, \hat{\mathbf{b}}\right) = \arg \min_{\mathbf{a} \in \mathcal{A}, \mathbf{b} \in \mathcal{B}} ||\mathbf{Y} - \mathbf{H}\mathbf{X}_{\mathbf{a},\mathbf{b}}||^2. \tag{3.3}$$

Therefore using (2.3) we can write the pairwise error probability on the joint message set as,

$$P_e(\text{SNR}, \{\mathbf{a}, \mathbf{b}\} \longrightarrow \{\hat{\mathbf{a}}, \hat{\mathbf{b}}\}) \doteq \text{SNR}^{-M_r \text{rank}(\mathbf{B}(\mathbf{X}_{\mathbf{a},\mathbf{b}}, \mathbf{X}_{\hat{\mathbf{a}},\hat{\mathbf{b}}}))}. \tag{3.4}$$

The pairwise error probability between two codewords in the message set $\mathcal{A}$ is given by,

$$P_e(\text{SNR}, \mathbf{a} \longrightarrow \hat{\mathbf{a}}) \doteq \text{SNR}^{-M_r \min_{\mathbf{b}_1, \mathbf{b}_2 \in \mathcal{B}} \text{rank}(\mathbf{B}(\mathbf{X}_{\mathbf{a},\mathbf{b}_1}, \mathbf{X}_{\hat{\mathbf{a}},\mathbf{b}_2}))}. \tag{3.5}$$

Since we are dealing with fixed rate codebooks, by using a simple union bound argument we can write the diversity d_a for message set $\mathcal{A}$ as,

$$d_a = M_r \min_{\mathbf{a}_1 \neq \mathbf{a}_2 \in \mathcal{A}} \min_{\mathbf{b}_1, \mathbf{b}_2 \in \mathcal{B}} rank(\mathbf{B}(\mathbf{x}_{\mathbf{a}_1,\mathbf{b}_1}, \mathbf{x}_{\mathbf{a}_2,\mathbf{b}_2})) \tag{3.6}$$

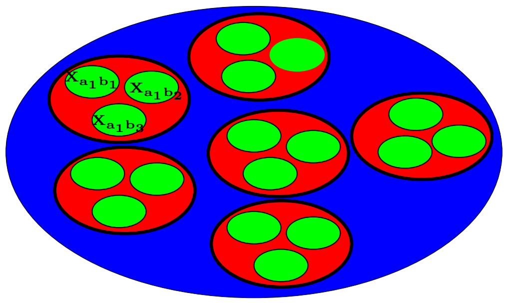


Figure 2: Codeword Clusters

Hence for the design constraint to be met, we need

$$\min_{\mathbf{a}_1 \neq \mathbf{a}_2 \in \mathcal{A}} \min_{\mathbf{b}_1, \mathbf{b}_2 \in \mathcal{B}} rank(\mathbf{B}(\mathbf{x}_{\mathbf{a}_1,\mathbf{b}_1}, \mathbf{x}_{\mathbf{a}_2,\mathbf{b}_2})) \geq D_a/M_r \tag{3.7}$$

In an identical manner, we can show for the message set $\mathcal{B}$, we need the following to hold.

$$\min_{\mathbf{b}_1 \neq \mathbf{b}_2 \in \mathcal{B}} \min_{\mathbf{a}_1, \mathbf{a}_2 \in \mathcal{A}} rank(\mathbf{B}(\mathbf{x}_{\mathbf{a}_1,\mathbf{b}_1}, \mathbf{x}_{\mathbf{a}_2,\mathbf{b}_2})) \geq D_b/M_r. \tag{3.8}$$

Basically, (3.7) implies that if we choose to transmit a particular message $\mathbf{a} \in \mathcal{A}$, regardless of which message is chosen in message set $\mathcal{B}$, we are ensured a diversity level of D_a for this message set. A similar argument will hold for message set $\mathcal{B}$.

Note that this criterion focuses *only* on the diversity level achieved by the code design. In general, the error probability depends *both* on the diversity order defined in Definition 2.1 as well as a coding gain [39]. The coding gain could be an important issue especially in low to moderate SNR regimes. As mentioned in Section 2.1, another important issue affecting error performance is that of the number of nearest neighbors. Also, the code design criterion is suitable only for fixed rate codebooks. When one needs to design codebooks whose rates grow with SNR [43], alternative code design criteria could be formulated.

3.3 Set partitioning of QAM and QPSK constellations

Let $\Gamma_1, \ldots, \Gamma_L$ be a L-level partition where Γ_i is a refinement of partition Γ_{i-1}. We view this as a rooted tree, where the root is the entire signal constellation and the vertices at level i are the subsets that constitute the partition Γ_i. In this paper we consider only binary partitions, and therefore subsets of partition Γ_i can be labelled by binary strings $a_1, \ldots, a_i$, which specify the path from the root to the specified vertex.

Signal points in QAM constellations are drawn from some realization of the integer lattice $\mathbb{Z}^2$. We focus on the particular realization shown in Figure 3, where the integer lattice has been scaled by $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ to give the lattice $D_2 = \{(a,b) | a, b \in \mathbb{Z}, a+b \equiv 0(\text{mod}2)\}$, and then translated by $(1,0)$.

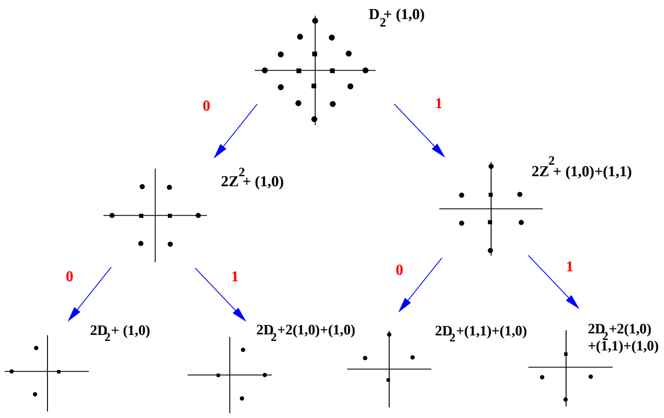


Figure 3: A binary partition of a QAM constellation

The constellation is formed by taking all the points from Λ that fall within a bounding region $\mathcal{R}$. The size of the constellation is proportional to the area of the bounding region, and in Figure 3, the bounding region encloses 16 points.

Binary partitions of QAM constellations are typically based on the following chain of lattices

$$D_2 \supset 2\mathbb{Z}^2 \supset 2D_2 \supset 4\mathbb{Z}^2 \supset \ldots 2^{i-1}D_2 \supset 2^i\mathbb{Z}^2 \supset 2^iD_2 \supset \ldots$$

In Figure 3, the subsets at level 1 are, to within translation, cosets of $2\mathbb{Z}^2$ in D_2 and the subsets at level 2 are cosets of $2D_2$. In general the subsets at level $2i$ are pairs of cosets of 2^iD_2 where the union is a coset of $2^i\mathbb{Z}^2$, and the subsets at level $2i+1$ are pairs of cosets of $2^{i+1}\mathbb{Z}^2$ where the union is a coset of 2^iD_2. Note that implicit in Figure 3 is a binary partition of QPSK, where the points $1, -1, i, -i$ are labelled $00, 01, 11, 10$ respectively. Binary partitions of PSK constellations are described in Section 3.4.

3.4 Algebraic properties of binary partitions

The QAM constellations can be represented through a lattice chain $\Lambda|\Lambda_1|\Lambda_2|\ldots$, where $\Lambda = \mathbb{Z}^2$ is the integer lattice. The lattices in the chain are produced with the generator matrix $\mathbf{G}^k$ where $\mathbf{G} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Given this, we can represent the 2^k-QAM constellation as $\Lambda|\Lambda_k$, *i.e.*, the coset representatives of Λ in Λ_k. The lattice Λ can also be written as the set of Gaussian integers $Z[i] = \{a + bi : a, b \in \mathbb{Z}\}$. Similarly we can write the lattice Λ_k as $\{(a + bi)(1 - i)^k : a, b \in \mathbb{Z}\}$. This decomposition of the QAM constellation is illustrated in Figure 4. Therefore, using this we can represent any point s in a 2^L-QAM constellation using a L-length bit string as

$$s - c(L) \equiv \sum_{l=0}^{L-1} b_l(1 - i)^l \mod (1 - i)^L, \tag{3.9}$$

where we define $f \equiv g \mod (1-i)^L$ if there exist $c, d \in \mathbb{Z}$ such that $f = (c+di)(1-i)^L + g$. Also in (3.9) the constant $c(L) = \frac{1}{2}$ for odd L and $\frac{1}{2}(1+i)$ for even L.

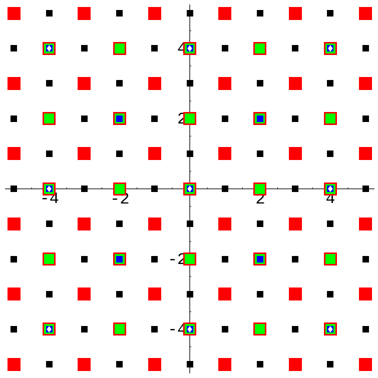


Figure 4: Decomposition for QAM constellations

Binary partitions of PSK constellations are based on a chain of subfields of the cyclotomic field $\mathbb{Q}(\xi_{2^L})$ obtained by adjoining $\xi_{2^L} = \exp(2\pi i/2^L)$ to the rational field $\mathbb{Q}$. Analogous to (3.9), points in the 2^L-PSK constellation can be represented as

$$s = \prod_{l=0}^{L-1} (\xi^{2^l})^{b_l}, \tag{3.10}$$

where $\xi = \xi_{2^L} = \exp(2\pi i/2^L)$ and ξ^{2^L} is a primitive element for $\mathbb{Q}(\xi_{2^{L-l}})$. Note that $1-\xi$ is prime in $\mathbb{Z}[\xi]$ and the quotient $\mathbb{Z}[\xi]/(1-\xi)$ is the field $\mathbb{Z}_2$.

The field $\mathbb{Q}(\xi_{2^L})$ is a degree 2^{L-1} extension of $\mathbb{Q}$. Every rational number is a quotient a/b, where $a, b \in \mathbb{Z}$, and every complex number in $\mathbb{Q}(i)$ is a quotient a/b, where a, b are Gaussian integers. In general every complex number in $\mathbb{Q}(\xi_{2^L})$ is a quotient a/b, where a, b are integer linear combinations of $1, \xi_{2^L}, \ldots, \xi_{2^L}^{2^{L-1}-1}$ and $b \neq 0$. For more details about cyclotomic fields see [41]. Note that $\xi_{2^L}^{2^{L-1}} = -1$, so that $\xi_{2^L}^{j} = -\xi_{2^L}^{2^{L-1}+j}$, for $j = 0, 1, \ldots, 2^{L-1}-1$.

We have a chain of fields

$$\mathbb{Q} = \mathbb{Q}(\xi_2) \subset \mathbb{Q}(i) = \mathbb{Q}(\xi_4) \subset \mathbb{Q}(\xi_8) \ldots \subset \mathbb{Q}(\xi_{2^L}).$$

The algebraic properties collected in the following lemma will be used to establish the performance of the multi-level diversity embedded codes introduces in Section 5.

Lemma 3.1 *Let $\xi = \xi_{2^L} = \exp(2\pi i/2^L)$, then,*

(1) If $\lambda = \sum_{t=0}^{2^{L-1}-1} \xi^t a_t$, where $a_t \in \mathbb{Z}$, then

$$\lambda \equiv \sum_{t=0}^{2^{L-1}-1} a_t \quad (mod\ 2) \tag{3.11}$$

(2) For $r,l \in \mathbb{Z}$*,*

$$\frac{\xi^l - \xi^r}{1-\xi} \equiv r - l \quad mod\ (1-\xi) \tag{3.12}$$

Proof.

(1) Since $\xi^t \equiv 1 \mod (1-\xi)$, for $t = 0, \ldots, 2^{L-1}-1$, we have

$$\lambda \equiv \sum_{t=0}^{2^{L-1}-1} a_t \mod (1-\xi) \tag{3.13}$$

(2) We have,

$$\begin{aligned}
\xi^l - \xi^r &= (1-(1-\xi))^l - (1-(1-\xi))^r \qquad (3.14)\\
&= \sum_{t=0}^{l}(-1)^t \binom{l}{t}(1-\xi)^t - \sum_{t=0}^{r}(-1)^t \binom{r}{t}(1-\xi)^t\\
&\equiv 1 - l(1-\xi) - 1 + r(1-\xi) \mod (1-\xi)^2\\
&\equiv (r-l)(1-\xi) \mod (1-\xi)^2
\end{aligned}$$

4 Linear code constructions

In this Section we restrict our attention to code designs that are linear over the complex field in order to be able to decode them efficiently using a sphere decoder [13]. In Section 4.1 we describe some general properties of the code constructions. Some specific constructions along with their properties are given in Section 4.2.

4.1 Additive code designs

We examine code designs that are additive, in that the transmitted signal is such that

$$\mathbf{X}_{\mathbf{a},\mathbf{b}} = \mathbf{X}_{\mathbf{a}} + \mathbf{X}_{\mathbf{b}}. \tag{4.1}$$

This allows the composite code to be linear in the complex field as well. Given the design criterion in (3.6), it is clear that to ensure that the message set $\mathcal{A}$ gets a diversity level of D_a, a necessary (but *not* sufficient) condition is that

$$\min_{\mathbf{a}_1 \neq \mathbf{a}_2 \in \mathcal{A}} rank(\mathbf{X}_{\mathbf{a}_1} - \mathbf{X}_{\mathbf{a}_2}) \geq D_a/M_r, \tag{4.2}$$

due to the linearity of the code in (4.1). The code design should be such that any perturbation of $\mathbf{x}_a$ due to message set $\mathcal{B}$ does not disturb this rank property.

We can now apply the code design criterion developed in Section 3.2 to the additive codewords model. For illustration, consider the case where we want $D_a = M_r M_t$, *i.e.* maximal diversity. Then we clearly need $(\mathbf{X}_{\mathbf{a}_1} - \mathbf{X}_{\mathbf{a}_2})$ to have full rank for $\mathbf{a}_1 \neq \mathbf{a}_2$. We also need to ensure that $(\mathbf{X}_{\mathbf{a}_1} - \mathbf{X}_{\mathbf{a}_2}) + (\mathbf{X}_{\mathbf{b}_1} - \mathbf{X}_{\mathbf{b}_2})$ is still full rank for *any* $\mathbf{b}_1, \mathbf{b}_2 \in \mathcal{B}$. This translates to the condition that there does not exist a vector $\mathbf{v} \in \mathbb{C}^{1 \times M_t}$ such that

$$\mathbf{v}(\mathbf{X}_{\mathbf{b}_1} - \mathbf{X}_{\mathbf{b}_2})(\mathbf{X}_{\mathbf{a}_1} - \mathbf{X}_{\mathbf{a}_2})^{\dagger} = -\mathbf{v}, \tag{4.3}$$

for all $\mathbf{b}_1, \mathbf{b}_2 \in \mathcal{B}, \mathbf{a}_1 \neq \mathbf{a}_2 \in \mathcal{A}$, where[4] $(\mathbf{Q})^{\dagger} = \mathbf{Q}^*(\mathbf{Q}\mathbf{Q}^*)^{-1}$ denotes the pseudo-inverse of a matrix. In the case of a square design, this translates to a question about existence of certain eigenvalues of the matrix defined in (4.3). Also, for square designs, the condition in (4.3) can be stated in terms of right eigenvectors in $\mathbb{C}^{M_t}$. The condition in (4.3) allows us to eliminate designs that perturb the full rank design for message set $\mathcal{A}$.

Additive codeword designs are amenable to computationally simpler lattice decoding strategies, such as the sphere decoder [13] and appear to be less susceptible to degradation in performance at low to moderate SNR caused by number of nearest neighbors [8]. Moreover, rank conditions may be easier to verify under these constraints. These reasons motivate us to examine constructions based on this model.

4.2 Code Examples

In this section we provide four code constructions that illustrate the design trade-off involved in diversity-embedded codes. We restrict our attention to examples with $M_t = 4$. We also start with a baseline code derived from the rate-$\frac{3}{4}$ orthogonal design based on Octonions [38, 5]. In all these examples, there are no restrictions on which constellations $\mathcal{S}_T$ are used to transmit codeword $\mathbf{X}_{\mathbf{a},\mathbf{b}}$. Therefore the proofs of diversity order are based on symbols from a complex field and do not use properties of specific constellations.

Example 1: Here $\mathcal{A}$ comes from the message set $\{a(0), a(1), a(2)\} \in \mathcal{S}_T$ and $\mathcal{B}$ comes from $b(0) \in \mathcal{S}_T$. This implies that $R_a = \frac{3}{4}\log|\mathcal{S}|$, and $R_b = \frac{1}{4}\log|\mathcal{S}_T|$.

$$\mathbf{X}_{\mathbf{a},\mathbf{b}} = \mathbf{X}_{\mathbf{a}} + \mathbf{X}_{\mathbf{b}} = \begin{bmatrix} a(0) & a(1) & a(2) & b(0) \\ -a^*(1) & a^*(0) & 0 & a(2) \\ -a^*(2) & 0 & a^*(0) & -a(1) \\ 0 & -a^*(2) & a^*(1) & a(0) \end{bmatrix}. \tag{4.4}$$

This code can be shown to achieve full diversity of $4M_r$ for the message set $\mathcal{A}$ and diversity M_r for message set $\mathcal{B}$.

The proof of this is simple and we illustrate it in order to demonstrate how the criterion in (3.6) can be applied. Let us denote $\tilde{\mathbf{a}} = \mathbf{a}_1 - \mathbf{a}_2$ and $\tilde{\mathbf{b}} = \mathbf{b}_1 - \mathbf{b}_2$. Since this is a square design, and $\mathbf{X}_{\mathbf{a}}$ is a unitary matrix, the condition given in (4.3) translates to

$$\nexists \mathbf{v} \text{ such that for } \tilde{\mathbf{a}} \neq \mathbf{0} \quad \mathbf{X}_{\tilde{\mathbf{b}}}\mathbf{X}^*_{\tilde{\mathbf{a}}}\mathbf{v} = -||\tilde{\mathbf{a}}||^2\mathbf{v}, \tag{4.5}$$

[4]For a matrix $\mathbf{X}$, we use the notation X^* to denote the Hermitian transpose of the matrix, which for scalars is just complex conjugation.

for all $\tilde{\mathbf{b}}$, where $||\tilde{\mathbf{a}}||^2 = |a(0)|^2 + |a(1)|^2 + |a(2)|^2 \neq 0$. The condition in (4.5) means that $\mathbf{X}_{\tilde{\mathbf{b}}}\mathbf{X}^*_{\tilde{\mathbf{a}}}$ should not have a negative real eigenvalue. It can easily be shown that for this design all the eigenvalues of this matrix are zero and hence the condition is met. At first glance, this design might seem uninteresting since we know that according to Theorem 2.2 (see [39]) there exist codes with rate $\log|\mathcal{S}_T|$ and diversity $M_r M_t$. However, we do not know of codes linear in the complex field, which are not designed for specific signal constellations, and have *no* constellation expansion, with such a property.

■

Example 2: Here $\mathcal{A}$ comes from the message set $\{a(0), a(1)\} \in \mathcal{S}_T$ and $\mathcal{B}$ comes from $\{b(0), b(1), b(2), b(3)\} \in \mathcal{S}_T$. This implies that $R_a = \frac{1}{2}\log|\mathcal{S}_T|$, and $R_b = \log|\mathcal{S}_T|$, leading to a total rate of $R_a + R_b = \frac{3}{2}\log|\mathcal{S}_T|$.

$$\mathbf{X}_{\mathbf{a},\mathbf{b}} = \mathbf{X}_{\mathbf{a}} + \mathbf{X}_{\mathbf{b}} = \begin{bmatrix} a(0) & a(1) & 0 & 0 \\ -a^*(1) & a^*(0) & 0 & 0 \\ b(0) & b(1) & a^*(0) & -a(1) \\ b(2) & b(3) & a^*(1) & a(0) \end{bmatrix}. \tag{4.6}$$

This code can be shown to achieve full diversity of $4M_r$ for the message set $\mathcal{A}$ and diversity M_r for message set $\mathcal{B}$. Therefore, this example achieves the tuple, $(\frac{1}{2}\log|\mathcal{S}|, 4M_r, \log|\mathcal{S}_T|, M_r)$.

The proof of this claim is quite simple and follows a similar argument as in Example 1. In this example the total rate is $\frac{3}{2}\log|\mathcal{S}_T|$, with a composite diversity of M_r, however it has a code of maximal diversity and rate $\frac{1}{2}\log|\mathcal{S}_T|$ embedded within it. Another question is whether such a code can be constructed by a TDM strategy. This is not possible for the following reason. A full-diversity code needs to occupy at least M_t symbols, and hence for a square design $T = M_t$. This implies that we cannot construct a switching scheme that has a full diversity code and has a rate beyond $\log|\mathcal{S}_T|$.

■

Example 3: Here $\mathcal{A}$ comes from the message set $\{a(0), a(1)\} \in \mathcal{S}_T$ and the message set $\mathcal{B}$ is given by $\{b(0), b(1), b(2), b(3)\} \in \mathcal{S}_T$. This implies that $R_a = \frac{1}{2}\log|\mathcal{S}_T|$, and $R_b = \log|\mathcal{S}_T|$, identical to Example 2.

$$\mathbf{X}_{\mathbf{a},\mathbf{b}} = \mathbf{X}_{\mathbf{a}} + \mathbf{X}_{\mathbf{b}} = \begin{bmatrix} a(0) & a(1) & b(2) & b(3) \\ -a^*(1) & a^*(0) & b^*(3) & -b^*(2) \\ b(0) & b(1) & a^*(0) & -a(1) \\ -b^*(1) & b(0) & a^*(1) & a(0) \end{bmatrix}. \tag{4.7}$$

This code can be shown to achieve diversity of $3M_r$ for the message set $\mathcal{A}$ and diversity $2M_r$ for message set $\mathcal{B}$. Therefore, this example achieves the tuple, $(\frac{1}{2}\log|\mathcal{S}|, 3M_r, \log|\mathcal{S}_T|, 2M_r)$.

The proof of this involves a set closely related to the quaternions (used in the Alamouti code [1]). For $a(0), a(1) \in \mathbb{C}$, define the following sets

$$\mathcal{Q} = \begin{bmatrix} a(0) & a(1) \\ -a^*(1) & a^*(0) \end{bmatrix}, \quad \tilde{\mathcal{Q}} = \begin{bmatrix} a(0) & a(1) \\ a^*(1) & -a^*(0) \end{bmatrix}. \tag{4.8}$$

The set $\mathcal{Q}$ is a division ring (called the Hamilton's biquaternionic group [5]), and $\tilde{\mathcal{Q}}$ is *not* closed under multiplicatiion. However, any $\mathbf{F} \in \tilde{\mathcal{Q}}$, can be written as $\mathbf{F} = \mathbf{DE}$, for some $\mathbf{E} \in \mathcal{Q}$, and $\mathbf{D} = \mathrm{diag}(1, -1)$. Note that the diagonal matrix $\mathbf{D}$ satisfies $\mathbf{DED} = \mathbf{E}'$, for some $\mathbf{E}, \mathbf{E}' \in \mathcal{Q}$. Also, the only matrix in $\mathcal{Q} \cap \tilde{\mathcal{Q}}$ is the zero matrix. Now, it can be easily seen that the following properties hold,

$$\begin{aligned} \mathbf{F} \in \tilde{\mathcal{Q}}, &\quad \mathbf{F}^{-1} \in \tilde{\mathcal{Q}} \\ \mathbf{E}, \mathbf{F} \in \tilde{\mathcal{Q}}, &\quad \mathbf{EF} \in \mathcal{Q}, \\ \mathbf{E} \in \mathcal{Q}, \mathbf{F} \in \tilde{\mathcal{Q}}, &\quad \mathbf{EF} \in \tilde{\mathcal{Q}} \\ \mathbf{E} \in \tilde{\mathcal{Q}}, \mathbf{F} \in \mathcal{Q}, &\quad \mathbf{EF} \in \tilde{\mathcal{Q}}. \end{aligned} \tag{4.9}$$

Given two distinct candidate message sets $\{\mathbf{a}_1, \mathbf{b}_1\}$ and $\{\mathbf{a}_2, \mathbf{b}_2\}$, let us denote $\tilde{\mathbf{a}} = \mathbf{a}_1 - \mathbf{a}_2$ and $\tilde{\mathbf{b}} = \mathbf{b}_1 - \mathbf{b}_2$. To show that $D_b = 2M_r$, we need to show that if $\tilde{\mathbf{b}} \neq \mathbf{0}$, then $\min\limits_{\tilde{\mathbf{b}} \neq \mathbf{0}} \min\limits_{\tilde{\mathbf{a}}} \mathrm{rank}(\mathbf{X}_{\tilde{\mathbf{a}}, \tilde{\mathbf{b}}}) = 2$. Observe that $D_b \geq 2M_r$ since the rows of $\begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_1 \end{bmatrix}$ are orthogonal. By choosing $\tilde{b}(0) = \tilde{b}(1) = \tilde{a}(0) = \tilde{a}(1) = 0$, we see that $D_b \leq 2M_r$, and hence $D_b = 2M_r$. Similarly, to show that $D_a = 3M_r$, we need to show that if $\tilde{\mathbf{a}} \neq \mathbf{0}$, then $\min\limits_{\tilde{\mathbf{a}} \neq \mathbf{0}} \min\limits_{\tilde{\mathbf{b}}} \mathrm{rank}(\mathbf{X}_{\tilde{\mathbf{a}}, \tilde{\mathbf{b}}}) = 3$. We can therefore re-write (4.7) as,

$$\mathbf{X}_{\mathbf{a}_1, \mathbf{b}_1} - \mathbf{X}_{\mathbf{a}_2, \mathbf{b}_2} = \mathbf{X}_{\tilde{\mathbf{a}}, \tilde{\mathbf{b}}} = \begin{bmatrix} \mathbf{E}[\tilde{a}(0), \tilde{a}(1)] & \mathbf{F}[\tilde{b}(2), \tilde{b}(3)] \\ \mathbf{E}[\tilde{b}(0), \tilde{b}(1)] & \mathbf{E}[\tilde{a}^*(0), -\tilde{a}(1)] \end{bmatrix} \stackrel{def}{=} \begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_1 \\ \mathbf{B}_2 & \mathbf{A}_2 \end{bmatrix}, \tag{4.10}$$

where we have indexed elements of the sets $\mathbf{E} \in \mathcal{Q}$ and $\mathbf{F} \in \tilde{\mathcal{Q}}$ in terms of the indeterminates $\{u, v\}$ as $\mathbf{E}[u, v]$ and $\mathbf{F}[u, v]$. Now, by doing the block triangularization we get

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{B}_2\mathbf{A}_1^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_1 \\ \mathbf{B}_2 & \mathbf{A}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_1 \\ \mathbf{0} & \mathbf{A}_2 - \mathbf{B}_2\mathbf{A}_1^{-1}\mathbf{B}_1 \end{bmatrix}. \tag{4.11}$$

Since $\mathbf{A}_2 \in \mathcal{Q}$ and $\mathbf{B}_2\mathbf{A}_1^{-1}\mathbf{B}_1 \in \tilde{\mathcal{Q}}$, it follows from (4.9) that $\mathbf{A}_2 \neq \mathbf{B}_2\mathbf{A}_1^{-1}\mathbf{B}_1$ and hence $D_a \geq 3M_r$. By choosing $\tilde{b}(0) = -\tilde{a}^*(1), \tilde{b}(1) = \tilde{a}^*(0), \tilde{b}(2) = \tilde{a}^*(1), \tilde{b}(3) = \tilde{a}(0)$, we see that $D_a \leq 3M_r$, showing that $D_a = 3M_r$.

The contrast with Example 2 is that although the rates for each layer are identical, here we achieve a higher diversity for the message set $\mathcal{B}$ but at the cost of a lower diversity for the message set $\mathcal{A}$. In this example as well, we cannot construct a TDM code that achieves this. If we need a code of diversity 3, it needs to occupy at least 3 time symbols. Hence for block size $T = M_t = 4$, we are left with 1 symbol in which to obtain diversity two and rate $\log|\mathcal{S}_T|$, which is clearly not possible.

■

The next example illustrates the case of a non-square design.

Example 4: In this example, $M_t = 4$ and the block size is $T = 5$ symbols, giving rise to a non-square code matrix design. Here $\mathcal{A}$ comes from the message set $\{a(0), a(1)\} \in \mathcal{S}$ and $\mathcal{B}$ comes from $\{b(0), b(1), b(2), b(3)\} \in \mathcal{S}$. This implies that $R_a = \frac{2}{5}\log|\mathcal{S}_T|$, and $R_b = \frac{4}{5}\log|\mathcal{S}_T|$, leading to a total rate of $R_a + R_b = \frac{6}{5}\log|\mathcal{S}_T|$.

$$\mathbf{X}_{\mathbf{a}, \mathbf{b}} = \mathbf{X}_{\mathbf{a}} + \mathbf{X}_{\mathbf{b}} = \begin{bmatrix} a(0) & -a^*(1) & b(0) & b(2) & -b^*(3) \\ a(1) & a^*(0) & b(1) & b(3) & b^*(2) \\ 0 & 0 & a^*(0) & a^*(1) & -b^*(1) \\ 0 & 0 & -a(1) & a(0) & b^*(0) \end{bmatrix}. \tag{4.12}$$

This code can be shown to achieve diversity of $4M_r$ for the message set $\mathcal{A}$ and diversity $2M_r$ for message set $\mathcal{B}$. Therefore, this example achieves the tuple, $(\frac{2}{5}\log|\mathcal{S}_T|, 4M_r, \frac{4}{5}\log|\mathcal{S}|, 2M_r)$.

Given the construction from Example 2, we can easily prove that for $\tilde{\mathbf{a}} \neq \mathbf{0}$, $\mathrm{rank}(\mathbf{X}_{\tilde{\mathbf{a}},\tilde{\mathbf{b}}}) = 4$, and hence for message set $\mathcal{A}$ we get a diversity order of $4M_r$. To prove the claimed diversity order for message set $\mathcal{B}$, we need to show the following

$$\min_{\substack{\tilde{\mathbf{a}} \\ \tilde{\mathbf{b}}\neq\mathbf{0}}} \mathrm{rank}(\mathbf{X}_{\tilde{\mathbf{a}},\tilde{\mathbf{b}}}) = 2. \tag{4.13}$$

Clearly, for $\tilde{a}(0) = \tilde{a}(1) = \tilde{b}(0) = \tilde{b}(1) = 0$, $\mathrm{rank}(\mathbf{X}_{\tilde{\mathbf{a}},\tilde{\mathbf{b}}}) = 2$ and hence the minimal rank is at most 2. We prove the result by contradiction. Suppose there exists $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}} \neq \mathbf{0}$ such that $\mathrm{rank}(\mathbf{X}_{\tilde{\mathbf{a}},\tilde{\mathbf{b}}}) = 1$. This implies that for such a choice, all the rows are multiples of one another, *i.e.*,

$$\begin{bmatrix} \tilde{a}(0) \\ -\tilde{a}^*(1) \\ \tilde{b}(0) \\ \tilde{b}(2) \\ -\tilde{b}^*(3) \end{bmatrix} = \alpha \begin{bmatrix} \tilde{a}(1) \\ \tilde{a}^*(0) \\ \tilde{b}(1) \\ \tilde{b}(3) \\ \tilde{b}^*(2) \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \\ \tilde{a}^*(0) \\ \tilde{a}^*(1) \\ -\tilde{b}^*(1) \end{bmatrix} = \beta \begin{bmatrix} 0 \\ 0 \\ -\tilde{a}(1) \\ \tilde{a}(0) \\ \tilde{b}^*(0) \end{bmatrix}, \quad \begin{bmatrix} \tilde{a}(1) \\ \tilde{a}^*(0) \\ \tilde{b}(1) \\ \tilde{b}(3) \\ \tilde{b}^*(2) \end{bmatrix} = \gamma \begin{bmatrix} 0 \\ 0 \\ \tilde{a}^*(0) \\ \tilde{a}^*(1) \\ -\tilde{b}^*(1) \end{bmatrix}$$

For the above to be true, we would need $\tilde{a}(0) = 0 = \tilde{a}(1)$ and $\tilde{b}(0) = \tilde{b}(1) = \tilde{b}(2) = \tilde{b}(3) = 0$, which results in a contradiction. Hence we get the result claimed in (4.13), giving $D_b = 2M_r$.

This code also cannot be constructed using a switching strategy. A full-diversity code needs to occupy at least M_t symbols, and hence the message set $\mathcal{A}$ needs to occupy $M_t = 4$ symbols. Hence for block size $T = 5$, we are left with 1 symbol in which to obtain diversity two and rate $\frac{4}{5}\log|\mathcal{S}_T|$, which is clearly not possible. ∎

5 Multi-level constructions

In Section 4 we described embedded diversity codes that are linear over the complex field. In this section, we develop a second design method based on multi-level (non-linear) space-time codes that are matched to a binary partition of a QAM or PSK constellation. We will give the construction and defer the proofs of code performance to the Appendix.

Given an L-level binary partition of a QAM or PSK signal constellation, a space-time codeword is an array $\mathbf{K} = \{\mathbf{K}_1, \mathbf{K}_2, \ldots, \mathbf{K}_L\}$ determined by a sequence of binary matrices, where matrix, $\mathbf{K}_i$ specifies the space-time array at level i. A *multi-level space-time code* is defined by the choice of the constituent sets of binary matrices $\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_L$. For $i = 1, \ldots, L$ the binary matrix $\mathbf{K}_i$ is required to be in the set $\mathcal{K}_i$.

El-Gamal and Hammons [26] showed that it is possible to design codes with a given level of diversity in the complex domain by placing constraints on the constituent sets of binary matrices $\{\mathcal{K}_i\}$, and their approach has been taken up by many authors (see Lu and Kumar [29] and the references there). Instead of coding across all levels at once, we develop the theory from the perspective of multi-level codes [27, 31, 4], in order to design the diversity embedding property.

Set-partitioning was developed by Ungerboeck [42], as the basis of code design for the AWGN channel where it provides a lower bound on the squared Euclidean distance between signals the depends only on the binary sum of signal labels. The importance of set-partitioning to code design for the wireless channel is that it provides a mechanism for translating constraints in the binary domain into lower bounds on the diversity protection in the complex domain.

Given two message sets $\mathcal{A}, \mathcal{B}$, they are mapped to the space-time codeword $\mathbf{X}$ as shown below.

$$\mathcal{A}, \mathcal{B} \xrightarrow{f_1} \mathbf{K} = \begin{bmatrix} K(1,1) & \dots & K(1,T) \\ \vdots & \vdots & \vdots \\ K(M_t,1) & \dots & K(M_t,T) \end{bmatrix} \xrightarrow{f_2} \mathbf{X} = \begin{bmatrix} x(1,1) & \dots & x(1,T) \\ \vdots & \vdots & \vdots \\ x(M_t,1) & \dots & x(M_t,T) \end{bmatrix}$$

where $K(m,n) \in \{0,1\}^{\log(|\mathcal{S}_T|)}$ *i.e.,* binary string and $x(m,n) \in \mathcal{S}_T$. This construction is illustrated in Figure 5 for a constellation size of L bits. The basic idea is that we specify sets $\mathcal{K}_1, \dots, \mathcal{K}_L$ from which the sequence of binary matrices which encode the constellations are chosen. These sets of binary matrices provide rank guarantees necessary to achieve the diversity orders required for each message set. For example, if we only require a single diversity order (*i.e.,* no diversity embedding) then we can choose all the sets of binary matrices to be identical. At the other extreme all the sets could be different, yielding L different levels of diversity embedding. Given the message set, we first choose the matrices $\mathbf{K}_1, \dots . \mathbf{K}_L$. The first mapping f_1 is obtained by taking matrices and constructing the matrix $\mathbf{K} \in \mathbb{C}^{M_t \times T}$ each of whose entries is constructed by concatenating the bits from the corresponding entries in the matrices $\mathbf{K}_1, \dots . \mathbf{K}_L$ into L-length bit-string. This matrix is then mapped to the space-time codeword through a constellation mapper f_2, for example the set-partition mapping given in Section 3.3.

As shown above, this structure can be used for one to L levels of diversity order. However, in this paper, we restrict our attention to two levels of diversity (as shown in Figure 1). For simplicity consider a 4-QAM constellation, with 2 levels of diversity order, D_a, D_b. Given that $L = 2$, we then assign layer 1 to diversity order D_a and layer 2 to D_b with $D_a \geq D_b$. We choose sets of binary matrices $\mathcal{K}_1, \mathcal{K}_2$ with rank guarantees $D_a/M_r, D_b/M_r$ respectively. Let the set sizes be $|\mathcal{K}_1| = 2^{TR_a}, |\mathcal{K}_2| = 2^{TR_b}$, yielding the appropriate rates R_a, R_b. Therefore, given a message $m_a \in \mathcal{A}, m_b \in \mathcal{B}$, we choose the matrices $\mathbf{K}_1 \in \mathcal{K}_1, \mathbf{K}_2 \in \mathcal{K}_2$ corresponding to the messages m_a and m_b respectively. Given $\mathbf{K}_1, \mathbf{K}_2$, we construct the space-time code $\mathbf{X}$ as illustrated in Figure 5. If we have constellations of size 2^L with $L > 2$ and we still need 2 levels of diversity, we can assign layers $1, \dots, L_a$ to choose matrices with the *same* binary set $\mathcal{K}_1$ with rank guarantee D_a/M_r and layers, $L_a + 1, \dots, L$ to choose matrices from the binary set $\mathcal{K}_2$ with rank guarantee D_b/M_r. We will see shortly that we can choose set cardinalities as $|\mathcal{K}_1| = 2^{T\frac{R_a}{L_a}}, |\mathcal{K}_2| = 2^{T\frac{R_b}{L_b}}$, where $L_b = L - L_1$. We get the corresponding rates R_a, R_b for the two diversity orders. Therefore, as before, given message $m_a \in \mathcal{A}, m_b \in \mathcal{B}$, we choose the sequence of matrices $\mathbf{K}_1, \dots, \mathbf{K}_{L_a} \in \mathcal{K}_1$ based on m_a and matrices $\mathbf{K}_{L_a+1}, \dots, \mathbf{K}_L \in \mathcal{K}_2$ based on m_b. Using this sequence of L matrices, we obtain the space-time codeword as seen in Figure 5.

In all this, the choice of the sets $\mathcal{K}_l, l = 1, \dots, L$ has been unspecified. A natural choice for these is

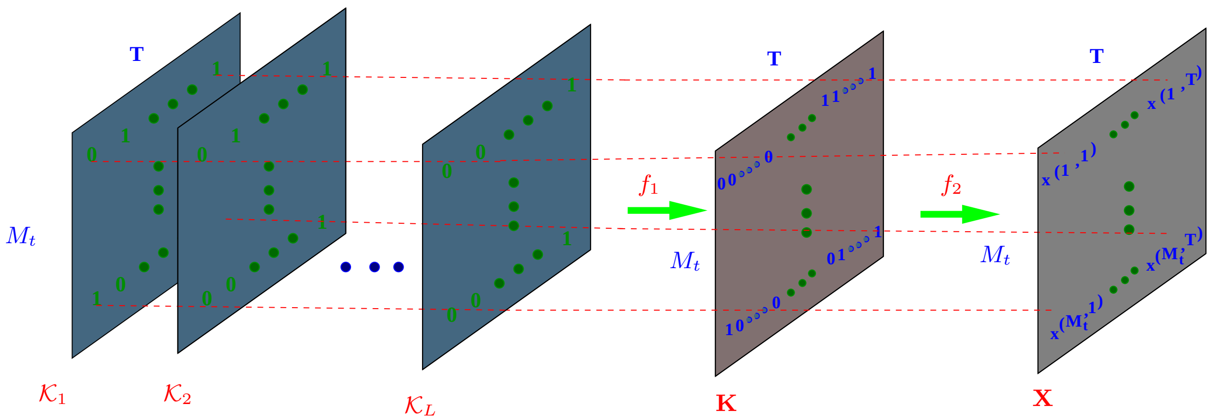


Figure 5: Schematic representation of the multi-level code construction.

a set $\mathcal{K}$ of binary $M_t \times M_t$ matrices such that for any distinct pair of matrices $\mathbf{A}, \mathbf{B} \in \mathcal{K}$ the rank of $\mathbf{A} - \mathbf{B}$ is at least $M_t - d$. The size of $\mathcal{K}$ is at most $2^{(d+1)M_t}$ since the first $d+1$ rows of $\mathbf{A}$ and $\mathbf{B}$ must be distinct, and there is a classical example [23] that achieves the bound. Let $f : x \to x^2$ be the Frobenius map on $GF(2^{M_t})$ and let

$$\mathcal{K} = \{\alpha_0 + \alpha_1 f + \ldots + \alpha_d f^d | \alpha_i \in GF(2^{M_t})\}.$$

Then $\mathcal{K}$ is closed under binary addition, and given a choice of basis for $GF(2^N)$ over $GF(2)$, each map in $\mathcal{K}$ may be viewed as an $M_t \times M_t$ binary matrix. There are at most 2^d elements of the null space of $\mathbf{A} - \mathbf{B}$, since each element is the root of a polynomial

$$\gamma_0 x + \gamma_1 x^2 + \ldots + \gamma_d x^{2^d}$$

with degree at most 2^d. Hence the rank of $\mathbf{A} - \mathbf{B}$ is at least $M_t - d$. Clearly, this construction can be easily extended to sets of $M_t \times T$ rectangular binary matrices for $T \geq M_t$ such that the difference of any two matrices in the set has rank $\lfloor M_t - r \rfloor$ over the binary field. Moreover, such sets have a cardinality of $|\mathcal{P}(M_t, T, r)| = 2^{T(r+1)}$.

These matrices yield a rate of $r + 1$ bits/transmission. In our construction we use these matrices, with $D_l = \lfloor M_t - r_l \rfloor$, where $D_1 \geq D_2 \geq \ldots \geq D_L$. This yields a rate of $R_l = r_l + 1$ in each layer. In the appendix we will show that this construction for QAM constellations achieves the rate tuple $(R_1, M_r D_1, \ldots, R_L, M_r D_L)$, with the overall equivalent single layer code achieving rate-diversity point, $(\sum_l R_l, M_r D_L)$. As described above, we can have the desired number of layers by choosing several identical diversity/rate layers.

Optimal decoding employs a maximum-likelihood decoder which jointly decodes the message sets. This is the decoder for which the performance is summarized in Theorem 5.1 and the proof is given in the Appendix. Note that one can also implement a successive cancellation (multi-stage) decoder by decoding each layer successively and incorporating the decision into the next stage just like it is done in classical multi-level codes [27, 4, 31]. Staged decoding reduces signal processing complexity at the expense of introducing additional nearest neighbors at each decoding stage, which reduces the performance.

Theorem 5.1 *Let $\mathcal{C}$ be a multi-level space-time code for a QAM or M-PSK constellation of size 2^L with M_t transmit antennas that is determined by constituent sets of binary matrices $\mathcal{K}_l, l = 1, \ldots, L$ with binary rank gurantees $D_1 \geq D_2 \ldots \geq D_L$. The input bits that select the codeword from the ith matrix $\mathcal{K}_i$ are guaranteed diversity D_i in the complex domain.*

This result shows that one can construct multi-level diversity embedded space-time codes where different information streams achieve different diversity levels. Another possible application of codes based on nested families of sets is balancing of the diversity gain and coding gain within a particular space-time code. The key observation here is that intra-subset distances increases with depth in the partition chain, so that less diversity may be required to achieve a given error rate at lower levels in the binary partition.

6 Diversity embedded trellis codes

The construction of diversity embedded trellis codes is quite similar to multi-level diversity embedded block codes. The idea is to construct the binary matrices with rank guarantees using binary convolutional codes. Using these sets of matrices the diversity embedding properties are ensured by appropriately choosing the underlying convolutional codes.

Consider transmission using a constellation size of 2^L. We choose sets $\mathcal{V}_1, \ldots, \mathcal{V}_L$, from which the sequence of binary matrices which encode the constellations are chosen, such that they have given rank guarantees. These guarantees reflect the ultimate diversity orders required for each message set. These $\mathcal{V}_i$ are chosen from specially constructed binary convolutional codes. Such constructions are possible and therefore we can construct diversity embedded trellis codes using such underlying convolutional codes. We describe these constructions in the next section and refer the reader to [29] for proofs of properties of rank binary convolutional codes.

We illustrate the idea by examining the construction for each of the layers. The construction is shown in Figure 6. Given the input stream for each layer i, the first block in the figure maps these inputs to the coefficients of R_i polynomials $u_{i,j}(D), j = 1, \ldots, R_i$ in $\mathbb{F}_2[D]$. The second block multiplies the input vector $\{u_{i,j}(D)\}_{j=1}^{R_i}$ by the generator matrix $\mathbf{G}_i(D)$, where the largest degree of a polynomial in $\mathbf{G}_i(D)$ is atleast M_t, and generates a vector $\mathbf{v}_i(D)$ of polynomials. The final block Ω then maps this vector $\mathbf{v}$ to a binary matrix $\mathbf{V}_i \in \mathbb{F}_2^{M_t \times T}$.

We define the set $\mathcal{V}_i$ to be the set of all output matrices $\mathbf{V}_i$ for all possible inputs on the stream i. *Note: These sets satisfy the rank property that the difference of any two matrices in a set $\mathcal{V}_i$ has rank at least D_i.*

Since we have the sets $\mathcal{V}_1, \ldots, \mathcal{V}_L$, given any L input bit streams the construction is similar to the constructions seen earlier and illustrated in Figure 5. Therefore, in a manner similar to the proof of Theorem 5.1 we can state the following result.

Theorem 6.1 *Let $\mathcal{C}_{trellis}$ be a multi-level space-time trellis code for a QAM or M-PSK constellation of size 2^L with M_t transmit antennas that is determined by constituent convolutional codes $\mathbf{G}_l(D), l = 1, \ldots, L$ with binary rank gurantees $D_1 \geq D_2 \ldots \geq D_L$. Any message, $\mathbf{u}_i$, is guaranteed diversity D_i in the complex domain.*

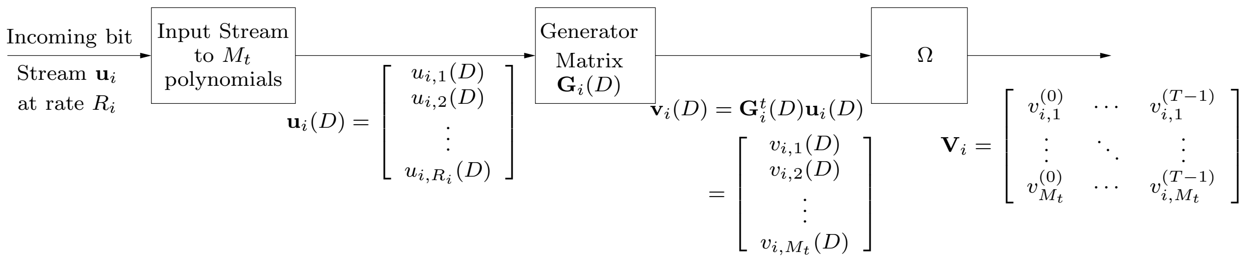


Figure 6: Binary matrices for each layer

6.1 Binary convolutional codes

We will now investigate the known schemes for construction of such sets of binary matrices. Explicit construction of full diversity maximum rate binary convolutional codes was first shown in [26]. This was extended for general points on the rate diversity tradeoff in [29]. We reproduce the construction over here for completeness.

Consider the construction for a particular layer above. We will see the construction of rate R symbols/Tx, and diversity $d = M_t - R + 1$ binary codes. This is basically a R input and M_t output convolutional codes.

Represent the generator matrix or transfer function matrix $\mathbf{G}$ for this code by a $R \times M_t$ generator matrix given by,

$$\mathbf{G} = \begin{bmatrix} g_1^{(1)}(D) & g_1^{(2)}(D) & \cdots & g_1^{(M_t)}(D) \\ g_2^{(1)}(D) & g_2^{(2)}(D) & \cdots & g_2^{(M_t)}(D) \\ \vdots & \ddots & \ddots & \vdots \\ g_R^{(1)}(D) & g_R^{(2)}(D) & \cdots & g_R^{(M_t)}(D) \end{bmatrix} \quad (mod\ f(D)) \tag{6.1}$$

where $g_u^{(v)}(D)$ represents the transfer function from the input u to the transmit antenna v and $f(D)$ is a primitive polynomial of degree E such that $E \geq M_t$. We use the construction suggested in [29] and choose $g_u^{(v)}(D) = D^{(v-1)\cdot 2^{(u-1)}}$.

The input message polynomial is represented by the vector of message polynomial,

$$\mathbf{u}(D) = \begin{bmatrix} u_1(D) & u_2(D) & \vdots & u_R(D) \end{bmatrix}^T \tag{6.2}$$

where $u_i(D) \in \mathbb{F}_2[D]$.

The code polynomial vector is given by,

$$\mathbf{v}(D) = \mathbf{G}^t(D)\mathbf{u}(D) \tag{6.3}$$

$$= \begin{bmatrix} v_1(D) & v_2(D) & \vdots & v_{M_t}(D) \end{bmatrix}^T \tag{6.4}$$

Since we require that the degree of $v_i(D) \leq (T-1)$ we have that $deg(u_i(D)) \leq (T-1) - (E-1)$ i.e. $deg(u_i(D)) \leq (T-E)$. *Note that this convolutional code corresponds to a rate of,*

$$\frac{\log\left(2^{(T-E)+1} \cdot 2^R\right)}{T} = \frac{R\,(T-E+1)}{T} \quad bits/Tx \tag{6.5}$$

which tends to R as $T \to \infty$.

The $M_t \times T$ code matrix which is actually transmitted on the antenna is given by,

$$\mathbf{V} = \begin{bmatrix} v_1^0 & \cdots & v_1^{T-1} \\ \vdots & \ddots & \vdots \\ v_{M_t}^0 & \cdots & v_{M_t}^{T-1} \end{bmatrix} \tag{6.6}$$

where v_i^j is the j^{th} coefficient of the polynomial v_i in (6.4). We make a distinction between $\mathbf{v}(D)$ which is a vector of polynomials in D and $\mathbf{V}$ which is a binary matrix. This mapping is denoted by Ω *i.e.* $\Omega(\mathbf{v}(D)) = \mathbf{V}$

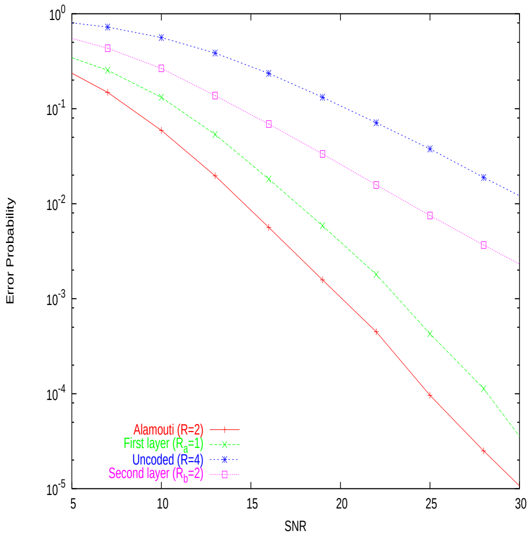
Since we consider a linear code it suffices to prove that for any non-zero input $\mathbf{u}(D)$, the output matrix $\mathbf{V}$ has rank $\geq d$. With these parameters the rank guarantees on the matrices were proved in [29]. In particular we state the following theorem without proof.

Theorem 6.2 *For all full rank matrices* $\mathbf{B}$ *of size* $R \times M_t$ *we have that* $det\left(\mathbf{B}\mathbf{G}^t(D)\right) \neq 0$*, where the determinant is taken over the vector space* $\mathbb{F}_2(D)$*.*

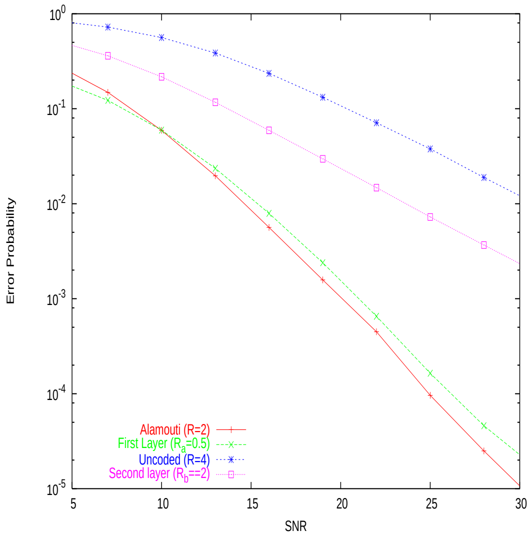
7 System implications of diversity embedding

Given that we can construct diversity embedded codes, we would like to examine how such codes can impact the wireless communication system design. We can see three immediate applications of diversity embedded codes: (i) A natural application would be for applications requiring unequal error protection (UEP). For example, image, audio or video transmission might need multiple levels of error protection for sensitive and less sensitive parts of the message. (ii) A second application could be to improve the overall throughput by opportunistically using the good channel realizations without channel state feedback. (iii) A third application could be in reducing delay in packet transmission using the different diversity orders for prioritized scheduling. In this section, we examine the impact of diversity embedding for each of these applications by comparing it to conventional single layer codes. In all the numerical results we have $M_r = 1$. Since all the constructions in this paper are for the *fixed* transmit alphabet, we make all our comparisons with the single-layer code for the *same* fixed transmit alphabet. We can also do other comparisons, such as those based on total rate, or matching the rate on particular diversity layers, and consequently use different transmit alphabets for the single-layer and diversity-embedded codes. We do not examine this here, since we believe that it would be an unfair comparison, even though such a comparison would be beneficial for the diversity embedded codes.

In Figure 7, we give the performance of a diversity embedded code which is designed for $M_t = 2$ and 4-QAM signal constellation. In Figure 7(a), the embedded code has $R_a = 1$ bit/transmission at diversity order $D_a = 2$ and $R_b = 2$ bits/transmission at diversity order $D_b = 1$. We compare this with a "full-rate", maximal diversity order code (Alamouti code with $R = 2$ bits/transmission and diversity order $D = 2$). We also plot the performance of uncoded transmission with rate $R = 4$ bits/transmission and diversity order $D = 1$. Qualitatively, we can see that the embedded code gives two levels of diversity. Note that as expected, we do pay a penalty in rate (or error



(a)



(b)

Figure 7: UEP performance of diversity embedded codes.

performance) over a single-layer code designed for the specific diversity order, but the penalty can be made smaller by cutting the rate for one of the diversity layers as demonstrated in Figure 7(b).

In Figure 8, we show the information-theoretic bounds for outage for a superposition diversity embedded code as studied in [19]. We plot the outage probability of the superposition diversity embedded code for a *fixed* transmission rates. We have two superposed layers with rates, 1 bit per transmission for the higher diversity layer and 2 bits per transmission for the lower diversity layer. As can be seen, in Figure 8 the superposition ratio has been chosen so that the outage probability has diversity orders are approximately 2 and 1 respectively for the higher and lower diversity layers. Even though our constructions were for the fixed alphabet constraint and not for the information-theoretic criterion, we see that the performance of the diversity embedded codes constructed in this paper are close to the information-theoretic bounds. In particular if we compare Figure 7 (a) with Figure 8, we see that for outage probabilities of 10^{-3} and 10^{-1} for the higher and lower diversity layers, the multi-level constructions are approximately 2.5 dB away from the information-theoretic bounds.

In Figure 9, we illustrate the advantage of diversity embedded codes in terms of opportunistically utilizing the channel conditions without feedback. In Figure 9(a), we see that for 4-QAM, and $M_t = 2$, the diversity embedded code outperforms the Alamouti single-layer code designed for full diversity. However, at very high SNR, single-layer transmission designed at the lower diversity order outperforms a diversity embedded code since it transmits at a higher rate. This illustrates that for moderate SNR regimes, diversity embedded codes outperform single layer codes in terms of average throughput. This is illustrated using transmission of packets of size 240 bits over a quasi-static channel and the average throughput is measured over a sequence of such packets. Figure 9(b) for 8-QAM shows that this regime increases with constellation size.

In the final application, we examine the delay behavior if we combine a rudimentary ACK/NACK feedback about the transmitted information along with space-time codes. In the single-layer code

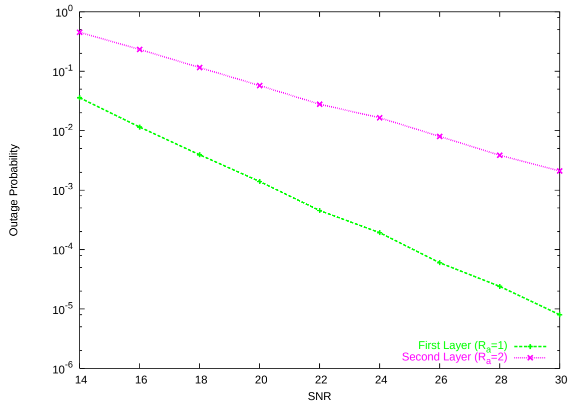
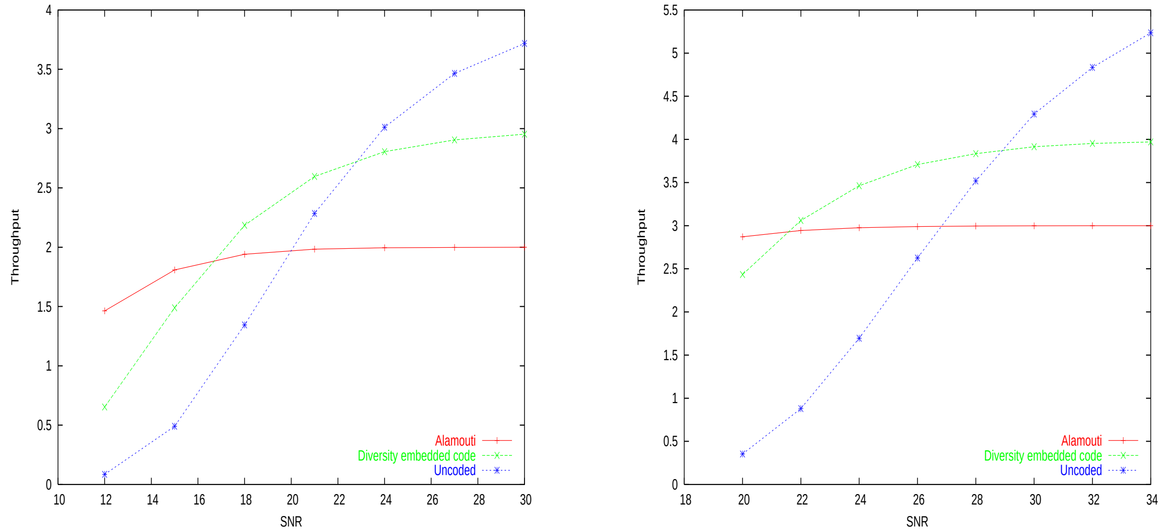


Figure 8: Outage performance of diversity embedded superposition codes studied in [19].



(a) 4-QAM

(b) 8-QAM

Figure 9: Throughput comparison of diversity embedded codes with single layer codes.

the traditional ARQ protocol is used wherein if a packet is in error, it is re-transmitted. In a diversity-embedded code, since different parts of the information get unequal error protection, we can envisage an alternative use of the ARQ. For two diversity levels, we assume that ACK/NACK is received separately for each diversity layer. The proposed mechanism is illustrated in Figure 10. The information is sent along two streams, one on the higher diversity level and the other on the lower diversity level. If the packet on the higher diversity level goes through but the lower diversity level fails, then in the next transmission, the failed packet is sent on the higher diversity level and therefore receives a higher “priority”. Therefore the lower priority packet opportunistically rides along with the higher priority packet thereby opportunistically uses the channel to reduce the delay.

We examine the utility of diversity embedding for scheduling by studying a stochastic arrival of packets to be delivered at the encoder. We have an exponential inter-arrival time point process

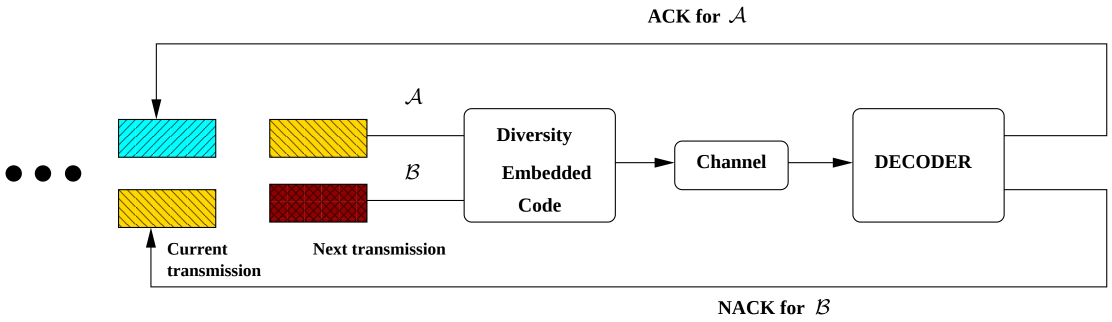
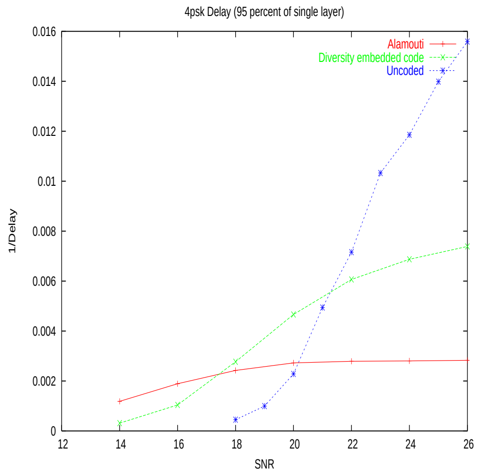


Figure 10: ARQ mechanism for diversity embedded codes using prioritized scheduling.

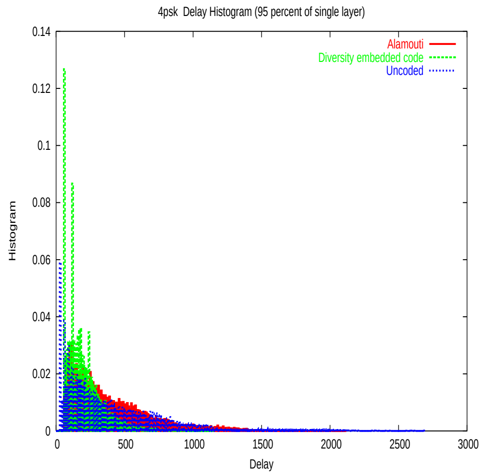
carrying an input packet to be delivered with each arrival. The average arrival rate of the packets are set to be a constant fraction of the average throughput of the full-diversity single-layer code. For such an arrival process, in Figure 11, we examine the impact of the ARQ mechanism illustrated in Figure 10 with a comparison to single layer schemes. In Figure 11 we transmit both the single-layer and the diversity embedded code using the same transmit 4-QPSK alphabet. Due to this, for a fixed packet size, the number of transmissions needed for each packet is *different* for the single-layer (full-diversity code) and the diversity embedded code as well as the single-layer (lower-diversity code). This is *specifically* taken into account in the average delay incurred for packet transmission. For example, if we use 4-QPSK constellations, for a packet size of 240 bits, the single-layer (full-diversity code) takes 120 transmissions. Whereas the diversity embedded code takes 240 transmissions, which also includes one packet on the full-diversity layer and two packets in the lower-diversity layer. This means that the packet delay is larger for a single transmission for the diversity embedded code, but as can be seen in Figure 11 the opportunistic scheduling more than makes up for this delay hit. We assume that the diversity embedded code gets ACK/NACK feedback on both levels separately. Figure 11(b) illustrates the delay histogram for these schemes at an SNR of 19.7dB.

Since the throughput of the schemes increases with SNR, it is natural to expect (as seen in Figure 11) that there is an SNR regime (in this case between $17-21$ dB where diversity embedded codes give lower average total delay than the single layer codes. Qualitatively this is similar to the throughput maximization of Figure 9. However, in Figure 11(b) which is plotted for an SNR of 19.7dB, the delay histogram is also much better for diversity embedded codes as compared to the single-layer codes. A similar comparison is done for a 8-QAM constellation in Figure 13, where qualitatively the same results are seen, but at a higher SNR range.

We have also examined the components of the delay, in terms of the transmission delay and queueing delay. The average transmission delay is the time it takes to successfully transmit the packet once it is put on one of the diversity layers. The average queueing delay is the time from the arrival of the packet to when it is assigned to one of the diversity layers. As can be expected, the transmission delay for diversity embedded codes is larger, since the time it takes to transmit even one packet is larger than the single-layer codes. This disadvantage is compensated for by much better queueing delay characteristics as seen in Figure 12 for a moderate SNR range. This shows that for a fixed alphabet, we are making a fair comparison between the single-layer and diversity embedded codes.

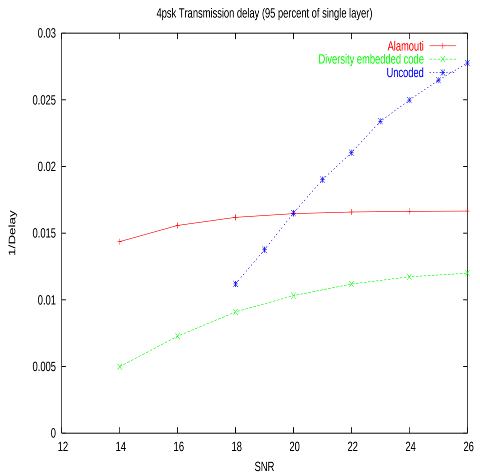


(a) Average delay

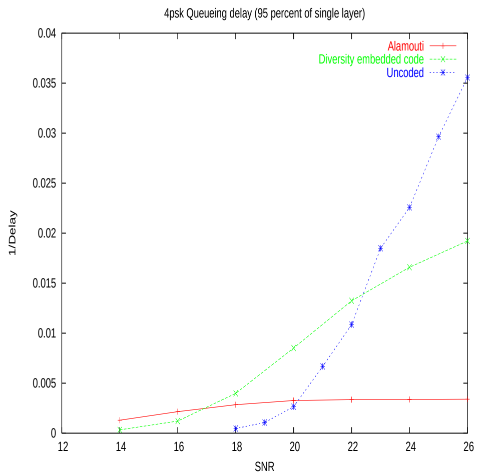


(b) Delay histogram

Figure 11: Delay comparison of diversity embedded codes with single layer codes when used with ARQ. The arrival process has exponential inter-arrival time which is 95% of the average throughput of the Alamouti code. In Figure (a), we plot the inverse average delay (in number of transmissions) of the scheduler for both single-layer and diversity embedded code implemented with opportunistic scheduling strategy. Figure (b) shows the histogram of the delays of the single-layer and diversity embedded schedulers. In both figures, the transmission alphabet is 4-QPSK.

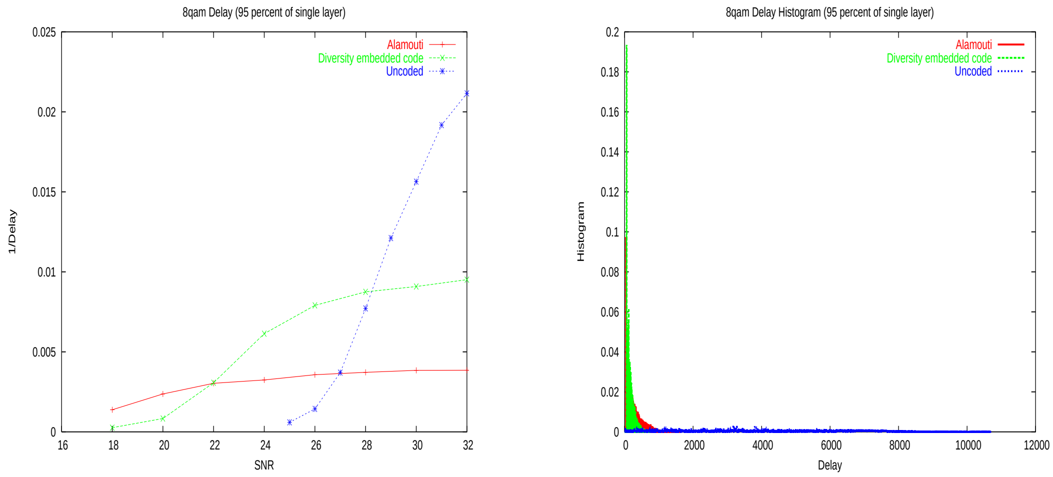


(a) Average transmission delay



(b) Average queueing delay

Figure 12: This figure compares the transmission delay and queueing delays, which together constitute the total average delay shown in Figure 11. In both figures, the transmission alphabet is 4-QPSK. In Figure (a), the transmission delay for the diversity embedded code is larger because it takes 240 transmissions for the higher layer packet, where as the single layer full-diversity code takes 120 transmissions, causing a higher delay. However in Figure (b), shows that the opportunistic scheduling has lower queueing delay in a certain SNR regime. In fact this compensates for the larger transmission delay giving a lower total average delay, for moderate SNR regime as seen in Figure 11.



(a) Average delay (b) Delay histogram

Figure 13: Delay comparison of diversity embedded codes with single layer codes when used with ARQ. The arrival process has exponential inter-arrival time which is 95% of the average throughput of the Alamouti code. In Figure (a), we plot the inverse average delay (in number of transmissions) of the scheduler for both single-layer and diversity embedded code implemented with opportunistic scheduling strategy. Figure (b) shows the histogram of the delays of the single-layer and diversity embedded schedulers. In both figures, the transmission alphabet is 8-QAM.

8 Discussion

In this paper we have designed a novel physical layer link that is better able to support applications with very different QoS requirements. The main idea is that of diversity embedded space-time codes in which a high-rate space-time code is designed to have a (lower-rate) high-diversity code embedded within it. This enables opportunistic use of a quasi-static channel realization without requiring channel knowledge at the transmitter. The constructions outperform simple time-sharing. There are several open questions including characterization of the achievable region, extremal code constructions and practical performance under complexity constraints.

A Proof of Theorem 5.1

We now verify the diversity embedding properties stated in Theorem 5.1 for multi-level codes employing QAM and PSK constellations.

As seen in Section 3.4, we can represent any point s in a 2^L-QAM constellation using a L-length bit string as

$$s - c(L) \equiv \sum_{l=0}^{L-1} b_l (1-i)^l \mod (1-i)^L, \tag{A.1}$$

where as before $f \equiv g \mod (1-i)^L$ if there exist $c, d \in \mathbb{Z}$ such that $f = (c+di)(1-i)^L + g$.

For 2^L-QAM constellations, consider space-time codewords $\mathbf{X}, \mathbf{X}^{'}$ determined by sequences of L binary matrices $\{\mathbf{K}_j\}_{j=1}^L, \{\mathbf{K}_j^{'}\}_{j=1}^L$ and specified by L message inputs $\{\mu_j\}, \{\mu_j^{'}\}$. If $\mathbf{K}_j, \mathbf{K}_j^{'} \in \mathcal{K}_j$, then by definition the rank of $\mathbf{K}_j - \mathbf{K}_j^{'}$ over the binary field $\mathbb{F}_2$ is at least D_j, for $\mathbf{K}_j \neq \mathbf{K}_j^{'}$. Also note that we have $D_1 \geq D_2 \geq \ldots D_L$. Note that this result applies to the general case where there can be L levels of diversity. If we have a smaller number of diversity orders, some of the sets chosen might be the same (as discussed in Section 5) but the argument for the diversity orders achieved remains the same.

Let us start with the first layer. If $\mu_1 \neq \mu_1^{'}$, then by construction we have $\mathbf{K}_1 \neq \mathbf{K}_1^{'}$ and therefore we need to show that $\text{rank}(\mathbf{X} - \mathbf{X}^{'}) \geq D_1$ when this happens. If we examine the element-wise difference of $\mathbf{X} - \mathbf{X}^{'}$ we see from the relationship in (A.1) that,

$$\left[x(p,q) - x^{'}(p,q)\right] \equiv \sum_{l=1}^{L} [\mathbf{K}_l(p,q) - \mathbf{K}_l^{'}(p,q)](1-i)^{l-1} \equiv [\mathbf{K}_1(p,q) - \mathbf{K}_1^{'}(p,q)] \mod (1-i). \tag{A.2}$$

Combining these element-wise results gives

$$\left[\mathbf{X} - \mathbf{X}^{'}\right] \equiv (\mathbf{K}_1 - \mathbf{K}_1^{'}) \mod (1-i), \tag{A.3}$$

which relates the codeword difference matrix to the difference of the binary matrices. This observation is key to the proof that the desired diversity order is achieved since it connects the properties in the binary domain to those in the complex domain. Now, we prove the desired property by contradiction. Suppose the rank of $(\mathbf{X} - \mathbf{X}^{'})$ is smaller than D_1, so there exist linearly independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{M_t - D_1 + 1}$ such that, $(\mathbf{X} - \mathbf{X}^{'})\mathbf{v}_l = \mathbf{0}$, $l = 1, \ldots, M_t - D_1 + 1$. Therefore, we have,

$$\left[(\mathbf{X} - \mathbf{X}^{'})\mathbf{v}_l\right] \equiv \mathbf{0} \mod (1-i), \quad l = 1, \ldots, M_t - D_1 + 1. \tag{A.4}$$

Since the elements of $(\mathbf{X} - \mathbf{X}^{'})$ are Gaussian integers, the entries of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{M_t - D_1 + 1}$ may be taken to be Gaussian integers. Note that we may suppose $\mathbf{v}_l \neq \mathbf{0} \mod (1-i)$ since we may divide through to eliminate common factors. Reducing (A.4) modulo $(1-i)$ gives $M_t - D_1 + 1$ linearly independent binary vectors $\{\bar{\mathbf{v}}_l\}$ for which $(\mathbf{K}_1 - \mathbf{K}_1^{'})\bar{\mathbf{v}}_l \equiv \mathbf{0} \pmod 2$. Hence we see that unless $\mathbf{K}_1 = \mathbf{K}_1^{'}$ we have a contradiction. This argument shows that we get the desired diversity order of D_1 for the first layer.

Now, we continue with an inductive argument. Suppose that $\mathbf{K}_l = \mathbf{K}_l^{'}$, for $l = 1, \ldots, n-1$. Then, analogous to (A.2), we see that element-wise,

$$\left[x(p,q) - x^{'}(p,q)\right] = \sum_{l=n}^{L-1} (\mathbf{K}_l(p,q) - \mathbf{K}_l^{'}(p,q))(1-i)^{l-1} \equiv (\mathbf{K}_n(p,q) - \mathbf{K}_n^{'}(p,q))(1-i)^{n-1} \mod (1-i)^n. \tag{A.5}$$

Equivalently, in matrix form we have,

$$\frac{\left[\mathbf{X} - \mathbf{X}^{'}\right]}{(1-i)^{n-1}} \equiv (\mathbf{K}_n - \mathbf{K}_n^{'}) \mod (1-i). \tag{A.6}$$

If the rank of $\mathbf{X}-\mathbf{X}^{'}$ is smaller than D_n, then there exist linearly independent vectors $\mathbf{w}_l$, for $l=1,\ldots,M_t-D_n+1$ such that,

$$\left[\mathbf{X}-\mathbf{X}^{'}\right]\mathbf{w}_l=\mathbf{0},\ \ l=1,\ldots,M_t-D_n+1. \tag{A.7}$$

Just as in the argument for $n=1$ before, we can suppose that $\mathbf{w}_l\neq\mathbf{0} \mod (1-i)$. The binary vectors $\{\bar{\mathbf{w}}_l\}$ obtained from $\{\mathbf{w}_l\}$ by reduction modulo $(1-i)$ are linearly independent over $\mathbb{Z}_2$. Reducing (A.6) modulo $(1-i)$ gives,

$$\left[\mathbf{K}_n-\mathbf{K}_n^{'}\right]\bar{\mathbf{w}}_l=\mathbf{0}\quad (\mathrm{mod}\ 2),\ \ l=1,\ldots,M_t-D_n+1. \tag{A.8}$$

Hence we see that unless $\mathbf{K}_n=\mathbf{K}_n^{'}$, we get a contradiction. In other words, it follows by induction that $\mathbf{K}_j=\mathbf{K}_j^{'}, j=1,\ldots,n$.

The proof for the 2^L-PSK constellations is similar to the case of QAM constellations given above. Given $\mathbf{X}\neq\mathbf{X}^{'}$, we use induction to prove that if the rank of $\mathbf{X}-\mathbf{X}^{'}$ is less than D_j, then $\mathbf{K}_l-\mathbf{K}_l^{'}$ for $l=1,\ldots,j$. Note that $D_1\geq D_2\geq\ldots\geq D_j$. We write the (p,q)-th elements of $\mathbf{X},\mathbf{X}^{'}$ as

$$x(p,q)=\xi^{\sum_{m=1}^{L}2^{m-1}\mathbf{K}_m(p,q)},\ \ x^{'}(p,q)=\xi^{\sum_{m=1}^{L}2^{m-1}\mathbf{K}_m^{'}(p,q)}.$$

Part (2) of Lemma 3.1 implies that,

$$\frac{x(p,q)-x^{'}(p,q)}{1-\xi}\equiv\mathbf{K}_1(p,q)-\mathbf{K}_1^{'}(p,q)\quad \mathrm{mod}\ (1-\xi)$$

so that by combining these element-wise relationships we get,

$$\frac{\mathbf{X}-\mathbf{X}^{'}}{1-\xi}\equiv\mathbf{K}_1-\mathbf{K}_1^{'}\quad \mathrm{mod}\ (1-\xi).$$

As above it follows that $\mathbf{K}_1=\mathbf{K}_1^{'}$. Now, suppose that $\mathbf{K}_l=\mathbf{K}_l^{'}$ for $l=1,\ldots,n-1$, then we can write the following element-wise relationships,

$$\frac{x(p,q)-x^{'}(p,q)}{1-\alpha}=\gamma\frac{\alpha^{\sum_{m=1}^{L-n}2^{m-1}\mathbf{K}_{m+n}(p,q)}-\alpha^{\sum_{m=1}^{L-n}2^{m-1}\mathbf{K}_{m+n}^{'}(p,q)}}{1-\alpha},$$

where $\alpha=\xi^{2^n}$ and $\gamma=\xi^{\sum_{m=1}^{n}2^{m-1}\mathbf{K}_m(p,q)}$. We extend part (2) of Lemma 3.1 to obtain,

$$\frac{x(p,q)-x^{'}(p,q)}{1-\xi^{2^n}}\equiv\left[\mathbf{K}_n(p,q)-\mathbf{K}_n^{'}(p,q)\right]\quad \mathrm{mod}\ (1-\xi^{2^n}),$$

and so by combining these element-wise relationships into a matrix we get,

$$\frac{\mathbf{X}-\mathbf{X}'}{1-\xi^{2^n}}\equiv\left[\mathbf{K}_n-\mathbf{K}_n^{'}\right]\quad \mathrm{mod}\ (1-\xi).$$

As above, it follows that $\mathbf{K}_n=\mathbf{K}_n^{'}$ and the result follows by induction.

References

[1] S.M. Alamouti. A simple transmit diversity technique for wireless communications. *IEEE Journal on Selected Areas in Communications*, 16(8):1451–1458, October 1998.

[2] Michael Artin, *Algebra.* Prentice Hall, Inc., New York, 1991.

[3] E. Biglieri, J. Proakis, and S. Shamai. Fading channels: Information-Theoretic and Communications aspects. *IEEE Transactions on Information Theory*, 44(6):2619–2692, October 1998.

[4] A.R. Calderbank, Multilevel codes and multistage decoding *IEEE Transactions on Communications*, 37(3):222 - 229, March 1989.

[5] A. R. Calderbank and A. F. Naguib. Orthogonal designs and third generation wireless communication. In J. W. P. Hirschfeld, editor, *Surveys in Combinatorics 2001, London Mathematical Society Lecture Note Series #288*, pages 75–107. Cambridge Univ. Press, 2001.

[6] A R. Calderbank and N. Seshadri. Multilevel codes for Unequal Error Protection. *IEEE Transactions on Information Theory*, 39(4):1234–1248, July 1993.

[7] A. R. Calderbank, S. N. Diggavi and N. Al-Dhahir. Space-Time Signaling based on Kerdock and Delsarte-Goethals Codes. *IEEE International Conference on Communications (ICC)*, pp 483 - 487, Paris, June 2004.

[8] J. Chui and A. R. Calderbank Effective coding gain for space-time codes. *IEEE International Symposium on Information Theory (ISIT)*, Seattle, July 2006.

[9] T. M. Cover. Broadcast channels. *IEEE Transactions on Information Theory*, 18:2–14, January 1972.

[10] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley and Sons, Inc., New York, 1991.

[11] R.V. Cox, J. Hagenauer, N. Seshadri and C.-E.W. Sundberg, Subband speech coding and matched convolutional channel coding for mobile radio channels, *IEEE Transactions on Signal Processing,* 39(8):1717–1731, August. 1991.

[12] I. Csiszar and J. Korner. *Information Theory: Coding theorems for Discrete Memoryless Channels.* Academic, New York, 1982.

[13] M O. Damen, A. Chkeif, and J.-C. Belfiore. Lattice codes decoder for space-time codes. *IEEE Communication Letters*, 4:161–163, May 2000.

[14] S N. Diggavi. On achievable performance of spatial diversity fading channels. *IEEE Transactions on Information Theory*, 47(1):308–325, January 2001.

[15] S. N. Diggavi, N. Al-Dhahir, and A. R. Calderbank. Diversity embedding in multiple antenna communications, advances in network information theory. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, pages 285-301, 2004

[16] S. N. Diggavi, N. Al-Dhahir, and A. R. Calderbank. Diversity embedded space-time codes. *IEEE GLOBECOM Conference*, San Francisco, December 2003, pp 1909–1914

[17] S N. Diggavi, S. Dusad, A. R. Calderbank, N. Al-Dhahir, "On Diversity Embedded codes," *Proceedings of Allerton Conference on Communication, Control, and Computing*, Illinois, September 2005.

[18] S. N Diggavi and D. Tse, *On successive refinement of diversity*, invited paper, Allerton Conference, October, 2004.

[19] S. N Diggavi and D. Tse, *Fundamental Limits of Diversity-Embedded Codes over Fading Channels*, IEEE International Symposium on Information Theory (ISIT), pp 510–514. September, 2005.

[20] S. N Diggavi and D. Tse, *On opportunistic codes and broadcast codes with degraded message sets*, IEEE Information Theory Workshop (ITW), pp 227–231, March, 2006.

[21] H. El-Gamal, G. Caire, G and M O. Damen, The diversity-multiplexing-delay tradeoff in MIMO ARQ channels *Proceedings International Symposium on Information Theory, 2005. ISIT 2005.* pp 1823–1827, Sept. 2005.

[22] G.J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Labs Technical Journal*, 1(2):41–59, September 1996.

[23] E. Gabidulin, Theory of codes with maximum rank distance. *Probl. Per. Inform.*, 21:3–16, Jan/March 1985.

[24] J-C. Guey, M. P. Fitz, M. R. Bell, and W-Y. Kuo. Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels. *IEEE Transactions on Communications*, 47(4):527–537, April 1999.

[25] J. Hagenauer. Rate Compatable Punctured Convolutional Codes (RCPC codes) and their applications. *IEEE Transactions on Communications*, 36(4):389–400, April 1988.

[26] A. R. Hammons, Jr., H. El-Gamal, On the theory of space-time codes for PSK modulation, *IEEE Transactions on Information Theory*, Vol 46, No. 2, pp 524–542, March 2000.

[27] H. Imai, S. Hirakawa, A new multilevel coding method using error-correcting codes, *IEEE Transactions on Information Theory,* 23(3):371–377 , May 1977.

[28] H. F. Lu and P. V. Kumar. Rate-diversity trade-off of space-time codes with fixed alphabet and optimal constructions for PSK modulation. *IEEE Transactions on Information Theory*, 49(10):2747–2752, October 2003.

[29] H. F. Lu and P. V. Kumar A unified construction of space-time codes with optimal rate-diversity tradeoff, *IEEE Transactions on Information Theory*, 51(5):1709–1730, May 2005.

[30] B. Masnick and J K. Wolf, On linear unequal error protection codes, *IEEE Transactions on Information Theory,* 13(4):600–607, October 1967.

[31] G.J. Pottie, D.P. Taylor, Multilevel codes based on partitioning, *IEEE Transactions on Information Theory*, 35(1):87–98, Jan. 1989.

[32] R. van Nobelen, N. Seshadri, J. Whitehead, and S. Timiri. An adaptive radio link protocol with enhanced data rates for GSM evolution. *IEEE Personal Communications*, 6(1), February 1999.

[33] L H. Ozarow, S Shamai, and A. D. Wyner. Information theoretic considerations for cellular mobile radio. *IEEE Transactions on Vehicular Technology*, 43(2):359–378, May 1994.

[34] A. Seeger. *Hierarchical Channel Coding for Digital Video Broadcasting.* PhD thesis, Technical University of Munich, Munich, Germany, 1999.

[35] S. Shamai (Shitz). A broadcast strategy for the Gaussian slowly fading channel. In *IEEE International Symposium on Information Theory (ISIT)*, Ulm, Germany, p 150, June 1997.

[36] S. Shamai (Shitz) and A. Steiner. Single user broadcasting in a MIMO channel. In *Information Theory Workshop (ITW2003)*, Paris, France, pp 38–41, April 2003.

[37] I E. Telatar. Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications*, 10(6):585–596, November-December 1999.

[38] V. Tarokh, H. Jafarkhani, and A.R. Calderbank. Space-time block codes from orthogonal designs. *IEEE Transactions on Information Theory*, pages 1456–1467, July 1999.

[39] V. Tarokh, N. Seshadri, and A.R. Calderbank. Space-time codes for high data rate wireless communications: Performance criterion and code construction. *IEEE Transactions on Information Theory*, 44(2):744–765, March 1998.

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

[41] L. Washington *Introduction to cyclotomic fields*, *Springer Verlag*, 2nd edition June 1997.

[42] G. Ungerboeck, Channel coding with multilevel/phase signals, *IEEE Transactions on Information Theory*, 28(1):55–67, Jan 1982.

[43] L. Zheng and D N C. Tse. Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels. *IEEE Transactions on Information Theory*, 49(5):1073–1096, May 2003. Available from http://degas.eecs.berkeley.edu/ dtse/pub.html.