SPACE-TIME CODES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED MATHEMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MURAT KARAÇAYIR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
CRYPTOGRAPHY

MAY 2010

Approval of the thesis:

SPACE-TIME CODES

submitted by **MURAT KARAÇAYIR** in partial fulfillment of the requirements for the degree of **Master of Science in Department of Cryptography, Middle East Technical University** by,

Prof. Dr. Ersan Akyıldız _____
Director, Graduate School of **Applied Mathematics**

Prof. Dr. Ferruh Özbudak _____
Head of Department, **Cryptography**

Prof. Dr. Ferruh Özbudak _____
Supervisor, **Department of Mathematics**

Examining Committee Members:

Assoc. Prof. Dr. Ali Doğanaksoy _____
Department of Mathematics, METU

Prof. Dr. Ferruh Özbudak _____
Department of Mathematics, METU

Assist. Prof. Dr. Zülfükar Saygı _____
Department of Mathematics, TOBB ETU

Dr. Muhiddin Uğuz _____
Department of Mathematics, METU

Dr. Burcu Gülmez Temür _____
Department of Mathematics, Atılım University

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: MURAT KARAÇAYIR

Signature :

ABSTRACT

SPACE-TIME CODES

Karaçayır, Murat
M.S., Department of Cryptography
Supervisor : Prof. Dr. Ferruh Özbudak

May 2010, 72 pages

The phenomenon of fading constitutes a fundamental problem in wireless communications. Researchers have proposed many methods to improve the reliability of communication over wireless channels in the presence of fading. Many studies on this topic have focused on diversity techniques. Transmit diversity is a common diversity type in which multiple antennas are employed at the transmitter. Space-time coding is a technique based on transmit diversity introduced by Tarokh et alii in 1998.

In this thesis, various types of space-time codes are examined. Since they were originally introduced in the form of trellis codes, a major part is devoted to space-time trellis codes where the fundamental design criteria are established. Then, space-time block coding, which presents a different approach, is introduced and orthogonal space-time block codes are analyzed in some detail. Lastly, rank codes from coding theory are studied and their relation to space-time coding are investigated.

Keywords: Wireless Communications, Diversity, Space-Time Trellis Codes, Orthogonal Designs, MRD Codes

ÖZ

UZAY-ZAMAN KODLARI

Karaçayır, Murat
Yüksek Lisans, Kriptografi Bölümü
Tez Yöneticisi : Prof. Dr. Ferruh Özbudak

Mayıs 2010, 72 sayfa

Sönümleme kavramı kablosuz haberleşmede temel bir sorun teşkil etmektedir. Araştırmacılar sönümlemenin varlığında kablosuz kanallar üzerinden iletişimin güvenilirliğini artırmak için çeşitli yöntemler önermişlerdir. Bu konu üzerindeki birçok çalışma, çeşitleme tekniklerine odaklanmıştır. İletim çeşitlemesi, ileticide birden çok sayıda antenin kullanıldığı yaygın bir çeşitleme türüdür. Uzay-zaman kodlama, 1998'de Tarokh ve diğerleri tarafından ortaya atılan, iletim çeşitlemesini temel alan bir tekniktir.

Bu tezde, uzay-zaman kodlarının çeşitli türleri incelenmiştir. Bu kodlar orijinal olarak kafes kodları olarak ortaya çıktığından, önemli bir bölüm uzay-zaman kafes kodlarına ayrılmış ve burada ayrıca temel tasarım ölçütleri ortaya konmuştur. Daha sonra, farklı bir yaklaşım sunan uzay-zaman blok kodlama tanıtılmış ve ortogonal uzay-zaman blok kodları belli bir detayla ele alınmıştır. Son olarak, kodlama kuramından rank kodları çalışılmış ve uzay-zaman kodlarıyla ilişkileri araştırılmıştır.

Anahtar Kelimeler: Kablosuz Haberleşme, Çeşitleme, Uzay-Zaman Kafes Kodları, Ortogonal Tasarımlar, MRD Kodları

Aileme

ACKNOWLEDGMENTS

First of all, I would like to thank my thesis supervisor Prof. Dr. Ferruh Özbudak for introducing me to this topic and also for his patience and understanding towards me. Without the motivation inspired by him, this thesis could not have started.

Secondly, it is a pleasure to express my gratitude to my family, whose unwavering love and support I have always felt with myself throughout my life. Their contribution on this product cannot be exaggerated.

Lastly, I would like to thank The Scientific and Technological Research Council of Turkey (TÜBİTAK), who made the writing of this thesis easier by supporting me financially during my graduate study.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

Wireless communication has always been a part of life since the invention of radio. The reasons of the excitement caused by the idea of communicating over long distances through air are rather understandable. Today, recent developments on the topic, especially the acclaimed 3G technology, have given rise to an enthusiasm wave which is influencing increasingly more people. The issue should not be interpreted solely as entertainment; on the contrary, hundreds of millions of people are relying on wireless communications for serious tasks. To name a few, communication via cell phones, TV broadcasting, monetary transactions, registration to courses and taking exams are wireless applications which are already perceived as "ordinary". These examples give enough evidence that in the near future, the matter will reach, if it has not yet, such extents that it will not be a matter of choice for most people to design their lives according to wireless applications. Even if we put all these aside, the reality today is that the demand for wireless applications does not seem to have any prospect of diminishing.

Since wireless communications is such an important topic, the works of many authors for several decades have produced a giant literature on this field. This literature shows that wireless communications is a highly interdisciplinary area of research. Results from information theory, coding theory, mathematics, statistics and computer science have been used thoroughly by researchers both to draw conclusions related to the subject and to devise new techniques improving the quality of communication greatly.

1.1 Problem Definition

In wireless channels, a signal sent from a transmitter does not follow a single path before it reaches its destination. Instead, objects present in the environment cause it to traverse many different paths by means of physical effects such as reflection and refraction. Thus, multiple versions of the transmitted signal reach the receiver. The observed signal at the receiver is a sum of these multiple signals, and it is typically different from the originally transmitted one. Furthermore, in real applications, the relative positioning of transmitter-receiver pairs and the overall state of the objects between them may vary frequently in time, causing a change in the multiple paths that signals follow. As a result, it is not rare that the signal observed by a receiver does not suffice to recover the actually transmitted signal. This factor, known as "multipath fading" or simply as "fading", is a fundamental problem in wireless communication.

Space-time coding was first described in [1] by Tarokh, Seshadri and Calderbank as a solution to this problem. It claims to increase the reliability of data transmission in wireless communication systems. Like many other wireless communication schemes, it is based on a technique known as *diversity*. We proceed with a brief description of this technique.

1.2 Diversity

As expressed before, the paradigm of fading and its time-varying nature constitute a fundamental problem when communicating over wireless channels. During some time periods, the transmitted symbol can well be recovered by the receiver despite the presence of fading; while during other time periods, fading may reach extents that make faultless transmission of data impossible. This latter case is referred to as *deep fading*. Therefore, it is reasonable to assert that a communication scheme is likely to suffer from errors if it depends on the strength of a single signal path. One way to remove this dependency is to ensure that each individual symbol is sent over several paths which undergo independent fading. By this way, correct transmission of an information symbol is achieved as long as one of its paths is strong. This resource is known as diversity.

Diversity has several types, but only two of them are relevant in our discussion of space-time codes. These are *time diversity* and *space diversity*.

Time Diversity

The fading characteristics of a wireless communication channel can be viewed as a function of time. Intuitively, information symbols transmitted with a small time difference will undergo similar amounts of fades. In other words, as two signals are transmitted further apart in time, the fades acting on them tend to behave more independently from one another. Both for theoretical discussions and real applications, a more precise explanation is provided by the so called *coherence time*. Roughly speaking, coherence time can be defined as the time duration over which the state of a channel remains predictable. Therefore, two signals are assumed to undergo dependent fades if they are transmitted within a time period shorter than the coherence time; otherwise they are assumed to experience independent fades.

Time diversity (also called *temporal diversity*) combats fading by making proper use of *coding* and *interleaving*. Information is encoded using an error-correcting code, each codeword is divided into L parts and consecutive parts are transmitted inside different coherence periods. More explicitly, if $\mathbf{x} = (x_1, x_2, \ldots, x_L)$ is a codeword and x_i and x_j are any two parts of $\mathbf{x}$ transmitted at times t_i and t_j respectively, then the difference $|t_i - t_j|$ should be greater than T_c, coherence time of the channel. Then, the definition of coherence time implies that all parts of $\mathbf{x}$ undergo independent fades. Even if some of them get lost due to presence of deep fade, depending on the error correcting capability of the code, others may suffice to correctly recover the original information. In this context, L is called the number of *diversity branches*. The simplest realization of a time diversity scheme occurs when one uses the repetition code as the error-correcting code. In this case, each codeword consists of L repeated copies of an information symbol. The same symbol is simply transmitted L times from the same antenna, each transmission being inside a different coherence period.

Although it increases the reliability of data transmission significantly, time diversity is easily seen to have major drawbacks. First of all, its data rate is low: it takes L symbol times to transmit one symbol. In addition, due to the necessity of spreading

a codeword over different coherence periods, decoding at the receiver cannot start without a certain amount of delay. This is clearly a problem in applications where long decoding delays are not tolerable (e.g. in voice transmission over mobile phones). This means further that employing solely time diversity is not a good option when the channel fading varies too slowly with time. All these drawbacks motivate the use of space diversity together with time diversity.

Space Diversity

Time variation of multipath fading mainly comes from the fact that receivers in real applications are mobile objects. Time diversity makes use of this fact to ensure that signals representing the same information is sent over independently fading paths. Clearly, a receiver's mobility is not the only way to achieve this effect. It is known that sufficiently separated antennas cause multipaths which fade more or less independently. This resource is referred to as *space diversity* (or *spatial diversity*) and is a special case of the more general *antenna diversity*. Space diversity is called *transmit diversity* if multiple transmit antennas are used and *receive diversity* if multiple receive antennas are used. In this thesis we will primarily be dealing with schemes employing transmit diversity, while receive diversity will only be optional.

A simple scheme which combines time diversity and space diversity can be as follows: Suppose we want to transmit a symbol x. Again, we encode x with an error correcting code and obtain $\mathbf{x} = (x_1, x_2, \ldots, x_L)$. Then we use L different transmit antennas to transmit $\mathbf{x}$. At symbol time $t = i$, x_i is transmitted by transmit antenna i and the other antennas are silent. If the error correcting code is the repetition code, then this scheme amounts to transmitting the same information symbol through L transmit antennas over L symbol times. Note that this time we do not have to wait between consecutive symbols since variation in fading results from the use of different transmit antennas. Therefore, this scheme is better suited in cases where there is a strict delay requirement.

Diversity is an important resource in wireless communications. For that reason, many schemes combine several types of diversity. As the name suggests, space-time codes

employ space and time diversity to increase the reliability of wireless communication. A more thorough treatment of these and other diversity techniques can be found in [6, Chapter 3]. A reader who is particularly interested in transmit diversity may also consult [7].

1.3 Parameters Related to Wireless Communications

What do we mean when we talk about *improving* the *performance* of a wireless communication system? Is performance something measurable, and if it is, what are the measures of performance? These fundamental questions have been the subject of many studies so far. The results of these have produced a good deal of theory on the field, which provides quite satisfactory explanations on the aspects that can be used to evaluate the performance of a wireless communication system. In this part, we give brief information on some of these aspects. Our presentation will mainly be based on heuristic explanations rather than mathematical ones.

Data Rate, Error Rate

The emergence of a communication theory was prompted by the fact that the nature is not fully controllable. Regardless of which medium one conveys information through, there is always some unpredictable element which may prevent the intended receiver from receiving the information exactly in the form it was sent. This brings about a positive probability of communication error and the main purpose of any communication scheme is to reduce this probability as much as possible. Nevertheless, no communication scheme is error free and thus, errors are expected to occur with some rate depending on which scheme we are using. This is called *error rate*. The most obvious performance measure of any communication scheme-wireless or not-is the error rate. Of course, the relation is inverse; the smaller the error rate, a higher performance the communication scheme has.

In wireless communications, information is carried by electromagnetic signals. The receiver may not observe this signal exactly as it was sent, partly due to the unpredictable element mentioned in the previous paragraph, which is called *noise*. Whether

or not the receiver will be able to obtain the sent information depends on the relative strength of the received signal and the noise. Therefore, the ratio

$$\frac{P_{\text{signal}}}{P_{\text{noise}}},$$

where P_{signal} and P_{noise} are the average powers of the signal and the noise respectively, is very important and is termed *signal-to noise ratio*(SNR). SNR is usually defined in a logarithmic scale given by

$$\mathsf{SNR}_{\text{dB}} = 10 \log_{10} \frac{P_{\text{signal}}}{P_{\text{noise}}},$$

where dB stands for "decibels". Intuitively, as SNR increases, the error rate decreases. Therefore, many communication schemes aim at increasing the SNR. There are two straightforward means to achieve this: decreasing the noise power and increasing the signal power. The first option is not likely to work because the part of noise which is controllable in a communication device is generally idealized in advance and the rest is out of human control. The second option is also flawed. Most wireless devices are not able to increase the power. For instance, cell phones rely on a battery to function, whose power is very limited. Even if power is increased, both the signal and the noise grows in strength in certain cases. This can be felt when one increases the volume of a radio. Since the two most obvious ways of increasing SNR do not work in general, one needs more elaborated techniques instead, of which diversity is one.

The idea that information is a measurable phenomenon was realized by the groundbreaking work of Claude E. Shannon in the late 1940s. In [2], he introduced the so called *Shannon capacity*, which draws a limit to the amount of information that can be sent over a channel in unit time with arbitrarily low error rate. The amount of information sent in unit time is called *data rate*. According to Shannon's Noisy-channel Coding Theorem, the typical way of achieving data rates close to channel capacity without sacrificing the reliability of communication is to increase the amount of information sent at a time. On wireless channels, however, this cannot be done due to delay limitations. Furthermore, the element of fading implies that the capacity of certain wireless channels is practically zero(see the note in [6, page 218]). Therefore, for wireless channels, a more accurate performance measure is provided by *outage capacity*. Outage capacity is the maximum rate for which the probability of error is less

than a given threshold. Limits of wireless communication in terms of outage capacity have been investigated by several authors. Interested reader may see [3, 4].

Diversity Gain, Coding Gain

We have made a conceptual description of diversity but we have not yet provided mathematical means to define it. Diversity is indeed measurable and is closely related to the error rate of a channel. We can use several diversity techniques in a wireless communication scheme but how much performance is improved by doing so is subject to mathematical analysis. In other words, how much we *gain* by introducing diversity to a system can be expressed mathematically and is called *diversity gain*. A definition of diversity gain commonly used in the literature is given by

$$-\lim_{\mathsf{SNR}\to\infty}\frac{\log(\mathrm{PEP})}{\log(\mathsf{SNR})},$$

where PEP is the pairwise error probability, i.e. the probability that the receiver decodes the transmitted information erroneously. This probability is bounded from above by an expression whose leading term is a multiple of $\left(\frac{1}{\mathsf{SNR}}\right)^L$ where L is the diversity gain(see [6, page 76]).

A similar performance measure is *coding gain*. Coding gain is an approximate measure of the advantage provided by a coded system over an uncoded one having the same diversity gain[1, page 6]. Coding adds redundancy to information by mapping information sequences to longer(when expressed as a binary vector) code sequences or codewords. Then the distance(in a predefined sense) of two codewords is larger than the distance of the information sequences to which they correspond. This reduces the probability of confusing different information pieces, which is the reward of adding redundancy. Coding gain is a quantization of this and closely related to the minimum distance of distinct codeword pairs. Like diversity gain, coding gain also shows itself in an upper bound on PEP. Coding gain also has a nice geometric interpretation. Performance of a wireless system is often illustrated by a PEP versus $\mathsf{SNR}_{\mathrm{dB}}$ graph. Coding gain of a system is the horizontal shift of its performance graph compared to that of an uncoded system and hence measured in decibels.

All of the performance measures mentioned in this section can be found in much

greater detail in many textbooks on wireless communications. The reader who would like to develop insight into these fundamental concepts is referred to [5, 6].

1.4 Thesis Outline

The rest of these thesis is about three different types of space-time codes. In Chapter 2, we first describe the channel model which will be valid for the other chapters as well. We then describe the encoding process of space-time trellis codes in detail and discuss their decoding briefly. Performance criteria for space-time codes are also established in this chapter. Among these criteria, the *rank criterion* applies to all the codes in the thesis.

Next two chapters are about space-time block codes, where each of them focuses on a different type of construction. In Chapter 3, orthogonal space-time block codes are introduced, which are constructed by making use of some results from orthogonal design theory. The subject of Chapter 4 is how maximal rank distance codes from coding theory can be used to construct "good" space-time block-codes. In both chapters, codes under discussion are examined according to the rank criterion.

CHAPTER 2

SPACE-TIME TRELLIS CODES

Space-time coding was first proposed in [1] by Tarokh, Seshadri and Calderbank as a method in wireless communications. It claims to increase the reliability of communication by using several antennas at the transmitter. Therefore, it is a transmit diversity scheme. Since the codes given in [1] is based on trellis codes, these are called *space-time trellis codes*.

In Section 2.1, we define some terms related to wireless channels and describe the channel assumptions which will be valid throughout the rest of the thesis. A detailed description of space-time trellis codes is the topic of Section 2.2. In Section 2.3, three performance criteria about space-time codes are established. How different parameters about space-time codes are linked is covered in Section 2.4. A simple example of space-time trellis codes is given in Section 2.5.

2.1 Channel Model

When communicating over wireless channels, the numerosity and complexity of the physical mechanisms behind multipath fading make it certain that having control over all the channel parameters is practically impossible. Although base stations and large-scale objects such as buildings remain stationary in very long time durations, receivers and relatively small objects present in the environment move in an unpredictable manner. Therefore, it is more or less necessary to deal with the channel as a whole and describe it using an appropriate statistical model. Still, models which characterize cases where the number, location and geometry of all the reflectors are known have

also been analyzed in the literature. Interested reader is referred to [8, Section 2.2.1] for a description of these models.

Suppose we want to send information using a transmit antenna. The baseband representation $x_b(t)$ of the information at time t and its corresponding modulated signal $x(t)$ are related by

$$x(t) = x_b(t)e^{i2\pi f_c t}, \tag{2.1}$$

where $i = \sqrt{-1}$ and f_c is the frequency of the carrier wave. Since the signal takes some time to reach the receiver, each path introduces a delay as well as an attenuation factor. More explicitly, if $M(t)$ is the set of all multipaths at time t, then the received signal at time t is expressed by

$$y(t) = \eta(t) + \sum_{p \in M(t)} a_p x(t - \tau_p), \tag{2.2}$$

where a_p and τ_p are respectively the strength and delay of the signal path p from the transmit antenna to the receive antenna. $\eta(t)$ is a sample of additive white Gaussian noise (AWGN) affecting the received signal at time t, which we ignore for now. Rather than having to contend with the effect of each individual path separately, we are interested in the aggregate effect of the channel on the transmitted signals and we denote this effect by $\alpha(t)$, called the *path gain*. Then (2.2) can be written simply as

$$y(t) = \alpha(t)x(t) + \eta(t). \tag{2.3}$$

Note that we are not using any delay term since we assume that the strength and delay associated with all signal paths are already considered in the calculation of $\alpha(t)$. For a more complete description of path gain, one can see [6, Chapter 2].

One important point about schemes employing space-time codes is that they do not treat time as a continuous concept. Transmissions in these schemes occur inside separate time slots. For that reason, (2.3) should not be interpreted as a continuous function of time in our context. Rather, it should be interpreted as a function of the set $\{1, 2, \ldots, T\}$ where T is the number of time slots in which transmission will occur. To emphasize this nuance, instead of (2.3),

$$y_t = \alpha_t x_t + \eta_t \tag{2.4}$$

will be used from now on.

Since a wireless channel is effectively described by its path gains, it is important to describe the behaviour of these fade coefficients statistically. Among all such models the most commonly used is the *Rayleigh fading*, which assumes that there are many small reflectors in the environment. In such a case, the fade coefficients can be modeled as a zero-mean complex Gaussian random variable with the same variance for both (real and imaginary) dimensions[6, Section 2.4.2]. An important assumption of Rayleigh fading is that none of the signal paths is dominant over the others. This means in particular that there is no line of sight between the transmitter and the receiver since this path would dominate the propagation medium. These explanations make it clear that it is reasonable to use a Rayleigh fading model for heavily built-up urban areas. If one wishes to model cases where there is a line of sight between the transmitter and the receiver, *Rician fading* should be used instead. We will assume a scenario which is well approximated by Rayleigh fading throughout this thesis.

Another important aspect of a wireless channel is how fast it varies. More precisely, we are interested in the coherence time and its relation to the delay constraint of the application. If coherence time is large relative to the delay constraint, the resulting situation is called *slow fading*. In slow fading, the fade coefficients within a coherence period are dependent, possibly equal. If they are equal and vary from one coherence period to another, *quasistatic fading* occurs. The opposite case occurs when the coherence time is smaller than the delay constraint of the channel and it is known as *fast fading*. Slow fading corresponds to situations where fading is mainly due to shadowing caused by large-scale stationary objects such as buildings and mountains. In mobile applications such as GPRS, however, the presence of many mobile users causes a fast variation in the channel and hence such channels exemplify fast fading. Although fast fading applies to such a wide class of cases, we will mainly be interested in quasistatic fading for simplicity.

A confusion might arise about the relation between the time variation of a channel and the statistical model we use for the fade coefficients. In our context, time variation of a channel refers to a change in the actual values of its path gains, whereas the statistical model is just a probabilistic distribution of these path gains. As the overall state of the channel varies with time, the actual fade coefficients vary; but their probabilistic description may tend to become further from or closer to a given model, say, Rayleigh

fading, independently of the changes in these coefficients. This distinction will become more evident in the next section, where multiple path gains will be in question due to the presence of multiple transmit antennas.

The last classification of wireless channels related to our discussion is about *frequency selectivity*. It is a known fact that any waveform is a sum of periodic signals of different frequencies. As far as wireless communication is concerned, any signal sent over a wireless channel is composed of components having different frequencies. If these components are affected differently by the channel, the fading is said to be *frequency selective*. Otherwise all frequency components of a signal undergo the same amount of fading and the channel is called a *flat fading* channel. Our interest will be in flat fading channels.

To summarize, whenever the channel characteristics will be important for us, we will be using a channel model where the fading is quasistatic, flat and path gains are selected according to a Rayleigh distribution. It should be stressed that these details will not be mentioned much in what follows. They are nevertheless important in order to prevent misguidance because what is true for a particular type of channel may not hold for other channels. We continue with a description of space-time trellis codes.

2.2 Space-Time Trellis Codes

In this section we briefly describe space-time trellis codes as they were originally introduced in [1]. First, the underlying system model will be explained. Secondly, a description of the scheme will be made. Then, the structure of the encoder will be explained in detail. Lastly, decoding of space-time trellis codes will be discussed.

2.2.1 System Model

In a mobile communication system with many users, it is most cost effective to equip the base station with multiple antennas and having a small number(possibly one) of antennas in mobile units[9]. Let us consider such a system where the base station is equipped with $n_T > 1$ antennas and each mobile has $n_R \geq 1$ antennas. Data

transmission can be either from the base station to the mobile unit or from the mobile unit to the base station. However, we are only interested in the former case. Therefore, this is a transmit diversity scheme where receive diversity is only optional.

Suppose we want to make data transmission over such a system during l symbol times. l is called the *frame length*, meaning that there will be l different time slots in which data transmission will be made. At time slot t, the transmitted signal from transmit antenna i for $i = 1, 2, \ldots, n_T$ is x_t^i and the signal received by receive antenna j is

$$y_t^j = \eta_t^j + \sum_{i=1}^{n_T} \alpha_{i,j} x_t^i \sqrt{E_s} \tag{2.5}$$

for $j = 1, 2, \ldots, n_R$. Here $\sqrt{E_s}$ is a real scaling factor to ensure that the average energy per signal is 1. η_t^j is the noise at time t associated with receive antenna j and it is selected from independent samples of a complex Gaussian random variable with mean zero and variance equal to $N_0/2$ for both (real and imaginary) dimensions. $\alpha_{i,j}$ is the path gain from the transmit antenna i to the receive antenna j and it is assumed to be quasistatic, flat and Rayleigh distributed with variance 0.5 for both dimensions. In addition, these path gains are assumed to be independent for different (i, j) pairs. This corresponds to the assumption that signals transmitted from different transmit antennas as well as signals received by different receive antennas undergo independent fades.

The last assumption we make about the channel is the availability of complete channel state information at the receiver. In reality, path gains at a particular time can be known only with an error due to the time variation of the channel. In our model, since the channel fade coefficients are assumed to be constant during a data frame, it is reasonable to assume that these coefficients are known by the receiver during that frame. For the methods of obtaining channel state information at the receiver, one can see [6, page 65]. The effects of not knowing the channel has also been established in the literature. The reader is referred to [10] for details on this issue.

2.2.2 Description of the Scheme

Now we turn our attention from the channel properties to the transmitter. Given binary data, the problem as to how this data will be converted to signals needs to

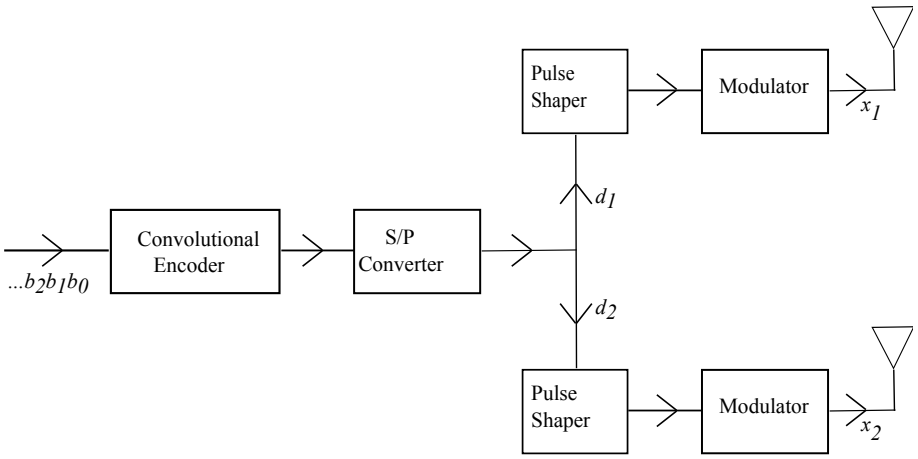

Figure 2.1: Block diagram of the transmitter.

be addressed. This is achieved through coding followed by modulation. First, binary data is encoded by a convolutional encoder. Then, the encoded data is split into n_T parts of fixed size by a serial to parallel converter. After that, each part is converted to a digital signal by means of a pulse shaper. The outputs of the pulse shaper are then modulated and modulated signal i is transmitted by transmit antenna i for $i = 1, 2, \ldots, n_T$. Since a convolutional code can be represented by a trellis diagram, the class of codes defined by this scheme is known as *space-time trellis codes*. Figure 2.1 illustrates the block diagram of a transmitter with $n_T = 2$ antennas.

Although the above description of the scheme is valid for any modulation, it will be useful to fix a modulation type for the understandability of the discussion. For this purpose we choose M-PSK modulation since the examples given in [1] are mostly for M-PSK modulation. The constellation size of M-PSK modulation is equal to $M = 2^p$ where p is a positive integer, and its elements are the M^{th} roots of unity scaled by a factor. More explicitly, the constellation for M-PSK modulation is given by the set

$$\{ae^{i2\pi\frac{k}{M}} : a \in \mathbb{R}, k = 0, 1, \ldots, M-1\}. \tag{2.6}$$

The commonly used 4-PSK and 8-PSK constellations are shown in Figure 2.2, where the complex elements are mapped to integers modulo 4 and 8, respectively, as in the figure. Other mappings are also possible, but we will use this one. From now on, elements of M-PSK constellation will be represented by integers specified by the mapping $ae^{i2\pi\frac{k}{M}} \longrightarrow k$.

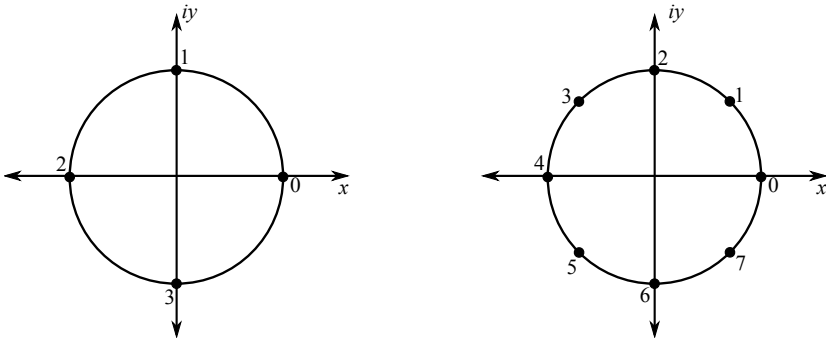

Figure 2.2: 4-PSK and 8-PSK constellations.

One point which is still ambiguous about the scheme is the frame length. In fact, Figure 2.1 does not impose any restriction on the frame length. At any time slot, a number, say b, of bits arrive at the encoder. Using these bits and the bits in its memory, the encoder produces pn_T bits and these bits are divided into n_T data streams $d_1, d_2, \ldots, d_{n_T}$, each having p bits. Then the d_i's are modulated and sent from different transmit antennas simultaneously. As long as b new bits arrive at the encoder, the transmitter will continue to transmit data. Therefore, this scheme can be used with as large a frame length as possible and so we will not impose any restriction on the frame length. It should be noted, however, that our assumption of quasistatic fading comes from the fact that mobile applications are generally low-delay applications. Since using long data frames causes long decoding delays, such an action would not be consistent with our assumptions.

Next, we describe the convolutional encoder in more detail.

2.2.3 Structure of the Encoder

Figure 2.1 illustrates all the required elements in a space-time trellis encoder. However, the details of the pulse shaper and the modulator will not be of interest for us. Rather, we are interested in the mathematical relation between the input data bits and the output symbols belonging to the constellation alphabet of the selected modulation. For this reason, we consider the space-time trellis encoder as consisting of a single element which takes as input binary data and outputs constellation symbols. We illustrate this in Figure 2.3.

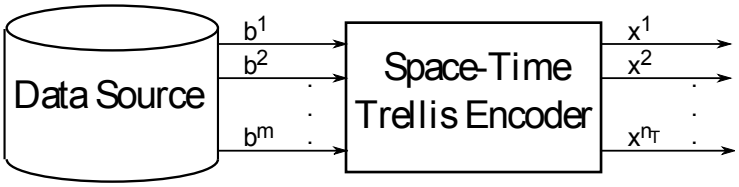

Figure 2.3: An input-output model of the space-time trellis encoder.

Before moving into details about the inner structure of the encoder, let us describe the input and the output explicitly. The input consists of m binary sequences of length l where l is the frame length. We write this as

$$\boldsymbol{b}^i = (b_1^i, b_2^i, \ldots, b_l^i) \tag{2.7}$$

for $i = 1, 2, \ldots, m$. Thus, the input can be viewed as consisting of m branches of data each of which moves through a unique part of the encoder. In this context m is called the *(transmission) rate* of the space-time trellis code. As for the output, it is composed of n_T sequences of constellation symbols, which are

$$\boldsymbol{x}^i = (x_1^i, x_2^i, \ldots, x_l^i) \tag{2.8}$$

for $i = 1, 2, \ldots, n_T$. The sequence $\boldsymbol{x}^i$ contains the symbols to be transmitted from transmit antenna i over l successive time slots. The following description holds for any modulation whose elements are suitably labeled, still we will assume M-PSK modulation whose elements are represented with the integer labeling as explained in the previous subsection.

We are now ready to describe the structure of the encoder. The encoder contains m shift registers. Let us denote the number of memory elements in shift register i by v_i. These have the property that $v_i \leq v_j$ for $i < j$ and $v_m - v_1 \leq 1$. By the *state* of a shift register at time slot t we mean the bits contained by its memory elements written in an ordered manner. Explicitly, if we denote the respective bits contained by the memory elements of shift register i at time slot t by $c_1^i, c_2^i, \ldots, c_{v_i}^i$, then the state of shift register i at time slot t is given by $\boldsymbol{s}_t^i = (c_1^i, c_2^i, \ldots, c_{v_i}^i)$. At the end of each time slot, the bit in the rightmost memory element is discarded from the encoder, bits in the other memory elements are shifted one unit to the right and the incoming bit is fed into the leftmost memory element. Thus, if we denote by b_t^i the incoming bit from

data subsequence $\boldsymbol{b}^i$ at time t, then the state of the encoder at time slot $t+1$ will be $\boldsymbol{s}^i_{t+1} = (b^i_t, c^i_1, \ldots, c^i_{v_i-1})$. Using this recursive definition for the state of a shift register and the description of the data subsequences $\boldsymbol{b}^i$ given in the previous paragraph, we can represent the state of the shift register i at time slot t in terms of the elements of $\boldsymbol{b}^i$ by

$$\boldsymbol{s}^i_t = (b^i_{t-1}, b^i_{t-2}, \ldots, b^i_{t-v_i}). \tag{2.9}$$

One restriction we impose is that the shift registers of the encoder should be in the all-zero state at the beginning and at the end of each frame. For this reason, for $i = 1, 2, \ldots, m$ we define $b^i_0 = b^i_{-1} = \ldots = b^i_{1-v_i} = 0$ and it turns out that $b^i_{l-v_i+1} = b^i_{l-v_i+2} = \ldots = b^i_l = 0$.

Let us now describe the output of the encoder at time slot t, given the input bits $\boldsymbol{b}_t = (b^1_t, b^2_t, \ldots, b^m_t)$ and the states of the shift registers at time t. Each input bit b^i_t and the content of each memory element can have an effect on each output symbol x^j_t for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n_T$. Let us define

$$I^* = \bigcup_{i=1}^{m} \bigcup_{j=0}^{v_i} \{(i, j)\}, \;\; N = \{1, 2, \ldots, n_T\}. \tag{2.10}$$

Here I^* is an index set consisting of integer pairs. The first coordinates of these pairs correspond to shift registers and the second correspond to positions within them, where the newcoming bits are given index 0. The relation between these encoder positions and the output symbols can be specified by a function of the form

$$g : I^* \times N \longrightarrow \{0, 1, \ldots, M-1\}. \tag{2.11}$$

To make it more clear, let us define $g^i_{j,k} := g((i, j), k)$ to be the *multiplying factor* associated with the encoder position (i, j) as explained above, which will be used to compute the value of output symbol x^k_t. Since the information bit contained in the encoder position (i, j) at time t is b^i_{t-j}, this means that the contribution of b^i_{t-j} to the value of x^k_t is equal to $g^i_{j,k} b^i_{t-j}$. Using the integer notation for the output symbols, we can write this shortly as

$$x^k_t = \sum_{i=1}^{m} \sum_{j=0}^{v_i} g^i_{j,k} b^i_{t-j} \pmod{M} \tag{2.12}$$

for $k = 1, 2, \ldots, n_T$. Alternatively, we can express (2.12) as the multiplication of certain matrices. Let $v = \sum_{i=1}^{m} v_i$ be the total number of memory elements in the

encoder. For $i = 1, 2, \ldots, m$ we define

$$\mathbf{u}_t^i = (b_t^i, b_{t-1}^i, \ldots, b_{t-v_i}^i) \tag{2.13}$$

to be the vector of length $1 + v_i$ consisting of the input bits from the i^{th} information branch which have an effect on the output at time t. Then we simply concatenate the elements of these vectors to obtain

$$\mathbf{u}_t = \begin{bmatrix} \mathbf{u}_t^1 & \mathbf{u}_t^2 & \ldots & \mathbf{u}_t^m \end{bmatrix}, \tag{2.14}$$

which is a $1 \times (v+m)$ row matrix consisting of all the input bits that affect the output at time t. Similarly we can gather all the multiplying factors in a single matrix as follows: In $\mathbf{u}_t$, we replace every entry by its multiplying factor corresponding to the output symbol x_t^k, i.e. we replace b_{t-j}^i by $g((i,j),k)$. Doing this for all $k = 1, 2, \ldots, n_T$, we obtain n_T row matrices of size $1 \times (v+m)$. Taking transpose of these gives us n_T column matrices $\mathbf{g}^1, \mathbf{g}^2, \ldots, \mathbf{g}^{n_T}$ of size $(v+m) \times 1$, where $\mathbf{g}^i$ consists of all the multiplying factors contributing to the value of x_t^i, written in an ordered manner. Then, like we did in the construction of $\mathbf{u}_t$, we simply concatenate the column matrices $\mathbf{g}^i$ to obtain a $(v+m) \times n_T$ matrix

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}^1 & | & \mathbf{g}^2 & | & \ldots & | & \mathbf{g}^{n_T} \end{bmatrix}. \tag{2.15}$$

In this setting, (2.12) together with the constructions (2.14) and (2.15) imply that

$$\mathbf{X}_t = \mathbf{u}_t \mathbf{G}, \tag{2.16}$$

where

$$\mathbf{X}_t = \begin{bmatrix} x_t^1 & x_t^2 & \ldots & x_t^{n_T} \end{bmatrix} \tag{2.17}$$

is the $1 \times n_T$ row matrix consisting of the output symbols at time t. $\mathbf{G}$ is called the *generator matrix* and it describes the space-time trellis code uniquely.

Figure 2.4 depicts the relation between the input bits and the output symbols in a space-time trellis encoder. The fact that each encoder position has n_T multiplying factors associated with it is indicated by labeling the multipliers with a vector having n_T entries. In fact, each multiplier should be thought of as n_T different multipliers, each having only one multiplying factor corresponding to a unique output symbol. The same is true for the adder(located to the right), which should be viewed as n_T

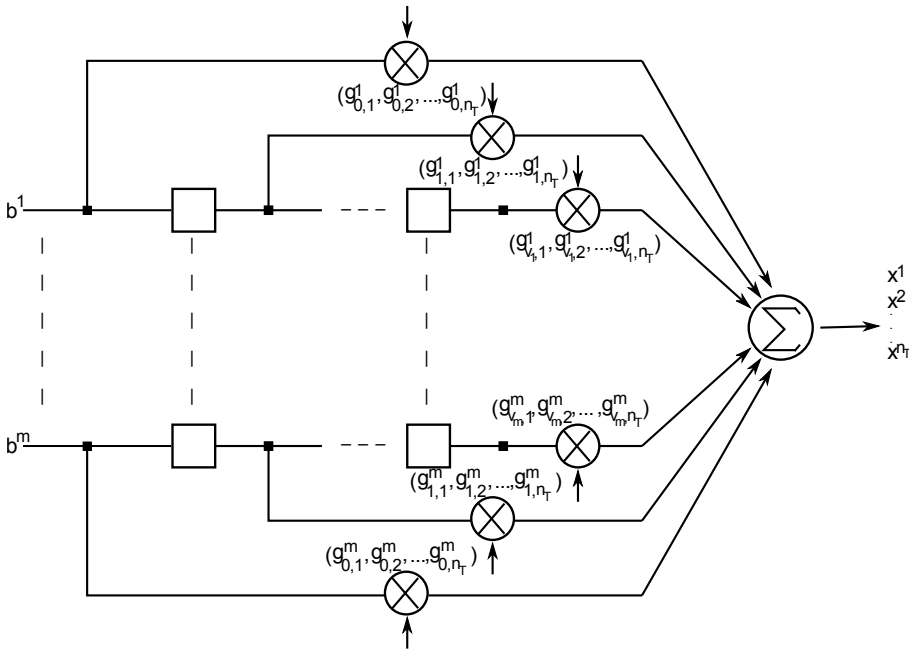

Figure 2.4: A schematic description of a space-time trellis encoder.

separate adders each of which corresponds to a unique output symbol. Then adder i only accepts the terms containing multiplying factors that correspond to the i^{th} output symbol. This can be better understood by looking at (2.12).

Although (2.16) gives a concise description of a space-time trellis encoder, it does not enable us to observe at a glance the transitions between different states of the shift registers. Given a space-time trellis code and its memory content at a certain time t, the set of output symbols can be viewed as a function of the set $(\mathbb{F}_2)^m$ of all the binary vectors of size m, or equivalently, the set of all possible values for the input $\boldsymbol{b}_t$ at time t. A visual representation of this aspect of space-time trellis codes is provided by *trellis diagrams*. Starting from $t = 1$, a trellis diagram combines each possible *state* of a convolutional encoder at time t to those ones which can immediately arise from that state. By this way, a trellis diagram explicitly shows the passage of time.

In order to proceed further, we have to make a precise definiton of the state of an encoder. Just like the state of a shift register consists of the memory elements of that shift register, the state $\boldsymbol{S}_t$ of an encoder consists of its memory content at time t. The difference comes from the order in which we write the memory elements, which we defined as being from left to right in (2.9). This time we define the ordering of

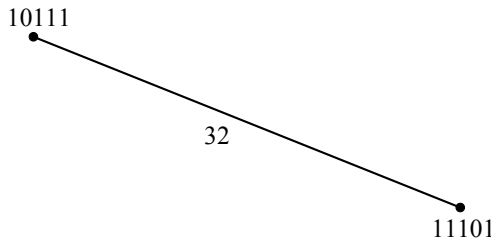

Figure 2.5: A sample transition taken from the trellis diagram of a space-time trellis encoder with 32 states and 2 transmit antennas where the modulation is the 4-PSK modulation.

the elements from right to left and then from top to bottom. To express this more precisely, we recall that

$$B_t = \bigcup_{i=1}^{m} \bigcup_{j=1}^{v_i} \{b_{t-j}^i\} \tag{2.18}$$

is the set of all memory contents of the encoder at time t. Let $b_{t-j_1}^{i_1}$ and $b_{t-j_2}^{i_2}$ be two elements of B_t. Then if $j_2 > j_1$, $b_{t-j_2}^{i_2}$ appears before $b_{t-j_1}^{i_1}$ in $\boldsymbol{S}_t$; while if $j_2 < j_1$, $b_{t-j_1}^{i_1}$ comes before $b_{t-j_2}^{i_2}$ in $\boldsymbol{S}_t$. If $j_1 = j_2$, then again there are two possibilities. If $i_2 > i_1$, $b_{t-j_1}^{i_1}$ comes before $b_{t-j_2}^{i_2}$ in $\boldsymbol{S}_t$; while if $i_2 < i_1$, $b_{t-j_2}^{i_2}$ appears before $b_{t-j_1}^{i_1}$ in $\boldsymbol{S}_t$. This explanation defines $\boldsymbol{S}_t$ uniquely. Given $\boldsymbol{S}_t$, each possible value of the input bits $\boldsymbol{b}_t$ at time t results in a different value for the next state $\boldsymbol{S}_{t+1}$ of the encoder. Since there are a total of 2^m possibilities for $\boldsymbol{b}_t$, this means that the number of possible next states $\boldsymbol{S}_{t+1}$ is equal to 2^m. In a trellis diagram, states are represented by dots and possible transitions between states are represented by line segments joining them. Each transition has a label associated with it, namely the output produced by the encoder during that transition. Since the input causing a specific transition can be read of from the last m entries of the new state, there is no need to specify it on the diagram. The reader is referred to [17] for a more mathematical discussion of states.

Figure 2.5 illustrates a sample transition taken from a space-time trellis code with $v_1 = 2$ and $v_2 = 3$. The number of states is equal to $2^v = 32$. For ease of illustration, states of this encoder are represented by binary streams of length 5 instead of binary vectors of the same length. In the figure, the encoder moves from state 10111 to state 11101 and outputs the 4-PSK symbol 3 for the first transmit antenna, and 2 for the second transmit antenna. This is indicated by placing "32" just under the line segment joining the two states, although a different approach may be required for a complete trellis diagram. The input bits causing this transition can be understood by

looking at the last two bits of the new state, namely "01". Thus, the input bit from the first information branch is 0 and the one from the second is 1. We will talk about trellis diagrams in more detail in Section 2.5.

2.2.4 Decoding of Space-Time Trellis Codes

Let us assume that the output of a space-time trellis encoder is the codeword

$$\boldsymbol{c} = (c_1^1, c_1^2, \ldots, c_1^{n_T}, c_2^1, c_2^2, \ldots, c_2^{n_T}, \ldots, c_l^1, c_l^2, \ldots, c_l^{n_T}) \tag{2.19}$$

where c_t^i is the symbol transmitted by transmit antenna i at time t. Then the symbol received by receive antenna j at time t is

$$r_t^j = \eta_t^j + \sum_{i=1}^{n_T} \alpha_{i,j} c_t^i \sqrt{E_s} \tag{2.20}$$

for $j = 1, 2, \ldots, n_R$. In our model, the channel gains $\alpha_{i,j}$ from transmit antenna i to receive antenna j are assumed to be known by the receiving party. Another assumption is that the noise values η_t^j are independent from each other. Under this scenario, the decoding problem of a space-time trellis code is to determine the transmitted codeword $\boldsymbol{c}$ from the received symbols r_t^j.

Since the noise variables are selected from a zero-mean normal distribution, noise values with small magnitudes are more likely to appear than those with large magnitudes. Furthermore, since noise variables are independent over time, it may appear logical to determine the output symbols $c_t^1, c_t^2, \ldots, c_t^{n_T}$ in a particular time slot t by just looking at the symbols $r_t^1, r_t^2, \ldots, r_t^{n_R}$ received in the same time slot, and to determine $\boldsymbol{c}$ by combining the results for all time slots $t = 1, 2, \ldots, l$. This goes as follows: If we denote by $\mathcal{M}$ the constellation alphabet with M elements, then for each $\boldsymbol{q} = (q_1, q_2, \ldots, q_{n_T}) \in \mathcal{M}^{n_T}$ and for each $j = 1, 2, \ldots, n_R$ we calculate

$$\boldsymbol{q}[j] = \sum_{i=1}^{n_T} \alpha_{i,j} q_i, \tag{2.21}$$

the symbol that would be received by receive antenna j in the absence of noise, in case $\boldsymbol{q}$ is the encoder's output at time t. Here we ignore the scaling factor $\sqrt{E_s}$, which is common to all received symbols. Then, at a particular time t, the noise associated with receive antenna j corresponding to an estimation $\boldsymbol{q}$ for the output is equal to

$$\eta_t^j(\boldsymbol{q}) = r_t^j - \boldsymbol{q}[j] \tag{2.22}$$

for $j = 1, 2, \ldots, n_R$. In view of this observation, it is not plausible to try to minimize the magnitude $\left|\eta_t^j(\boldsymbol{q})\right|$ for all j's because the independence of the path gains corresponding to different receive antennas does not allow simultaneous minimization of them. Therefore, we may instead attempt to minimize the sum of these magnitudes, namely

$$\sum_{j=1}^{n_R} \left| r_t^j - \sum_{i=1}^{n_T} \alpha_{i,j} q_i \right| \tag{2.23}$$

over all possible outputs $\boldsymbol{q} = (q_1, q_2, \ldots, q_{n_T}) \in \mathcal{M}^{n_T}$ and decide in favour of the output vector minimizing this sum for each time slot $t = 1, 2, \ldots, l$. This approach fails for the following reason: Suppose that we have decoded all the received symbols up to a particular time slot t_1 according to this rule. At the beginning of t_1, the encoder is in a particular state $\boldsymbol{S}_{t_1}$. Suppose now that our decoding rule decides in favour of a specific output $\boldsymbol{q}_{t_1}$ for time slot t_1, from the received symbols $r_{t_1}^1, r_{t_1}^2, \ldots, r_{t_1}^{n_R}$. In the trellis diagram of the code, none of the 2^m transitions leaving the encoder's state $\boldsymbol{S}_{t_1}$ in time slot t_1 may correspond to $\boldsymbol{q}_{t_1}$. In such a case, our decoding rule fails because its result is not a valid codeword. Therefore, our approach of treating each time slot independently and combining the results of all time slots does not work.

The solution of the decoding problem lies in recognition of the fact that the outputs of the encoder corresponding to different time slots are not independent entities. Therefore, rather than trying to minimize the sum of magnitudes of noise vectors for each time slot separately, we treat this quantity as a *branch metric* and try to minimize the sum of this branch metric. More explicitly, corresponding to each transition labeled $c_t^1 c_t^2 \ldots c_t^{n_T}$ we associate the quantity

$$\sum_{i=j}^{n_R} \left| r_t^j - \sum_{i=1}^{n_T} \alpha_{i,j} c_t^i \right|^2, \tag{2.24}$$

where the Euclidean distance is squared this time. Since codewords of a code are in a one-to-one correspondence with the paths in the trellis diagram, the path with the smallest accumulated metric with respect to (2.24) gives the decision of the decoder. In other words, we accept the codeword which minimizes the sum

$$\sum_{t=1}^{l} \sum_{j=1}^{n_R} \left| r_t^j - \sum_{i=1}^{n_T} \alpha_{i,j} c_t^i \right|^2 \tag{2.25}$$

over all possible values of $\boldsymbol{c}$, as the transmitted codeword. The determination of the

path with the smallest accumulated metric is achieved by use of the *Viterbi algorithm*. For details on the Viterbi decoding algorithm see [11, 12].

Lastly, we note that the original information bits can be recovered from the transmitted codeword in a straightforward manner: Each transition in the final codeword's path has a corresponding new state. Last m bits of the new state at the end of each transition gives the information bits which caused that transition. Doing this for all the transitions and combining the results, we can obtain the original information bits. We end our discussion on the decoding of space-time trellis codes here.

2.3 Design Criteria for Space-Time Trellis Codes

In the previous section, we described the structure of a space-time trellis code. However, we have not yet provided any criteria with respect to which different codes can be compared to each other. Therefore, currently we do not have any means to classify some codes as "good" codes. In this section, this issue will be addressed. We exhibit three design criteria for space-time trellis codes, two of which were first derived in [1].

Let us begin by recalling the channel parameters which are of import for us. The channel fading is modeled as Rayleigh fading in which the path gains $\alpha_{i,j}$ are independent for different (i, j) pairs and assumed to be constant during a frame of length l. The noise values η_t^j associated with receive antenna j at time t are assumed to be independent samples of a complex Gaussian random variable with mean equal to 0 and variance equal to $N_0/2$ for both dimensions.

Suppose that, under this scenario, we are making data transmission using a space-time trellis encoder with n_T transmit and n_R receive antennas, where the path gains are known by the receiver. Suppose further that the data at the receiver is being decoded with a maximum-likelihood decoder. There is a positive probability that the decoder erroneously decides in favour of a codeword

$$\boldsymbol{e} = e_1^1 e_1^2 \dots e_1^{n_T} e_2^1 e_2^2 \dots e_2^{n_T} \dots e_l^1 e_l^2 \dots e_l^{n_T} \tag{2.26}$$

when the actually transmitted codeword is

$$\boldsymbol{c} = c_1^1 c_1^2 \dots c_1^{n_T} c_2^1 c_2^2 \dots c_2^{n_T} \dots c_l^1 c_l^2 \dots c_l^{n_T}, \tag{2.27}$$

where we denote the codewords as a sequence of code symbols for ease of illustration. In [1], this probability is upper bounded as

$$P(\boldsymbol{c} \to \boldsymbol{e}) \leq \exp(-d^2(\boldsymbol{c}, \boldsymbol{e})E_s/4N_0), \tag{2.28}$$

where $\sqrt{E_s}$ is the scaling factor in (2.5) and

$$d^2(\boldsymbol{c}, \boldsymbol{e}) = \sum_{t=1}^{l} \sum_{j=1}^{n_R} \left| \sum_{i=1}^{n_T} \alpha_{i,j}(c_t^i - e_t^i) \right|^2. \tag{2.29}$$

If we denote by $\eta_t^j(\boldsymbol{x})$ the noise value associated with receive antenna j at time t corresponding to a codeword estimate $\boldsymbol{x}$, then in view of (2.21) and (2.22) we can write (2.29) as

$$d^2(\boldsymbol{c}, \boldsymbol{e}) = \sum_{t=1}^{l} \sum_{j=1}^{n_R} \left| \eta_t^j(\boldsymbol{e}) - \eta_t^j(\boldsymbol{c}) \right|^2. \tag{2.30}$$

The description of the decoder made in Section 2.2.4 implies that

$$\sum_{t=1}^{l} \sum_{j=1}^{n_R} \left| \eta_t^j(\boldsymbol{e}) \right|^2 \leq \sum_{t=1}^{l} \sum_{j=1}^{n_R} \left| \eta_t^j(\boldsymbol{c}) \right|^2. \tag{2.31}$$

In other words, the actually transmitted codeword $\boldsymbol{c}$ is not the codeword that corresponds to the path with the smallest accumulated metric. As $d^2(\boldsymbol{c}, \boldsymbol{e})$ becomes large, the decoder's decision of $\boldsymbol{e}$ instead of $\boldsymbol{c}$ becomes more unlikely. This gives an intuitional explanation of the upper bound (2.28).

In order to derive design criteria from the upper bound (2.28), an analysis is carried out in [1]. Especially important in this analysis are the $n_T \times l$ matrix

$$\mathbf{B}(\boldsymbol{c}, \boldsymbol{e}) = \begin{bmatrix} c_1^1 - e_1^1 & c_2^1 - e_2^1 & \dots & c_l^1 - e_l^1 \\ c_1^2 - e_1^2 & c_2^2 - e_2^2 & \dots & c_l^2 - e_l^2 \\ \vdots & \vdots & \ddots & \vdots \\ c_1^{n_T} - e_1^{n_T} & c_2^{n_T} - e_2^{n_T} & \dots & c_l^{n_T} - e_l^{n_T} \end{bmatrix} \tag{2.32}$$

and the $n_T \times n_T$ square matrix $\mathbf{A}(\boldsymbol{c}, \boldsymbol{e}) = \mathbf{B}(\boldsymbol{c}, \boldsymbol{e})\mathbf{B}^*(\boldsymbol{c}, \boldsymbol{e})$, where * denotes the transpose conjugate. Let us now stop for a moment and present some properties of these matrices which are relevant to our analysis.

First, we introduce some new notation to simplify matters. By d_t^i we mean $c_t^i - e_t^i$, the (i, t) entry of $\mathbf{B}(\boldsymbol{c}, \boldsymbol{e}) = \mathbf{B}$. $\boldsymbol{d}^i$ denotes the $1 \times l$ row matrix containing elements in the i^{th} row of $\mathbf{B}$ and $\boldsymbol{d}_j$ denotes the column matrix containing elements in the j^{th}

column of $\mathbf{B}$. Similarly $\boldsymbol{a}^i$ denotes the $1 \times n_T$ row matrix containing elements in the i^{th} row of $\mathbf{A}(\boldsymbol{c}, \boldsymbol{e}) = \mathbf{A}$. From its definition, the (i, j) entry of $\mathbf{A}$ is equal to

$$\mathbf{A}_{ij} = \boldsymbol{d}^i (\boldsymbol{d}^j)^*. \tag{2.33}$$

Now, we observe that $\mathbf{B}$ and $\mathbf{A}$ have equal rank. To show this let us assume

$$\sum_{j=1}^{k} \beta_{i_j} (d_1^{i_j}, d_2^{i_j}, \ldots, d_l^{i_j}) = \mathbf{0} \tag{2.34}$$

is a nontrivial linear relation satisfied by k rows of $\mathbf{B}(\boldsymbol{c}, \boldsymbol{e})$ having indices in the set $I = \{i_1, i_2, \ldots, i_k\}$ and $\mathbf{0}$ is the zero vector of length l. We denote by $d_{j \to I}$ the $k \times 1$ column matrix formed by taking only those entries of $\boldsymbol{d}_j$ which are also in d^i for some $i \in I$. If we collect the (complex) scalars in a single $1 \times k$ row matrix $\boldsymbol{\beta}$, then we can express this relation by l different equalities as

$$\boldsymbol{\beta}\, \boldsymbol{d}_{1 \to I} = 0, \boldsymbol{\beta}\, \boldsymbol{d}_{2 \to I} = 0, \ldots, \boldsymbol{\beta}\, \boldsymbol{d}_{l \to I} = 0. \tag{2.35}$$

Let us now consider the linear combination of k rows of $\mathbf{A}$ given by

$$\boldsymbol{a}^I = \beta_{i_1} \boldsymbol{a}^{i_1} + \beta_{i_2} \boldsymbol{a}^{i_2} + \ldots + \beta_{i_k} \boldsymbol{a}^{i_k}, \tag{2.36}$$

which is a vector of size n_T. Then using (2.33) and the distributive property of matrix multiplication over addition we can express the j^{th} entry of $\boldsymbol{a}^I = (a_1^I, a_2^I, \ldots, a_{n_T}^I)$ as follows:

$$\begin{aligned} a_j^I &= \beta_{i_1} \mathbf{A}_{i_1 j} + \beta_{i_2} \mathbf{A}_{i_2 j} + \ldots + \beta_{i_k} \mathbf{A}_{i_k j} \\ &= \beta_{i_1} \boldsymbol{d}^{i_1} (\boldsymbol{d}^j)^* + \beta_{i_2} \boldsymbol{d}^{i_2} (\boldsymbol{d}^j)^* + \ldots + \beta_{i_k} \boldsymbol{d}^{i_k} (\boldsymbol{d}^j)^* \\ &= (\beta_{i_1} \boldsymbol{d}^{i_1} + \beta_{i_2} \boldsymbol{d}^{i_2} + \ldots + \beta_{i_k} \boldsymbol{d}^{i_k}) (\boldsymbol{d}^j)^*. \end{aligned} \tag{2.37}$$

Here the term $\beta_{i_1} \boldsymbol{d}^{i_1} + \beta_{i_2} \boldsymbol{d}^{i_2} + \ldots + \beta_{i_k} \boldsymbol{d}^{i_k}$ in the last expression is a row matrix $\boldsymbol{d}^I = [d_1^I\ d_2^I\ \ldots\ d_l^I]$ having dimension $1 \times l$ whose j^{th} entry is equal to

$$\begin{aligned} d_j^I &= \beta_{i_1} d_j^{i_1} + \beta_{i_2} d_j^{i_2} + \ldots + \beta_{i_k} d_j^{i_k} \\ &= \boldsymbol{\beta}\, \boldsymbol{d}_{j \to I} \\ &= 0 \end{aligned} \tag{2.38}$$

in view of (2.35). Thus, $\boldsymbol{d}^I$ is the zero matrix of size $1 \times l$. Substituting in (2.37) gives

$$a_j^I = \mathbf{0} (\boldsymbol{d}^j)^* = 0 \tag{2.39}$$

for all $j = 1, 2, \ldots, n_T$. Since all its entries are zero, $\boldsymbol{a}^I$ is the zero vector of size n_T. Therefore, (2.36) defines a linear dependence of k rows of $\mathbf{A}$. This means that $\text{rank}(\mathbf{B}) \geq \text{rank}(\mathbf{A})$. In a similar manner we can show that $\text{rank}(\mathbf{B}) \leq \text{rank}(\mathbf{A})$. This proves that the ranks of $\mathbf{B}$ and $\mathbf{A}$ are equal, say r.

Next, we establish some observations related to the eigenvalues of $\mathbf{A}$. Since $\mathbf{A} = \mathbf{A}^*$ and $\mathbf{A}$ has a square root –which is $\mathbf{B}$–, a result from linear algebra (see e.g. [13, Chapter 4]) states that $\mathbf{A}$ has exactly r nonzero eigenvalues, which are all positive real numbers, say $\lambda_1, \lambda_2, \ldots, \lambda_r$. Also related to our discussion are the principal $r \times r$ submatrices of $\mathbf{A}$, which are the $r \times r$ matrices consisting of the elements lying in the intersection of a set of r rows with a set of r columns having the same indices. Since r rows can be selected out of n_T rows in $\binom{n_T}{r}$ different ways, there are $\binom{n_T}{r}$ principal submatrices of $\mathbf{A}$. The determinant of a principal submatrix is called a principal minor. The sum of the $r \times r$ principal minors of $\mathbf{A}$ is denoted by $E_r(\mathbf{A})$. According to another result in linear algebra (see [13, page 42]), the eigenvalues of a matrix are related to $E_r(\mathbf{A})$. In our context, $\lambda_1 \lambda_2 \ldots \lambda_r = E_r(\mathbf{A})$. We continue to summarize the analysis in [1].

After a series of mathematical manipulations (see [1, pages 5-6] for details) the upper bound (2.28) can be expressed as

$$P(\boldsymbol{c} \to \boldsymbol{e}) \leq \left(\prod_{i=1}^{r} \lambda_i \right)^{-n_R} (E_s/4N_0)^{-rn_R}. \tag{2.40}$$

It should be pointed out that this is just the probability that the received symbols are wrongly decoded as $\boldsymbol{e}$ given that $\boldsymbol{c}$ is transmitted. The total probability of decoding error in case $\boldsymbol{c}$ is transmitted is equal to

$$P_e(\boldsymbol{c}) = \sum_{\boldsymbol{x} \in \mathcal{C}, \boldsymbol{x} \neq \boldsymbol{c}} P(\boldsymbol{c} \to \boldsymbol{x}) \tag{2.41}$$

where $\mathcal{C}$ is the set of all possible codewords. If $P(\boldsymbol{x})$ is the probability that $\boldsymbol{x}$ is transmitted, then

$$P_e(\mathcal{C}) = \sum_{x \in \mathcal{C}} P(\boldsymbol{x}) P_e(\boldsymbol{x}) \tag{2.42}$$

is the overall probability of decoding error. Note that if all codewords are equally likely, then $P_e(\mathcal{C}) = P_e(\boldsymbol{x})$ for any $\boldsymbol{x} \in \mathcal{C}$. In the light of these observations, assuming $n_T \leq l$, (2.40) we are left with the following two design criteria[1]:

- *Rank Criterion*: In order to achieve the maximum possible diversity advantage of $n_T n_R$, the matrices $\mathbf{B}(\boldsymbol{c}, \boldsymbol{e})$ has to be full rank for all distinct codeword pairs $\boldsymbol{c}$ and $\boldsymbol{e}$. If $\min_{(\boldsymbol{c},\boldsymbol{e})\in\mathcal{C}^2,\boldsymbol{c}\neq\boldsymbol{e}} \{\text{rank}[\mathbf{B}(\boldsymbol{c}, \boldsymbol{e})]\}$ is r, then a diversity advantage of rn_R is achieved. This minimum rank is referred to as the rank of the code and the code is said to be an *r-space-time trellis code*. This criterion is also known as the *diversity criterion*.

- *Determinant Criterion*: If the rank of the code is equal to r, the quantity $\min_{(\boldsymbol{c},\boldsymbol{e})\in\mathcal{C}^2,\boldsymbol{c}\neq\boldsymbol{e}} E_r(\mathbf{A}(\boldsymbol{c}, \boldsymbol{e}))$ should be maximized. If the minimum rank is the maximum possible n_T, then this corresponds to minimizing the determinant of $\mathbf{A}(\boldsymbol{c}, \boldsymbol{e})$ over all pairs of distinct codewords $\boldsymbol{c}$ and $\boldsymbol{e}$.

It should be noted that these two design criteria are valid under the assumption of quasistatic, flat Rayleigh fading where the fade coefficients $\alpha_{i,j}$ are independent from each other. Other fading scenarios yield possibly different design criteria. See [1, pages 6-8] for details.

It was shown in [14] that as the number n_R of receive antennas approaches infinity, the impact of fading on the performance tends to disappear. Following this observation, for codes satisfying $rn_R > 3$[1], the following design criterion was derived[14]:

- *Trace Criterion*: In order to have a maximum coding advantage, the sum of eigenvalues of the matrices $\mathbf{A}(\boldsymbol{c}, \boldsymbol{e})$ should be maximized over all pairs of distinct codewords $\boldsymbol{c}$ and $\boldsymbol{e}$. But the sum of eigenvalues of a matrix is equal to its trace. Thus, the sum

$$\begin{aligned} tr(\mathbf{A}) &= \sum_{i=1}^{n_T} \mathbf{A}_{ii} \\ &= \sum_{i=1}^{n_T} \sum_{t=1}^{l} |c_t^i - e_t^i|^2 \end{aligned} \tag{2.43}$$

  should be maximized over all pairs of distinct codewords $\boldsymbol{c}$ and $\boldsymbol{e}$.

Note that (2.43) is just the squared Euclidean distance of the codewords $\boldsymbol{c}$ and $\boldsymbol{e}$. When compared to (2.29), it is seen that the path gains no longer have an effect

[1] The condition that $rn_R > 3$ is not based solely on a mathematical analysis. Simulation results show that, as long as $rn_R > 3$, best codes based on the trace criterion outperform those based on the determinant criterion. Therefore 3 is commonly accepted as the boundary value in the literature.

on the probability of decoding error. In other words, it is now the distance between transmitted codewords that affect the probability of decoding error, not the distance between received symbols. When the product of the rank r of a code and the number n_R of receive antennas exceeds 3, the trace criterion is used instead of the determinant criterion.

2.4 Some Bounds on Code Parameters

Given a space-time trellis code, we would like its performance specified by some parameters to be as high as possible. For example, it is desirable to have a high data rate so that we can encode as many information bits as possible in a fixed time duration. Furthermore, in order to limit the probability of decoding error, in view of the rank criterion presented in Section 2.3, we would like the rank of the code to be as large as possible. We would also like the encoder to be simple so that decoding process is computationally efficient. It is not surprising that there exist tradeoffs between various aspects of a space-time trellis code. Some of these which were first established in [1] are the subject of this section.

We begin by presenting a bound that relates the rate of a space-time code to the size of the constellation alphabet. Let us assume that the size of the constellation alphabet $\mathcal{M}$ is equal to $M = 2^p$ for some positive integer p.

Theorem 2.4.1 ([1])**.** *Consider an r-space-time trellis code with n_T transmit antennas. Let l be the number of data frames. Then, the rate R of the code is upper bounded as*

$$R \leq \frac{\log[A_{M^l}(n_T, r)]}{l} \tag{2.44}$$

in bits per second per Hertz, where $A_{M^l}(n_T, r)$ is the maximum possible size of a code defined over an alphabet of size M^l and having block length equal to n_T and minimum Hamming distance equal to r.

Proof. The output of the encoder $(c_t^1, c_t^2, \ldots, c_t^{n_T})$ at time t has on average at most 2^R differrent values over a data frame of length l. Therefore, the set $\mathcal{C}$ of all codewords contains at most $(2^R)^l = 2^{lR}$ elements. The function f from $\mathcal{M}^{ln_T}$ to $(\mathcal{M}^l)^{n_T}$ which

maps the codeword

$$(c_1^1, c_1^2, \ldots, c_1^{n_T}, c_2^1, c_2^2, \ldots, c_2^{n_T}, \ldots, c_l^1, c_l^2, \ldots, c_l^{n_T}) \tag{2.45}$$

to

$$((c_1^1, c_2^1, \ldots, c_l^1), (c_1^2, c_2^2, \ldots, c_l^2), \ldots, (c_1^{n_T}, c_2^{n_T}, \ldots, c_l^{n_T})) \tag{2.46}$$

is clearly one-to-one so $f(\mathcal{C})$ also has at most 2^{lR} elements. $f(\mathcal{C})$ can be viewed as a code over an alphabet of size M^l having block length equal to n_T. Furthermore, we observe that for two distinct codewords $\boldsymbol{c}$ and $\boldsymbol{e}$, the rows of $\mathbf{B}(\boldsymbol{c}, \boldsymbol{e})$ are exactly the entries of $f(\boldsymbol{c}) - f(\boldsymbol{e})$ written as a vector of length l. Thus, the assumption $\text{rank}(\mathbf{B}) \geq r$ implies that $f(\boldsymbol{c}) - f(\boldsymbol{e})$ has at least r nonzero entries. In other words, the Hamming distance between $f(\boldsymbol{c})$ and $f(\boldsymbol{e})$ is not less than r. This shows that the minimum Hamming distance of the code $f(\mathcal{C})$ is r. Therefore, the size of $f(\mathcal{C})$ cannot exceed $A_{M^l}(n_T, r)$. Thus we have

$$2^{lR} \leq A_{M^l}(n_T, r).$$

Taking logarithm and dividing by l yields (2.44). □

The proof reveals the existence of a rather simple tradeoff between the rank and the size of a space-time trellis code. The bound given in (2.44) becomes even simpler when the code is of full rank n_T. We just observe that, over an alphabet of size M^l, the repetition code of length n_T has minimum Hamming distance equal to n_T and size equal to M^l. Furthermore, if we consider any $M^l + 1$ codewords over the same alphabet, for any fixed coordinate, some two of them have to agree in that coordinate by the pigeonhole principle. This shows that $A_{M_l}(n_T, n_T) \leq M^l$. But since the repetition code achieves this bound we have $A_{M_l}(n_T, n_T) = M^l$. Thus we have proved the following:

Corollary 2.4.2 ([1])**.** *Consider a full rank space-time trellis code employing a constellation $\mathcal{M}$ of size $M = 2^p$. Its rate R satisfies $R \leq p$.*

This shows that the maximum achievable rate of a full rank space-time trellis code is determined by its constellation size. For example, if 8-PSK constellation is used, the maximum possible rate of a full rank code is 3 bits/s/Hz.

We have seen that the constellation size, the rank and the rate of a space-time trellis code are related by a tradeoff. On the other hand, we have not yet related this tradeoff to the parameters of the encoder. One important parameter of a convolutional encoder is its *constraint length* K. Constraint length has several definitions in the literature, but we take it to be the length of the longest shift register. In terms of the notation introduced in Section 2.2.3, $K := \max_i\{v_i\}$. The following relates the rank of a space-time trellis code to its constraint length.

Lemma 2.4.3 ([1])**.** *The constraint length K of an r-space-time trellis code satisfies $K \geq r - 1$.*

Proof. Consider the two information sequences $\boldsymbol{I} = (1, 0, 0, \ldots, 0)$ and $\mathbf{0}_{lm}$, the all-zero sequence of length lm. In case $\boldsymbol{I}$ is the input of the encoder, the bit 1, which is the first bit of the first information subsequence, enters the encoder at $t = 2$ and stays in the encoder for at most K time intervals. This means that it is no longer in the encoder's memory at $t = K + 2$. Therefore, the encoder will always be in the all-zero state starting from $t = K + 2$. At any time t, if the encoder is in the zero state and the input is the all-zero input, it is clear from (2.12) that the output at time t is the all-zero output. Therefore, the output $\boldsymbol{c}$ corresponding to the input $\boldsymbol{I}$ is of the form $c_1^1 c_1^2 \ldots c_1^{n_T} c_2^1 c_2^2 \ldots c_2^{n_T} \ldots c_{K+1}^1 c_{K+1}^2 \ldots c_{K+1}^{n_T} 00 \ldots 0$. For the same reason, the output $\boldsymbol{e}$ corresponding to the all-zero input sequence $\mathbf{0}_{lm}$ is the all-zero output sequence $\mathbf{0}_{ln_T}$ of length ln_T. Thus we see that $n_T - (K+1)$ rows of the $n_T \times n_T$ matrix $\mathbf{B}(\boldsymbol{c}, \boldsymbol{e})$ have all of their entries equal to 0. Hence, the rank of $\mathbf{B}(\boldsymbol{c}, \boldsymbol{e})$ is at most $K + 1$. Since r is the minimum of ranks of these matrices, we have $r \leq K + 1$. □

The above lemma is important because the complexity of most known algorithms, including the Viterbi algorithm, is known to be growing too rapidly with the constraint length. Thus, minimizing the decoding error probability turns out to be in the cost of increasing the complexity of the decoder. The following is a more precise statement of this fact:

Lemma 2.4.4 ([1])**.** *Let $\mathcal{C}$ be an r-space-time trellis code having transmission rate m. The number of states of $\mathcal{C}$ is at least $2^{m(r-1)}$.*

Proof. First of all, we note that the number of states of the code is just 2^v, where

v is the total number of memory elements. If all the shift registers are of the same length, then $v = mK$ since we have m information branches. Thus the number of states is equal to 2^{mK}, which is at least $2^{m(r-1)}$ in view of Lemma 2.4.3. If the shift registers are not of the same length, we argue differently. At any time t, there are a total of 2^m possible binary vectors which can be the input to the encoder at time t. Therefore, in the trellis diagram of the code, the number of trellis branches leaving each state is equal to 2^m. Furthermore, the argument used in the proof of Lemma 2.4.3 makes it clear that two paths leaving the zero state in time slot 1 cannot meet at the same state until time slot r. In other words, no two of the paths can intersect during the first $r-1$ time slots. This means that, just before time slot r, each path in the trellis is in a unique state. A simple counting argument shows that the number of these states is $2^{m(r-1)}$. Therefore, the number of states of $\mathcal{C}$ cannot be less than this number. □

The number of states is closely related to the *trellis complexity* of an encoder. Indeed, we take it here as the definition of trellis complexity. The above lemma shows that trellis complexity grows exponentially with the rank and rate of a space-time trellis code. Considering this lemma with (2.44), we see that there is a fundamental tradeoff between transmission rate, rank, constellation size and trellis complexity.

One natural question might be as to whether the bounds presented above are tight. Codes satisfying all the above with equality are called *optimal* with respect to the fundamental tradeoffs between transmission rate, rank, constellation size and trellis complexity. The answer is affirmative for codes with two transmit antennas. Examination of such a code will be the topic of next section.

2.5 A Code for Two Transmit Antennas

In this section we analyze in some detail a space-time trellis code for two transmit antennas, which was constructed in [1]. The constellation is the 4-PSK constellation, where the elements of $\mathbb{Z}_4$ are used to label the signal points as shown in Figure 2.2. The encoder consists of two shift registers each having a single memory element. Thus we have $m = v = 2$. Under this setting, we can describe the encoder by four pairs of

multiplying factors given by

$$(g^1_{0,1}, g^1_{0,2}) = (0,1), (g^2_{0,1}, g^2_{0,2}) = (0,2), (g^1_{1,1}, g^1_{1,2}) = (1,0), (g^2_{1,1}, g^2_{1,2}) = (2,0).$$

At time t, the incoming bit from the first information branch is b^1_t and from the second information branch is b^2_t. Hence the content of the first shift register is b^1_{t-1} and the content of the second one is b^2_{t-1}. Using (2.12), we can express the symbols transmitted from the two transmit antennas at time t by

$$x^1_t = b^1_{t-1} + 2b^2_{t-1} \pmod 4,\ x^2_t = b^1_t + 2b^2_t \pmod 4 \tag{2.47}$$

or in a more compact form by

$$(x^1_t, x^2_t) = b^1_t(0,1) + b^2_t(0,2) + b^1_{t-1}(1,0) + b^2_{t-1}(2,0). \tag{2.48}$$

Thus, at any time t, the symbol transmitted from the first transmit antenna is affected only by the information bits which were input to the encoder at time $t-1$, while the symbol transmitted from the second antenna is affected only by the input at time t.

Let us now consider the trellis diagram of the code. The trellis diagram of a convolutional code is a network of branches showing the passage of time by visualizing all possible transitions between states of the encoder. In our case, we have the extra condition that the initial and the final states of the diagram are the all-zero state. In the first time stage of the trellis diagram, all paths in the trellis diverge from the zero state. Likewise, all paths merge in the zero state at the end of the last time stage. Therefore, the trellis diagram of a space-time trellis code consists of l stages, where at the beginning of the first and at the end of the last stage the only possible state is the all-zero state. Our aim is to find a way to visualize the code in a simpler manner.

Lemma 2.5.1. *Consider a space-time trellis code $\mathcal{C}$ with constraint length K and total number of memory elements equal to v. If the frame length l is not less than $2K$, the following holds: At the end of the K^{th} stage, all possible 2^v states appear in the trellis diagram of $\mathcal{C}$.*

Proof. An equivalent formulation of the final statement in the lemma is as follows: *Starting from the zero state, after K transitions, the encoder can be in any state.* Let m denote the number of information branches. To prove the assertion, for any initial

state at time t and any given objective state $\boldsymbol{S}$, we will construct an information sequence of mK elements, which causes the encoder to be in state $\boldsymbol{S}$ at time $t+K$. The state of an encoder at time t defined in Section 2.2.3 contains v_i elements for shift register i. Its content is given by (2.18). According to this, the elements corresponding to shift register i are $b^i_{t-v_i}, b^i_{t-v_i+1}, \ldots, b^i_{t-1}$. The state of shift register i is equal to

$$\boldsymbol{s}^i_t = (b^i_{t-1}, b^i_{t-2}, \ldots, b^i_{t-v_i}). \tag{2.49}$$

Suppose now that we are at time t. Keeping this observation in mind, in order to construct an information sequence which will result in the encoder's being in state $\boldsymbol{S}$ at time $t+K$, we consider the corresponding states of the shift registers

$$\boldsymbol{s}^i_{t+K} = (b^i_{t+K-1}, b^i_{t+K-2}, \ldots, b^i_{t+K-v_i}) \tag{2.50}$$

for $i = 1, 2, \ldots, m$. We form m information sequences of length K by first reversing each of these and then preceding the reversed vector with any $K - v_i$ bits. To be deterministic, we choose all these bits to be 0. Thus,

$$(\boldsymbol{b_S})^i = (\overbrace{0, 0, \ldots, 0}^{K-v_i times}, b^i_{t+K-v_i}, b^i_{t+K-v_i+1}, \ldots, b^i_{t+K-1}) \tag{2.51}$$

for $i = 1, 2, \ldots, m$. It is clear that feeding the shift register i with $(\boldsymbol{b_S})^i$ causes the shift register i to be in state $\boldsymbol{s}^i_{t+K}$ given by (2.50) at time $t+K$. Since the state of an encoder is uniquely determined by the states of all the shift registers, the information sequence of mK elements given by

$$(\boldsymbol{b_S}) = ((b_{\boldsymbol{S}})^1_1, (b_{\boldsymbol{S}})^2_1, \ldots, (b_{\boldsymbol{S}})^m_1, \ldots, (b_{\boldsymbol{S}})^1_K, (b_{\boldsymbol{S}})^2_K, \ldots, (b_{\boldsymbol{S}})^m_K), \tag{2.52}$$

where $(b_{\boldsymbol{S}})^i_j$ denotes the j^{th} element of $(\boldsymbol{b_S})^i$, causes the encoder to be in state $\boldsymbol{S}$ at time $t+K$. Thus, we have proved that, given any initial state at time t and any objective state $\boldsymbol{S}$, there is an information sequence of length mK which, when input to the encoder, causes the encoder to be in state $\boldsymbol{S}$ at time $t+K$.

The requirement $l \geq 2K$ is necessary for the following: Although an information sequence causing a specific state can be found, due to the fact that last few bits of each information subsequence is forced to be zero(see the explanation after (2.9)), it is not certain that the encoder can be fed with this information sequence. If $l \geq 2K$, however, the encoder can be fed with any of the possible 2^{mK} information sequences, starting from the initial all-zero state. Thus, the encoder can be in any state at the end of the K^{th} stage of the trellis diagram. □

Lemma 2.5.2. *Suppose that a space-time trellis encoder with v memory elements can be in any of the 2^v states at time t. If the input at time t can have all possible values, then the encoder can also be in any of the states at time $t+1$.*

Proof. As always, let m and K denote the number of information branches and the constraint length, respectively. First, we note that the input bits feeding the encoder at a given time can be identified with elements of $\mathbb{Z}_{2^m}$. We then consider the states in which the rightmost memory content of all the m shift registers is 0. Since the number of remaining memory elements is $v-m$, there are a total of 2^{v-m} such states. Let us denote these states by $\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_{2^{v-m}}$.

Let $\boldsymbol{S}_j(i)$ denote the state of the encoder after the state $\boldsymbol{S}_j$ is fed by the input i. Then, if

$$\mathcal{S} = \bigcup_{i=1}^{2^v} \{\boldsymbol{S}_i\} \tag{2.53}$$

is the set of all possible states, to prove the lemma it is enough to show

$$\mathcal{S} = \bigcup_{j=1}^{2^{v-m}} \bigcup_{i=0}^{2^m-1} \{\boldsymbol{S}_j(i)\}. \tag{2.54}$$

To prove this, first we note that for any j, $\boldsymbol{S}_j(i)$ are all different for $i = 1, 2, \ldots, 2^m$. This follows from the fact that i can be read of from the last m entries of the next state. Thus, if $\mathcal{S}_j$ is the set of all states that can arise from $\boldsymbol{S}_j$, the number $|\mathcal{S}_j|$ of elements of $\mathcal{S}_j$ is equal to 2^m. It remains to show $\mathcal{S}_j$ are all disjoint, i.e. for $i \neq j$

$$\mathcal{S}_i \cap \mathcal{S}_j = \varnothing. \tag{2.55}$$

Let $1 \leq j < j' \leq 2^{v-m}$. $\boldsymbol{S}_j$ and $\boldsymbol{S}_{j'}$ are different states by definition. Both have the property that the rightmost memory content of every shift register is 0. Therefore, their difference comes from other memory elements. Since only the rightmost bits leave the memory during a transition, all other memory content continues to remain in the memory. Therefore, $\boldsymbol{S}_j$ and $\boldsymbol{S}_{j'}$ cannot move to a common state after a single transition, meaning that there does not exist a pair $(i, i') \in (\mathbb{Z}_{2^m})^2$ such that $\boldsymbol{S}_j(i) = \boldsymbol{S}_{j'}(i')$. This shows that (2.55) holds. Thus, the union

$$\bigcup_{j=1}^{2^{v-m}} \mathcal{S}_j \tag{2.56}$$

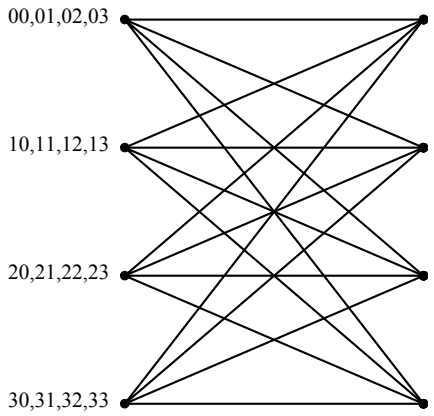

Figure 2.6: The trellis module of our code for two transmit antennas.

is disjoint and its cardinality is equal to $2^{v-m}2^m = 2^v$, the number of all possible states. Since the two sides of (2.54) have the same cardinality and the right-hand side is a subset of the left-hand side, (2.54) holds and the lemma is proved. □

The following is what these two lemmas are for:

Corollary 2.5.3. *Given a space-time trellis code with constraint length K and frame length l such that $l \geq 2K$. The portion of the trellis diagram between times K and $l - K$ consists of a regular pattern which repeats itself.*

Following [15], we call the repeated pattern mentioned in the corollary the *trellis module* of the code. The trellis module of a code shows all possible states on its left and right sides, and links the states on the left with the ones which can arise from them after a single transition. Although the trellis diagram of a code can be arbitrarily long, its structure can be described by a single trellis module except for a number of stages at the beginning and at the end of the trellis. Figure 2.6 depicts the trellis module of our code given in (2.48).

At first glance, one may not find the figure informative enough. We begin our explanation by the labeling of states, which are indicated by medium sized dots in the trellis module. Since states are just binary vectors of length v, they can naturally be mapped to binary integers from 0 up to $2^v - 1$. In the module, they appear from top to bottom in increasing order according to this integer mapping. Thus, for our case the topmost state is 00, the second state is 01, the third is 10, and the bottommost

state is 11, though not shown on the figure explicitly. In general, if the encoder is in state $s_1 s_2 \dots s_v$ and $i_1 i_2$ is the input, the next state is $s_3 s_4 \dots s_v i_1 i_2$. By this way, it is ensured that next states corresponding to a fixed initial state are kept together on the right side of the module. This fact cannot be directly observed for this code, where every state can be the next to any given state, due to the equality $2^m = 2^v$; or in words, the number of branches leaving each state is equal to the number of all possible states. For trellis modules of graphs satisfying $m < v$, one can see [1].

Next, we explain how the output corresponding to a transition is indicated in the trellis module. As explained in Section 2.2.3, the label $x_1 x_2$ corresponds to transmission of x_1 from transmit antenna 1 and transmission of x_2 from transmit antenna 2. Due to the numerosity of transition branches in the module, this time we do not label transitions as in Figure 2.5. Instead, to the left of each state on the left side, we keep a list of output labels. Each of these labels corresponds to a unique transition branch departing from that state. More precisely, the k^{th} output label from the left is the output corresponding to the k^{th} transition branch from the top. For example, if the initial state is 01, the transition corresponding to the input 10 is labeled 12. This means that, the input 10 causes the first and the second transmit antennas to transmit the 4-PSK symbols 1 and 2, respectively. The next state is 10. These can be verified by using the algebraic description (2.48) of the code.

We claim that this code is optimal with respect to the fundamental tradeoff between transmission rate, rank, constellation size and trellis complexity given in Section 2.4. The constraint length K of the code is equal to 1, total memory order v is 2 and the number of states equals $2^v = 4$. The transmission rate is 2 bits/s/Hz. The constellation consists of $4 = 2^2$ elements. The only unknown parameter is the rank r of the code. The value of r which would cause all the bounds of Section 2.4 to be satisfied with equality is easily seen to be 2. Thus, our claim is equivalent to the claim that the code is of full rank.

In order to show that the rank of the code is 2, we consider the $2 \times l$ matrix

$$\mathbf{B}(\boldsymbol{c}, \boldsymbol{e}) = \begin{bmatrix} c_1^1 - e_1^1 & c_2^1 - e_2^1 & \dots & c_l^1 - e_l^1 \\ c_1^2 - e_1^2 & c_2^2 - e_2^2 & \dots & c_l^2 - e_l^2 \end{bmatrix} \tag{2.57}$$

for distinct codewords $\boldsymbol{c}$ and $\boldsymbol{e}$. This matrix should have rank 2 for every such code-

word pairs. This can be proved by finding two columns which are linearly independent. We start the search for two such columns by noting that the paths corresponding to $\boldsymbol{c}$ and $\boldsymbol{e}$ are distinct since $\boldsymbol{c}$ and $\boldsymbol{e}$ are distinct. Both of these paths start and terminate with the all-zero state. Therefore, there must exist times t_1 and t_2 such that these two paths diverge at time t_1 and merge at time t_2. This means that the paths corresponding to these two codewords are in the same state at time t_1, and in different states at time $t_1 + 1$. Similarly, they are in different states at time $t_2 - 1$, and in the same state at time t_2. By the vector $\boldsymbol{x}_t = (x_t^1, x_t^2)$ we denote the portion of a codeword $\boldsymbol{x}$ output at time t. From Figure 2.6, we see that outputs corresponding to paths diverging from the same state differ only in the second coordinate and outputs corresponding to paths merging at the same state differ only in the first coordinate. Thus we have $c_{t_1}^1 = e_{t_1}^1, c_{t_1}^2 \neq e_{t_1}^2$ and $c_{t_2}^1 \neq e_{t_2}^1, c_{t_2}^2 = e_{t_2}^2$. Therefore, the linear relation

$$\beta_1(\boldsymbol{c}_{t_1} - \boldsymbol{e}_{t_1}) + \beta_2(\boldsymbol{c}_{t_2} - \boldsymbol{e}_{t_2}) = \boldsymbol{0} \tag{2.58}$$

does not have any nontrivial solutions. This shows that $t_1{}^{\text{th}}$ and $t_2{}^{\text{th}}$ columns of $\mathbf{B}(\boldsymbol{c}, \boldsymbol{e})$ are linearly independent, completing the proof of our claim that the code given by (2.48) is optimal with respect to the fundamental tradeoff between transmission rate, rank, constellation size and trellis complexity.

In Section 2.3, we established design criteria for space-time codes which serve as a guide in designing codes which achieve high performance. Since our code is having the maximum possible rank, which is 2, it satisfies the rank criterion. Now, we will compute the coding advantage. In our case, since the maximum possible diversity gain is achieved, the coding advantage corresponds to the minimum determinant among all the matrices $\mathbf{A}(\boldsymbol{c}, \boldsymbol{e})$ corresponding to distinct codewords $\boldsymbol{c}$ and $\boldsymbol{e}$. $\mathbf{A}(\boldsymbol{c}, \boldsymbol{e})$ can be expressed as

$$\mathbf{A}(\boldsymbol{c}, \boldsymbol{e}) = \sum_{t=1}^{l} \begin{bmatrix} |c_t^1 - e_t^1|^2 & (c_t^1 - e_t^1)(\overline{c_t^2 - e_t^2}) \\ (c_t^2 - e_t^2)(\overline{c_t^1 - e_t^1}) & |c_t^2 - e_t^2|^2 \end{bmatrix}, \tag{2.59}$$

where the horizontal bar over a symbol denotes the complex conjugate of that symbol. It is shown in [1, page 13] that the code is geometrically uniform. This means that the performance of the code is independent of the transmitted codeword and hence we can safely assume that one of the codewords, say $\boldsymbol{c}$, in $\mathbf{A}(\boldsymbol{c}, \boldsymbol{e})$ is the codeword corresponding to the all-zero information sequence. Thus, if expressed with the integer

labeling in Figure 2.2, $\boldsymbol{c}$ is the all-zero codeword. On the other hand, since we are interested in computing determinants, we take as codewords their actual complex values. Thus $\boldsymbol{c}$ is equal to $11\ldots1$, the all-one sequence of length ln_T. Substituting this in (2.59) gives

$$\mathbf{A}(\boldsymbol{c},\boldsymbol{e}) = \sum_{t=1}^{l} \begin{bmatrix} |1-e_t^1|^2 & (1-e_t^1)(\overline{1-e_t^2}) \\ (1-e_t^2)(\overline{1-e_t^1}) & |1-e_t^2|^2 \end{bmatrix}. \tag{2.60}$$

Each term of this sum corresponds to a stage of the trellis diagram. Let us denote the 2×2 matrix corresponding to time t by Θ_t. If the portion (e_t^1, e_t^2) of $\boldsymbol{e}$ which is output at time t is equal to the complex vector $(1,1)$, then Θ_t is the zero matrix of size 2×2. Otherwise Θ_t is not the zero matrix and contributes to the above sum. Let $I_t = \{t_1, t_2, \ldots, t_d\} \subseteq \{1,2,\ldots,l\}$ be the set of time instants in which the output corresponding to $\boldsymbol{e}$ is different from $(1,1)$ such that $t_i < t_j$ for $i<j$. Since $\boldsymbol{c}$ and $\boldsymbol{e}$ are distinct, $\boldsymbol{e}$ has to diverge from the zero state at time t_1 and remerge to the zero state at time t_d. Therefore, it is true that

$$\Theta_{t_1} = \begin{bmatrix} 0 & 0 \\ 0 & f \end{bmatrix},\ \Theta_{t_d} = \begin{bmatrix} s & 0 \\ 0 & 0 \end{bmatrix}, \tag{2.61}$$

where f and s are both equal to either 2 or 4. The time stages in which the trellis path corresponding to $\boldsymbol{e}$ leaves or moves to the zero state contribute to the sum (2.60) by matrices of this form. Let us denote the set of such time stages by I_0. The time stages t in which the trellis path corresponding to $\boldsymbol{e}$ neither leaves nor moves to the zero state, which we denote by I_+, contribute with a matrix of the form

$$\begin{bmatrix} a_t & b_t \\ \overline{b_t} & d_t \end{bmatrix}, \tag{2.62}$$

where a_t, d_t are nonnegative real numbers and $b\overline{b} = |b|^2 \le a_t d_t$. This can be verified by looking at (2.60) and recalling that each symbol in $\boldsymbol{e}$ is a 4-PSK symbol as illustrated in Figure 2.2. After these observations, we can write (2.60) as

$$\begin{aligned} \mathbf{A}(\boldsymbol{c},\boldsymbol{e}) &= \sum_{t\in I_0} \Theta_t + \sum_{t\in I_+} \Theta_t \\ &= \sum_{t\in I_0} \begin{bmatrix} s_t & 0 \\ 0 & f_t \end{bmatrix} + \sum_{t\in I_+} \begin{bmatrix} a_t & b_t \\ \overline{b_t} & d_t \end{bmatrix}. \end{aligned} \tag{2.63}$$

Adding the terms in the summation on the left side is easily seen to yield a matrix of the same form. Similarly, using the identity $\overline{z_1+z_2} = \overline{z_1} + \overline{z_2}$ for any two complex

numbers z_1 and z_2, and with some extra effort, the terms in the summation on the right side can be seen to preserve their structure after addition. This leaves us with

$$\mathbf{A}(\boldsymbol{c},\boldsymbol{e}) = \begin{bmatrix} S & 0 \\ 0 & F \end{bmatrix} + \begin{bmatrix} A & B \\ \overline{B} & D \end{bmatrix}, \tag{2.64}$$

where A, S, F, D are nonnegative real numbers and $|B|^2 \leq AD$. Further, since t_1 and t_d are in I_0, S and F are positive even integers. Hence, the coding advantage is given by

$$\min \left\{ \det \begin{bmatrix} S & 0 \\ 0 & F \end{bmatrix} + \begin{bmatrix} A & B \\ \overline{B} & D \end{bmatrix} \right\} \tag{2.65}$$

taken over all numbers as explained above. This determinant is equal to

$$\begin{aligned} \det \begin{bmatrix} S+A & B \\ \overline{B} & F+D \end{bmatrix} &= (S+A)(F+D) - |B|^2 \\ &= SF + SD + AF + AD - |B|^2, \end{aligned} \tag{2.66}$$

which is greater than SF since $|B|^2 \leq AD$ and F, S, A, D are nonnegative real numbers. Thus we can write

$$\min \left\{ \det \begin{bmatrix} S & 0 \\ 0 & F \end{bmatrix} + \begin{bmatrix} A & B \\ \overline{B} & D \end{bmatrix} \right\} = \min \left\{ \det \begin{bmatrix} S & 0 \\ 0 & F \end{bmatrix} \right\}. \tag{2.67}$$

This minimal value is clearly achieved when $S = F = 2$. Thus, the coding advantage of our code is equal to 4.

Although we have established design criteria for space-time trellis codes, we have not yet touched the matter of actually designing a code. The codes introduced in [1], including the example we gave here, are constructed by hand. On the other hand, the matrix representation (2.16) we derived in Section 2.2.3 gives a ground for systematic code search. Such an approach was adopted in [16]. There, for two transmit antennas, a computer search was performed over all possible generator matrices. In this case, among those yielding a code of full rank–which is 2 in this case–, the one giving the maximum coding advantage is chosen as the generator matrix of the code. By this method, codes outperforming the ones in [1] in terms of coding gain was found. See [16] for details.

CHAPTER 3

SPACE-TIME BLOCK CODES FROM ORTHOGONAL DESIGNS

The topic of the previous chapter was space-time trellis codes, a trellis coded modulation scheme over wireless channels which provides coding gain as well as diversity gain. Although they achieve high performance, space-time trellis codes have the drawback that, for a fixed number of transmit antennas, their decoding complexity grows exponentially with the transmission rate. One approach which addresses this issue is provided by *space-time block coding*. The first known examples of space-time block codes(STBC's) were introduced by Alamouti in [18] although the term was coined later in [19]. There, orthogonal designs were used to construct space-time block codes for both real and complex constellations.

Space-time block codes from orthogonal designs will constitute the subject of this chapter. Our treatment will run in parallel to the framework provided in [19]. Thus, the description of the scheme based on square orthogonal designs for real signal constellations are explained in Section 3.1. Derivation of basic results and the decoding process are also described in this section. Then, in Section 3.2, the scheme for real constellations are generalized to nonsquare case and fundamental questions related to this generalization and their answers are presented. Lastly, in Section 3.3, what has been covered in the first two sections are generalized to complex signal constellations.

3.1 STBC's from Real Orthogonal Designs

We begin our discussion on space-time block codes from orthogonal designs with a description of real orthogonal designs. First, we review the underlying system model which we are familiar to from the previous chapter. Secondly, we define real orthogonal designs and describe how they are used to construct space-time block codes. Lastly, we shortly explain the decoding of these codes.

3.1.1 System Model

Once again, $n_T > 1$ will denote the number of transmit antennas and $n_R \geq 1$ will denote the number of receive antennas. The number of data frames is l. If the symbol transmitted from transmit antenna i at time t is c_t^i, the symbol received by receive antenna j at time t equals

$$r_t^j = \eta_t^j + \sum_{i=1}^{n_T} \alpha_{i,j} c_t^i \sqrt{E_s}, \tag{3.1}$$

where this time the scaling factor $\sqrt{E_s}$ is chosen such that average energy of the transmitted symbols is $1/n_T$. As explained in the previous chapter, the noise variables η_t^j are modeled as independent samples of a Gaussian distribution and the path gains $\alpha_{i,j}$ are described by a model assuming quasistatic, flat Rayleigh fading. The channel state information, i.e. the value of all the $\alpha_{i,j}$'s are assumed to be known by the receiver. This description of the channel will be valid throughout the chapter.

The decoding scheme is the same as described in Section 2.2.4. The decoder decides in favour of the codeword which minimizes the metric

$$\sum_{t=1}^{l} \sum_{j=1}^{n_R} \left| r_t^j - \sum_{i=1}^{n_T} \alpha_{i,j} c_t^i \right|^2 \tag{3.2}$$

over all codewords $\boldsymbol{c} = c_1^1 c_1^2 \dots c_1^{n_T} c_2^1 c_2^2 \dots c_2^{n_T} \dots c_l^1 c_l^2 \dots c_l^{n_T}$.

Our objective in space-time code design will again be to minimize the probability of erroneous decoding. This means that the analysis summarized in Section 2.3 and the diversity criterion are valid in constructing space-time block codes from orthogonal designs. Therefore, in order to achieve the maximum possible diversity advantage

$n_T n_R$, the matrix

$$\mathbf{B}(\boldsymbol{c},\boldsymbol{e}) = \begin{bmatrix} c_1^1 - e_1^1 & c_2^1 - e_2^1 & \dots & c_l^1 - e_l^1 \\ c_1^2 - e_1^2 & c_2^2 - e_2^2 & \dots & c_l^2 - e_l^2 \\ \vdots & \vdots & \ddots & \vdots \\ c_1^{n_T} - e_1^{n_T} & c_2^{n_T} - e_2^{n_T} & \dots & c_l^{n_T} - e_l^{n_T} \end{bmatrix} \tag{3.3}$$

has to be full rank for any pair of distinct codewords $\boldsymbol{c}$ and $\boldsymbol{e}$. Since no coding will be included this time, no criterion related to the coding advantage is taken into account.

3.1.2 Real Orthogonal Designs and the Coding Scheme

A *real orthogonal design of size* n is an $n \times n$ orthogonal matrix having as entries the indeterminates $\pm x_1, \pm x_2, \dots, \pm x_n$. Thus, a real orthogonal design $\mathcal{O}$ of size n satisfies $\mathcal{O}\mathcal{O}^T = \mathcal{O}^T\mathcal{O} = \left[\sum_{i=1}^{n}(x_i)^2\right] I_n$ where I_n is the identity matrix of size n. Given a real orthogonal design, it is clear that negating a number of columns results in another real orthogonal design. Similarly, permuting the columns of a real orthogonal design does not affect orthogonality. Therefore, from any real orthogonal design of size n, by permuting columns and changing signs of certain columns where necessary, we can obtain an orthogonal design with first row $x_1, x_2, \dots, x_n$, which we call a *normalized* orthogonal design. The following is an example of a normalized real orthogonal design of size 4:

$$\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ -x_2 & x_1 & -x_4 & x_3 \\ -x_3 & x_4 & x_1 & -x_2 \\ -x_4 & -x_3 & x_2 & x_1 \end{bmatrix}. \tag{3.4}$$

Now we explain how real orthogonal designs are used to construct space-time codes. Let us assume that a real constellation $\mathcal{A}$ of 2^b elements is used. We decide on a one-to-one mapping f from the set $(\mathbb{F}_2)^b$ of all binary vectors of length b to the real constellation alphabet $\mathcal{A}$. Then, we fix a normalized real orthogonal design $\mathcal{O}$ of size n_T and use it as a template for our coding scheme as follows: At time slot 1, bn_T bits arrive at the encoder. n_T groups $u_1, u_2, \dots, u_{n_T}$ are formed from these information bits, each consisting of b bits. Then, each u_i is mapped to the constellation symbol $s_i = f(u_i)$ for $i = 1, 2, \dots, n_T$. In $\mathcal{O}$, we replace the entries x_i by s_i for

$i = 1, 2, \ldots, n_T$ and arrive at an orthogonal matrix $\mathcal{S} = \mathcal{O}(s_1, s_2, \ldots, s_{n_T})$ having as entries $\pm s_1, \pm s_2, \ldots, \pm s_{n_T}$. At time slot t, the (t, i) entry $\mathcal{S}_{ti}$ is transmitted from transmit antenna i for $i = 1, 2, \ldots, n_T$ and $t = 1, 2, \ldots, n_T$. Thus the frame length l is equal to n_T, the number of transmit antennas. Since it takes n_T time slots to transmit bn_T information bits, the transmission rate of this coding scheme is b bits/s/Hz.

Theorem 3.1.1 ([19]). *The above coding scheme attains the maximum possible diversity order $n_T n_R$.*

Proof. Let us call the matrices formed by replacing x_i by the constellation symbols s_i the *coding matrices* of the space-time block code. It is clear from the above description that all coding matrices are completely determined by their first rows. Thus, all coding matrices of the code can be uniquely represented by $\mathcal{O}(s_1, s_2, \ldots, s_{n_T})$ where the s_i are elements of the signal constellation $\mathcal{A}$.

In order to prove the assertion, we have to show that the difference of any two distinct code matrices has full rank n_T. Let $\mathcal{O}(s_1, s_2, \ldots, s_{n_T})$ and $\mathcal{O}(\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_{n_T})$ be two coding matrices corresponding to distinct information sequences so that $s_i \neq \tilde{s}_i$ for at least one i. Then their difference is easily seen to be equal to $\mathcal{B} = \mathcal{O}(s_1 - \tilde{s}_1, s_2 - \tilde{s}_2, \ldots, s_{n_T} - \tilde{s}_{n_T})$, the matrix constructed from the original orthogonal design $\mathcal{O}(x_1, x_2, \ldots, x_{n_T})$ by replacing x_i by $s_i - \tilde{s}_i$. Therefore $\mathcal{B}$ is an orthogonal matrix and hence the equality

$$\mathcal{B}\mathcal{B}^T = \Big[\sum_{i=1}^{n_T} (s_i - \tilde{s}_i)^2\Big] I_n \tag{3.5}$$

holds. Taking the determinant we see that

$$\det(\mathcal{B}\mathcal{B}^T) = \det(\mathcal{B})\det(\mathcal{B}^T) = \Big[\sum_{i=1}^{n_T} (s_i - \tilde{s}_i)^2\Big]^{n_T}. \tag{3.6}$$

Since any matrix has the same determinant as its transpose, we have

$$\det(\mathcal{B}) = \Big[\sum_{i=1}^{n_T} (s_i - \tilde{s}_i)^2\Big]^{n_T/2}. \tag{3.7}$$

Since $s_i - \tilde{s}_i$ is nonzero for at least one i, this determinant is nonzero and hence $\mathcal{B}$ is of full rank n_T. It follows that orthogonal space-time block codes satisfy the diversity criterion. □

The above shows that orthogonal space-time block codes achieve the highest possible diversity advantage. Furthermore, in view of Corollary 2.4.2 from the previous chapter, their transmission rate is also maximal. However, they have a major drawback expressed by the following:

Theorem 3.1.2. *A real orthogonal design of size n exists if and only if $n = 2, 4$ or 8.*

For a proof see [19, page 5]. We continue with the decoding of orthogonal space-time block codes.

3.1.3 Decoding of Orthogonal STBC's

The decoding rule of orthogonal space-time block codes are the same as the one given for space time trellis codes in Section 2.2.4. Assuming perfect channel state information, the receiver decides in favour of the codeword which minimizes the metric

$$\sum_{t=1}^{l}\sum_{j=1}^{n_R}\left| r_t^j - \sum_{i=1}^{n_T}\alpha_{i,j}c_t^i \right|^2 \tag{3.8}$$

over all possible codewords $c_1^1c_1^2\ldots c_1^{n_T}c_2^1c_2^2\ldots c_2^{n_T}\ldots c_l^1c_l^2\ldots c_l^{n_T}$, where the frame length l is known to be equal to n_T from the previous section. We have seen that any codeword is completely determined by its first n_T symbols. Since each symbol is taken from a constellation of size 2^b, the number of total codewords is equal to 2^{bn_T}. Therefore, unless any shortcut method is used the above metric should be calculated for this many codewords. As b and n_T gets larger, this process may become computationally infeasible.

Fortunately, orthogonality of the coding matrices allows us to carry out the minimization process with relative ease. We observe that, the original orthogonal design $\mathcal{O}$ which we use as template for the code has the property that each x_i appears exactly once in each row. Therefore, if we ignore the minus signs, each row of $\mathcal{O}$ corresponds to a permutation of the set $\{1, 2, \ldots, n_T\}$. Let us denote the permutation associated with row i by ϵ_i for $i = 1, 2, \ldots, n_T$. Thus, $\epsilon_t(i) = j$ means that the constellation symbol corresponding to x_i is transmitted from transmit antenna j at time t. Let

$\delta_j(i)$ denote the sign of x_i in the j^{th} row of $\mathcal{O}$. Then (3.8) can be written as

$$\sum_{t=1}^{n_T}\sum_{j=1}^{n_R}\left| r_t^j - \sum_{i=1}^{n_T}\alpha_{\epsilon_t(i),j}\delta_t(i)s_i \right|^2. \tag{3.9}$$

According to [19], using the orthogonality of columns of $\mathcal{O}$, minimizing this amounts to minimizing

$$\sum_{i=1}^{n_T} S_i \tag{3.10}$$

where

$$S_i = \left| \left[\sum_{t=1}^{n_T}\sum_{j=1}^{n_R} r_t^j \alpha^*_{\epsilon_t(i),j}\delta_t(i) \right] - s_i \right|^2 + \left(-1 + \sum_{i=1}^{n_T}\sum_{j=1}^{n_R}|\alpha_{i,j}|^2 \right)|s_i|^2 \tag{3.11}$$

and where * denotes the complex conjugate. Now, since $\mathcal{O}$ has been fixed, the value of $\epsilon_t(i)$ and $\delta_t(i)$ is unique for each (t,i) pair. For that reason, the above expression for S_i makes it clear that each S_i can be minimized individually. Thus, the receiver computes the decision metrics

$$R_i = \sum_{t=1}^{n_T}\sum_{j=1}^{n_R} r_t^j \alpha^*_{\epsilon_t(i),j}\delta_t(i) \tag{3.12}$$

for $i = 1, 2, \ldots, n_T$ and s_i is decided to be the constellation symbol which minimizes (3.11). This is written more explicitly as

$$s_i = \arg\min_{s\in\mathcal{A}} \left[|R_i - s|^2 + \left(-1 + \sum_{i=1}^{n_T}\sum_{j=1}^{n_R}|\alpha_{i,j}|^2 \right)|s|^2 \right]. \tag{3.13}$$

By this way, all the transmitted symbols $s_1, s_2, \ldots, s_{n_T}$ can be found. Since the transmitted codeword is completely determined by these n_T code symbols, decoding is thus completed. This calculation is significantly easier than (3.9). In addition, since the symbols s_i are computed independently, at most $n_T 2^b$ calculations are required this time. Still, there have been several improvements over this decoding scheme in the literature. Interested reader may refer to [20, 21].

3.2 STBC's from Generalized Real Orthogonal Designs

Real orthogonal designs can be used to construct a coding scheme which provides full diversity gain, maximum possible transmission rate and has a simple decoding algorithm. However, they are only available for 2, 4 or 8 transmit antennas. In order

to design orthogonal space-time block codes for other number of transmit antennas, the notion of real orthogonal designs needs to be generalized. In the present section, we make this generalization and establish some properties of these generalized designs.

3.2.1 Generalized Real Orthogonal Designs

We define a *generalized real orthogonal design of size* n to be a $p \times n$ matrix $\mathcal{G}$ with entries $0, \pm x_1, \pm x_2, \ldots, \pm x_k$ such that $\mathcal{G}^T\mathcal{G} = ((x_1)^2 + (x_2)^2 + \ldots + (x_k)^2)I_n$. The *rate* of $\mathcal{G}$ is defined to be equal to k/p.

It would be useful to draw attention to the differences between real orthogonal designs and generalized real orthogonal designs. In real orthogonal designs, the number of indeterminates is equal to the number of columns(and rows) and hence each indeterminate appears exactly once in each row. In generalized real orthogonal designs, on the other hand, the number k of indeterminates is not equal to n, the number of columns. In addition, 0 is allowed as an entry. The definition makes it clear that columns of generalized real orthogonal designs are still orthogonal to each other. Therefore, the simple decoding scheme introduced in Section 3.1.3 will be valid for space-time block codes constructed from them.

Given a $p \times n_T$ generalized real orthogonal design $\mathcal{G}$, it can be used as a template for a space-time block code as follows: kb bits arrive at the encoder and these bits are divided into k blocks of the same size b. Then using a fixed mapping, these blocks are mapped to symbols $s_1, s_2, \ldots, s_k$ from the constellation alphabet $\mathcal{A}$. Then for all $i = 1, 2, \ldots, k$, the entries x_i of $\mathcal{G}$ are replaced by s_i to form a new $p \times n_T$ matrix $\mathcal{C}$. As before, the entry $\mathcal{C}_{ti}$ is transmitted from transmit antenna i at time t. If $\mathcal{C}_{ti} = 0$, nothing is transmitted from transmit antenna i at time t. The proof of Theorem 3.1.1 holds also for generalized real orthogonal designs. Thus, this coding scheme achieves the maximum possible diversity advantage $n_T n_R$.

To prevent confusion, we find it useful to motivate the definiton of rate R for generalized real orthogonal designs. Since it takes p time slots to transmit kb bits, transmission rate of the above scheme is equal to kb/p. On the other hand, we know that for a space-time code having full diversity order $n_T n_R$, the highest possible transmission

rate is b bits/s/Hz. This means that at most pb bits can be transmitted in p time slots. For that reason, we define R to be equal to $kb/pb = k/p$. Note that the maximum achievable rate is 1 under this definition.

We have seen that generalized real orthogonal designs can be used in a straightforward manner to construct space-time block codes. Therefore, the problem of constructing generalized real orthogonal designs having desired properties becomes crucial. One of these desired properties is the maximization of R since we would like to be able to transmit as many information bits as possible. In addition to this, memory requirements should also be taken into account. Thus, given R and the number n_T of transmit antennas, p should be minimized. In fact, the minimum number p such that there exists a $p \times n$ orthogonal design of rate R is denoted by $A(R, n)$ and it is known as the *fundamental question of generalized orthogonal design theory.*

We continue with a short subsection on what is known in the mathematics literature as the *Hurwitz-Radon Theory.*

3.2.2 Hurwitz-Radon Theory

Related to our discussion on orthogonal designs is the Hurwitz-Radon Theory and so we reserve this section for a short introduction to the results from this theory which will be of interest to us.

We begin by defining a set of matrices which will turn out to be relevant to orthogonal designs. A *Hurwitz-Radon family of matrices of size* k is a set $\{B_1, B_2, \ldots, B_k\}$ of $n \times n$ matrices satisfying the following properties:

$$B_i^T B_i = I_n \text{ for } i = 1, 2, \ldots, k$$

$$B_i^T = -B_i \text{ for } i = 1, 2, \ldots, k$$

$$B_i B_j = -B_j B_i \text{ for } 1 \leq i < j \leq k. \tag{3.14}$$

The above definition was first made in [22]. The following collects several results about Hurwitz-Radon families which were obtained in the same paper. We state them in the form given in [19].

Theorem 3.2.1. *Let $n = 2^a b$ be a positive integer, where b is odd and $a = 4c + d$ with $0 \leq d < 4$. Any Hurwitz-Radon family of $n \times n$ matrices contains less than $\rho(n) = 8c+2^d$ matrices. Furthermore $\rho(n) \leq n$. A Hurwitz-Radon family of size $n-1$ containing $n \times n$ matrices exists if and only if $n = 2, 4$ or 8.*

The following shows that $\rho(n)$ given above can be defined as the smallest number k for which a Hurwitz-Radon family of size k containing $n \times n$ matrices does not exist.

Theorem 3.2.2. *For any positive integer n, there exists a Hurwitz-Radon family of size $\rho(n) - 1$ containing $n \times n$ integer matrices; i.e. matrices having all their entries from the set $\{-1, 0, 1\}$.*

The proof is constructive and interested reader may see [19, page 5] for it.

3.2.3 Construction and Some Results

In this subsection, we first tackle the issue of constructing generalized real orthogonal designs having full rate. The following construction is taken from [19].

Let $\mathbf{X} = (x_1, x_2, \ldots, x_p)^T$ be the $p \times 1$ column vector consisting of p indeterminates. From Theorem 3.2.2 we know that there exists a Hurwitz-Radon family of size $\rho(p)-1$ whose members are $p \times p$ integer matrices. Let us assume that $\{A_1, A_2, \ldots, A_{\rho(p)-1}\}$ be such a family of matrices. Letting $A_0 = I_p$ and $n_T \leq \rho(p)$, we construct a $p \times n_T$ matrix $\mathcal{G}$ by setting its i^{th} column to be equal to $A_{i-1}\mathbf{X}$ for $i = 1, 2, \ldots, n_T$.

Lemma 3.2.3. *The $p \times n_T$ matrix $\mathcal{G}$ constructed above is a generalized real orthogonal design with rate 1.*

Proof. Let $\boldsymbol{g}_i$ denote the i^{th} column of $\mathcal{G}$. Then we can express the (i, j) element of $\mathcal{G}^T\mathcal{G}$ by

$$\begin{aligned}(\mathcal{G}^T\mathcal{G})_{ij} &= \boldsymbol{g}_i^T \boldsymbol{g}_j \\ &= (A_{i-1}\mathbf{X})^T A_{j-1}\mathbf{X} \\ &= (\mathbf{X}^T A_{i-1}^T) A_{j-1}\mathbf{X} \\ &= \mathbf{X}^T (A_{i-1}^T A_{j-1})\mathbf{X} \qquad (3.15)\end{aligned}$$

For $i = j$, by the first Hurwitz-Radon condition given in (3.14), it is seen that

$$(\mathcal{G}^T\mathcal{G})_{ij} = \mathbf{X}^T I_p \mathbf{X} = \mathbf{X}^T\mathbf{X} = (x_1)^2 + (x_2)^2 + \ldots + (x_p)^2. \tag{3.16}$$

As for $i \neq j$, since $(\mathcal{G}^T\mathcal{G})_{ij}$ is a number, we can view it as a 1×1 matrix. Then its transpose is equal to itself. This goes as:

$$\begin{aligned}(\mathcal{G}^T\mathcal{G})_{ij} &= [(\mathcal{G}^T\mathcal{G})_{ij}]^T = [(\mathbf{X}^T A_{i-1}^T)(A_{j-1}\mathbf{X})]^T \\ &= (A_{j-1}\mathbf{X})^T(\mathbf{X}^T A_{i-1}^T)^T = \mathbf{X}^T A_{j-1}^T A_{i-1}\mathbf{X}\end{aligned} \tag{3.17}$$

On the other hand, using the second and the third of the Hurwitz-Radon conditions given by (3.14), we obtain

$$A_{j-1}^T A_{i-1} = -A_{j-1}A_{i-1} = A_{i-1}A_{j-1} = -A_{i-1}^T A_{j-1} \tag{3.18}$$

Substituting this in (3.17) and using (3.15) yields

$$(\mathcal{G}^T\mathcal{G})_{ij} = \mathbf{X}^T(-A_{i-1}^T A_{j-1})\mathbf{X} = -(\mathbf{X}^T A_{i-1}^T A_{j-1}\mathbf{X}) = -(\mathcal{G}^T\mathcal{G})_{ij}. \tag{3.19}$$

Therefore we have $(\mathcal{G}^T\mathcal{G})_{ij} = 0$ for $i \neq j$. Together with (3.16), this shows that $\mathcal{G}$ is a generalized real orthogonal design. Furthermore, the number of indeterminates and the number of rows of $\mathcal{G}$ are both equal to p by construction. Thus, $\mathcal{G}$ is a $p \times n_T$ generalized real orthogonal design having full rate. □

In Section 3.2.1 we posed the fundamental question of generalized orthogonal design theory. Since our objective is constructing codes having full rate, we are especially interested in the answer to this question for the case $R = 1$. In the above construction, the only relation between the number n_T of transmit antennas and the number p of time slots was $n_T \leq \rho(p)$. Thus, the lemma makes it clear that $A(1, n) \leq p$ for any p with $\rho(p) \geq n$. In other words, $A(1, n) \leq \min_{\rho(p)\geq n}(p)$. The following shows that this is indeed an equality.

Theorem 3.2.4 ([19])**.** *The value of $A(1, n)$ is the smallest number p such that $\rho(p) \geq n$. In other words, $A(1, n) = \min_{\rho(p)\geq n}(p)$.*

Proof. We have already seen that $A(1, n) \leq \min_{\rho(p)\geq n}(p)$. All we now need to show is $A(1, n) \geq \min_{\rho(p)\geq n}(p)$. Let $\mathbf{X} = (x_1, x_2, \ldots, x_p)^T$ as before and let $\mathcal{G}$ be a full rate $p \times n$ generalized real orthogonal design with minimal size, i.e. $p = A(1, n)$. The i^{th}

column of $\mathcal{G}$ can be expressed as $B_i\mathbf{X}$ for $i = 1, 2, \ldots, n$ where B_i is a $p \times p$ matrix. Then the orthogonality of $\mathcal{G}$ implies

$$\boldsymbol{g}_i^T \boldsymbol{g}_i = (B_i\mathbf{X})^T B_i\mathbf{X} = \mathbf{X}^T B_i^T B_i\mathbf{X} = \sum_{k=1}^{p}(x_k)^2 = \mathbf{X}^T\mathbf{X} \tag{3.20}$$

for $i = 1, 2, \ldots, n$ and

$$\begin{aligned} \boldsymbol{g}_i^T \boldsymbol{g}_j &= (B_i\mathbf{X})^T B_j\mathbf{X} = \mathbf{X}^T B_i^T B_j\mathbf{X} = 0 = -(\mathbf{X}^T B_i^T B_j\mathbf{X})^T \\ &= -\mathbf{X}^T(\mathbf{X}^T B_i^T B_j)^T = -\mathbf{X}^T[(B_i\mathbf{X})^T B_j]^T = -\mathbf{X}^T B_j^T B_i\mathbf{X} \end{aligned} \tag{3.21}$$

for $i \neq j$, where $\boldsymbol{g}_i$ denotes the i^{th} column of $\mathcal{G}$ as before. These identities imply that $B_i^T B_i = I_p$ for $i = 1, 2, \ldots, n$ and $B_i^T B_j = -B_j^T B_i$ for $1 \leq i < j \leq n$. Let us now consider the set $\{A_2, A_3, \ldots, A_n\}$ of $p \times p$ matrices where $A_i = B_1^T B_i$ for $i = 2, 3, \ldots, n$. Then these matrices satisfy

(i) for $i = 2, 3, \ldots, n$

$$\begin{aligned} A_i^T A_i &= (B_1^T B_i)^T B_1^T B_i = (B_i^T B_1)B_1^T B_i = (-B_1^T B_i)B_1^T B_i \\ &= -B_1^T(B_i B_1^T)B_i = -B_1^T(-B_1 B_i^T)B_i \\ &= (B_1^T B_1)(B_i^T B_i) = I_p I_p = I_p, \end{aligned} \tag{3.22}$$

(ii) for $i = 2, 3, \ldots, n$

$$A_i^T = (B_1^T B_i)^T = B_i^T B_1 = -B_1^T B_i = -A_i, \tag{3.23}$$

(iii) for $2 \leq i < j \leq n$

$$\begin{aligned} A_i A_j &= B_1^T B_i B_1^T B_j = B_1^T(B_i B_1^T)B_j = B_1^T(-B_1 B_i^T)B_j \\ &= -(B_1^T B_1)B_i^T B_j = -I_p B_i^T B_j = -B_i^T B_j \\ &= B_j^T B_i = -A_j A_i. \end{aligned} \tag{3.24}$$

Therefore, the set $\{A_2, A_3, \ldots, A_n\}$ is a Hurwitz-Radon family of size $n-1$. Since any Hurwitz-Radon family of $p \times p$ matrices cannot have more than $\rho(p) - 1$ members, it is true that $n - 1 \leq \rho(p) - 1$ and consequently $n \leq \rho(p)$. Thus, $p = A(1, n)$ cannot be less than the smallest integer p' having the property $\rho(p') \geq n$. In other words, $A(1, n) \geq \min_{\rho(p)\geq n}(p)$. Since we had previously seen that $A(1, n) \leq \min_{\rho(p)\geq n}(p)$, it follows that $A(1, n) = \min_{\rho(p)\geq n}(p)$. □

Corollary 3.2.5 ([19]). *If we define* $M_n = \{(c,d) : 0 \leq c, 0 \leq d < 4, 8c + 2^d \geq n\}$, *then* $A(1,n) = \min_{(c,d)\in M_n}(2^{4c+d})$.

Proof. First we show that $A(1,n) = p$ is a power of 2. Let $p = 2^a b$ where b is an odd integer. From the definition of ρ given in Theorem 3.2.1 it is clear that $\rho(2^a) = \rho(p)$. Since $\rho(p) \geq n$ by Lemma 3.2.3, we have $\rho(2^a) \geq n$. Since Theorem 3.2.4 says that p is the smallest number having this property, we must have $p \leq 2^a = p/b$. It follows that $b = 1$ and hence $A(1,n)$ is a power of 2. The result follows from the explicit formula given for ρ in Theorem 3.2.1. □

The above results are about how many time slots are required in a full rate orthogonal space-time block code using a fixed number of transmit antennas. In general, $p \times n$ generalized real orthogonal designs with rate R satisfying $p = A(R,n)$ are called *delay-optimal*. Since we are interested in designing full rate space-time block codes, of special interest for us is the designs attaining the value $A(1,n)$ for their number of rows. If one desires to use six transmit antennas, for instance, the number of time slots required is $A(1,6) = 8$. One can apply the construction described in the paragraph preceding Lemma 3.2.3 to realize such a design. The design taken from [19]

$$\mathcal{G}_6 = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ -x_2 & x_1 & x_4 & -x_3 & x_6 & -x_5 \\ -x_3 & -x_4 & x_1 & x_2 & x_7 & x_8 \\ -x_4 & x_3 & -x_2 & x_1 & x_8 & -x_7 \\ -x_5 & -x_6 & -x_7 & -x_8 & x_1 & x_2 \\ -x_6 & x_5 & -x_8 & x_7 & -x_2 & x_1 \\ -x_7 & x_8 & x_5 & -x_6 & -x_3 & x_4 \\ -x_8 & -x_7 & x_6 & x_5 & -x_4 & -x_3 \end{bmatrix} \tag{3.25}$$

is a delay-optimal 8×6 design with rate 1. For other examples see [19, 23].

3.3 Orthogonal STBC's for Complex Constellations

So far in this chapter we have investigated orthogonal space-time coding schemes to be used over real signal constellations. In reality, however, most applications assume

complex signal constellations. Therefore, it is natural to generalize real orthogonal designs to complex orthogonal designs. As will be seen, this generalization can be made in a rather straightforward manner. First we define complex orthogonal designs, which are analogous to real orthogonal designs defined in Section 3.1.2. Then, we generalize complex orthogonal designs to nonsquare matrices and establish some results related to them. We end the section with a short description of Alamouti's code.

3.3.1 Complex Orthogonal Designs

A *complex orthogonal design of size* n where $n > 1$ is an $n \times n$ orthogonal matrix $\mathcal{O}_c$ with entries the indeterminates $\pm x_1, \pm x_2, \ldots, \pm x_n$, complex conjugates of these indeterminates $\pm x_1^*, \pm x_2^*, \ldots, \pm x_n^*$, and multiples of these indeterminates by $\boldsymbol{i} = \sqrt{-1}$. As before, without loss of generality we may assume that $\mathcal{O}_c$ is normalized, i.e. the i^{th} entry $(\mathcal{O}_c)_{1i}$ in the first row of $\mathcal{O}_c$ is equal to x_i for all $i = 1, 2, \ldots, n$.

The encoding method introduced in Section 3.1.2 can be applied to complex orthogonal designs to obtain a space-time coding scheme over any complex signal constellation. As before, the resulting transmit diversity scheme achieves the maximum possible rate and the maximum diversity advantage $n_T n_R$. The decoding method presented in Section 3.1.3 can be applied to decode complex orthogonal space-time block codes.

We have seen that real orthogonal designs exist only for $2, 4$ or 8 dimensions. Therefore, it is reasonable to suspect that complex orthogonal designs also exist for a limited set of dimensions. A construction described in [19] proves useful for the examination of this existence problem. This construction goes as follows: Suppose that we have a complex orthogonal design $\mathcal{O}_c$ of size n. In view of the encoding scheme we have described, each indeterminate in $\mathcal{O}_c$ is replaced by a complex constellation symbol before data transmission. Thus, we can view the indeterminate x_i as a complex number of the form $x_i^1 + \boldsymbol{i} x_i^2$. In other words, to each complex indeterminate x_i we associate two real indeterminates x_i^1 and x_i^2. Based on this explanation, if the pair of real indeterminates associated with any entry of $\mathcal{O}_c$ is (x, y), we replace that entry of $\mathcal{O}_c$ with the 2×2 matrix

$$\begin{bmatrix} x & y \\ -y & x \end{bmatrix}. \tag{3.26}$$

More explicitly, the entries $\pm x_i, \pm x_i^*, \pm \boldsymbol{i} x_i$ are replaced by

$$\pm \begin{bmatrix} x_i^1 & x_i^2 \\ -x_i^2 & x_i^1 \end{bmatrix}, \pm \begin{bmatrix} x_i^1 & -x_i^2 \\ x_i^2 & x_i^1 \end{bmatrix}, \pm \begin{bmatrix} -x_i^2 & x_i^1 \\ -x_i^1 & -x_i^2 \end{bmatrix} \tag{3.27}$$

respectively. Let us denote by $\mathcal{O}$ the resulting $2n \times 2n$ matrix.

Lemma 3.3.1. *The matrix $\mathcal{O}$ constructed above is a real orthogonal design.*

Proof. Let us first introduce the notation that will be used throughout the proof. We can view $\mathcal{O}$ as an $n \times n$ matrix whose entries are 2×2 matrices. Let $\mathcal{R}^{ij}$ denote the matrix in the (i, j) position of $\mathcal{O}$. Thus, if $\mathcal{O}_{ij}$ denotes the (i, j) entry of $\mathcal{O}$ as usual, we have

$$\mathcal{R}^{ij} = \begin{bmatrix} \mathcal{O}_{2i-1,2j-1} & \mathcal{O}_{2i-1,2j} \\ \mathcal{O}_{2i,2j-1} & \mathcal{O}_{2i,2j} \end{bmatrix}. \tag{3.28}$$

In addition, let $\mathcal{O}_i$ denote the i^{th} column of $\mathcal{O}$ for $i = 1, 2, \ldots, 2n$ and let $\mathcal{R}_1^{ij}$ and $\mathcal{R}_2^{ij}$ denote the first and the second column of $\mathcal{R}^{ij}$, respectively.

We now examine the conditions imposed by the orthogonality of columns of $\mathcal{O}_c$. Since $\mathcal{O}_c$ is a complex orthogonal design, for $1 \leq i < j \leq n$ we have

$$(\mathcal{O}_c)_i^* (\mathcal{O}_c)_j = \sum_{k=1}^{n} (\mathcal{O}_c)_{ki}^* (\mathcal{O}_c)_{kj} = 0, \tag{3.29}$$

where $(\mathcal{O}_c)_i^*$ is the $1 \times n$ matrix obtained by taking the transpose conjugate of the i^{th} column of $\mathcal{O}_c$ and $(\mathcal{O}_c)_{ki}^*$ is the complex conjugate of the (k, i) entry of $\mathcal{O}_c$. The definition of a complex orthogonal design implies that each indeterminate appears exactly once in each column and each row of $\mathcal{O}_c$. Thus, the above summation is a sum of product of terms corresponding to different indeterminates. Furthermore, the product $(\mathcal{O}_c)_{ki}^* (\mathcal{O}_c)_{kj}$ corresponding to the entries in the k^{th} row has to be eliminated by the product $(\mathcal{O}_c)_{li}^* (\mathcal{O}_c)_{lj}$ corresponding to the entries in the l^{th} row for exactly one $l \neq k$. Let us denote this relation by $\epsilon_{ij}(k) = l$. Thus, for all distinct pairs i and j, ϵ_{ij} is a permutation of the set $N = \{1, 2, \ldots, n\}$ such that $\epsilon_{ij}(k) \neq k$ and $\epsilon_{ij}^2(k) = k$ for all $k \in N$.

We can now check if any two distinct columns of $\mathcal{O}$ are orthogonal. Let $\mathcal{O}_i$ and $\mathcal{O}_j$ be two such columns with $i < j$. The inner product of $\mathcal{O}_i$ and $\mathcal{O}_j$ can be expressed as

$$(\mathcal{O}_i)^T \mathcal{O}_j = \sum_{k=1}^{2n} \mathcal{O}_{ki} \mathcal{O}_{kj}. \tag{3.30}$$

If i is odd and $j = i+1$, then this sum can be written as

$$(\mathcal{O}_i)^T\mathcal{O}_j = \sum_{k=1}^{n}\left(\mathcal{O}_{2k-1,i}\mathcal{O}_{2k-1,j} + \mathcal{O}_{2k,i}\mathcal{O}_{2k,j}\right) = \sum_{k=1}^{n}(\mathcal{R}_1^{k,j/2})^T\mathcal{R}_2^{k,j/2}. \tag{3.31}$$

In this case, the description of the $\mathcal{R}^{ij}$'s given by (3.26) shows that $(\mathcal{R}_1^{ij})^T\mathcal{R}_2^{ij} = 0$ for all i and j. So $(\mathcal{O}_i)^T\mathcal{O}_j = 0$ in this case.

Suppose that either i is even or $j > i+1$ so that the elements of $\mathcal{O}_i$ and $\mathcal{O}_j$ lie in a disjoint set of 2×2 submatrices. More explicitly, the entries in $\mathcal{O}_i$ lie in the submatrices $\mathcal{R}^{k,\lfloor\frac{i+1}{2}\rfloor}$ and the entries in $\mathcal{O}_j$ lie in the submatrices $\mathcal{R}^{k,\lfloor\frac{j+1}{2}\rfloor}$ for $k = 1, 2, \ldots, n$. The condition we imposed on i and j guarantees that $p = \lfloor\frac{i+1}{2}\rfloor$ and $q = \lfloor\frac{j+1}{2}\rfloor$ are distinct integers. The orthogonality of $\mathcal{O}_c$ means that

$$\begin{aligned}(\mathcal{O}_c)_p^*(\mathcal{O}_c)_q &= \sum_{k=1}^{n}(\mathcal{O}_c)_{kp}^*(\mathcal{O}_c)_{kq} = 2\sum_{k=1}^{n}(\mathcal{O}_c)_{kp}^*(\mathcal{O}_c)_{kq} \\ &= \sum_{k=1}^{n}\left[(\mathcal{O}_c)_{kp}^*(\mathcal{O}_c)_{kq} + (\mathcal{O}_c)_{\epsilon_{pq}(k),p}^*(\mathcal{O}_c)_{\epsilon_{pq}(k),q}\right] = 0.\end{aligned} \tag{3.32}$$

Now, for an arbitrary $l \in \{1, 2, \ldots, n\}$ there are four possibilities which makes the equality

$$(\mathcal{O}_c)_{lp}^*(\mathcal{O}_c)_{lq} + (\mathcal{O}_c)_{\epsilon_{pq}(l),p}^*(\mathcal{O}_c)_{\epsilon_{pq}(l),q} = 0 \tag{3.33}$$

true. These are

(i) $(\mathcal{O}_c)_{\epsilon_{pq}(l),p} = (\mathcal{O}_c)_{lq}^*$ and $(\mathcal{O}_c)_{\epsilon_{pq}(l),q} = -(\mathcal{O}_c)_{lp}^*$.

(ii) $(\mathcal{O}_c)_{\epsilon_{pq}(l),p} = -(\mathcal{O}_c)_{lq}^*$ and $(\mathcal{O}_c)_{\epsilon_{pq}(l),q} = (\mathcal{O}_c)_{lp}^*$.

(iii) $(\mathcal{O}_c)_{\epsilon_{pq}(l),p} = \boldsymbol{i}(\mathcal{O}_c)_{lq}^*$ and $(\mathcal{O}_c)_{\epsilon_{pq}(l),q} = \boldsymbol{i}(\mathcal{O}_c)_{lp}^*$.

(iv) $(\mathcal{O}_c)_{\epsilon_{pq}(l),p} = -\boldsymbol{i}(\mathcal{O}_c)_{lq}^*$ and $(\mathcal{O}_c)_{\epsilon_{pq}(l),q} = -\boldsymbol{i}(\mathcal{O}_c)_{lp}^*$.

In every case, from the definition of the submatrices made in (3.26) and (3.27) we see that

$$(\mathcal{R}_1^{kp})^T\mathcal{R}_1^{kq} + (\mathcal{R}_1^{\epsilon_{pq}(k),p})^T\mathcal{R}_1^{\epsilon_{pq}(k),q} = 0, (\mathcal{R}_1^{kp})^T\mathcal{R}_2^{kq} + (\mathcal{R}_1^{\epsilon_{pq}(k),p})^T\mathcal{R}_2^{\epsilon_{pq}(k),q} = 0,$$

$$(\mathcal{R}_2^{kp})^T\mathcal{R}_1^{kq} + (\mathcal{R}_2^{\epsilon_{pq}(k),p})^T\mathcal{R}_1^{\epsilon_{pq}(k),q} = 0, (\mathcal{R}_2^{kp})^T\mathcal{R}_2^{kq} + (\mathcal{R}_2^{\epsilon_{pq}(k),p})^T\mathcal{R}_2^{\epsilon_{pq}(k),q} = 0.$$

for all $k = 1, 2, \ldots, n$. One of these four corresponds to the sum

$$\mathcal{O}_{2k-1,i}\mathcal{O}_{2k-1,j} + \mathcal{O}_{2k,i}\mathcal{O}_{2k,j} + \mathcal{O}_{2\epsilon_{pq}(k)-1,i}\mathcal{O}_{2\epsilon_{pq}(k)-1,j} + \mathcal{O}_{2\epsilon_{pq}(k),i}\mathcal{O}_{2\epsilon_{pq}(k),j}. \tag{3.34}$$

Therefore, if we denote this sum by $\mathcal{S}_k^{ij}$, we have

$$2\mathcal{O}_i^T\mathcal{O}_j = \sum_{k=1}^{n} \mathcal{S}_k^{ij} = 0 = \mathcal{O}_i^T\mathcal{O}_j. \tag{3.35}$$

Thus, we have shown that any two distinct columns of $\mathcal{O}$ are orthogonal. This completes the proof that $\mathcal{O}$ is a real orthogonal design of size $2n$. □

Since constructing complex orthogonal designs is not easier than constructing real ones, it is not profitable to use the above for construction purposes. Instead, it is useful in limiting the set of values for which a complex orthogonal design can exist. Since the above construction guarantees the existence of a real orthogonal design of size $2n$ corresponding to every complex orthogonal design of size n, from Theorem 3.1.2 we see that a complex orthogonal design of size n cannot exist unless $n = 2$ or 4. But, it is proved in [19] that a complex orthogonal design of size 4 does not exist. As for $n = 2$, Alamouti's scheme, which we will later discuss shortly, gives a complex orthogonal design of size 2. Thus, we have the following:

Theorem 3.3.2 ([19])**.** *A complex orthogonal design of size n exists if and only if $n = 2$.*

3.3.2 Generalized Complex Orthogonal Designs

We have seen that complex orthogonal designs exist only for two dimensions. Therefore, they can only be used to construct space-time block codes for two transmit antennas. For other number of antennas, as before, one needs to make generalization to nonsquare matrices. We define a *generalized complex orthogonal design of size n* to be a $p \times n$ matrix $\mathcal{G}_c$ with entries $\pm x_1, \pm x_1^*, \ldots, \pm x_k, \pm x_k^*$ and their product with $\boldsymbol{i}$ such that $\mathcal{G}_c^*\mathcal{G}_c = \mathcal{D}_c$, where $\mathcal{G}_c^*$ denotes the transpose conjugate of $\mathcal{G}_c$ and $\mathcal{D}_c$ is a diagonal matrix with entries

$$\mathcal{D}_{ii} = l_1^i|x_1|^2 + l_2^i|x_2|^2 + \ldots + l_k^i|x_k|^2 \tag{3.36}$$

for all $i = 1, 2, \ldots, n$. It can be shown that(see [19]) there exists a generalized complex orthogonal design of the same size as $\mathcal{G}_c$ whose diagonal entries are equal to $|x_1|^2 + |x_2|^2 + \ldots + |x_k|^2$. Thus, without loss of generality we can assume $\mathcal{G}_c$ has this property. As before, k/p is the rate of $\mathcal{G}_c$.

Given a generalized complex orthogonal design, the encoding method described before can be used to obtain a transmit diversity scheme achieving the maximum possible diversity advantage. Decoding is made by the method explained in Section 3.1.3. As for the rate, analogous to $A(R,n)$ defined in Section 3.2.1, we define $A_c(R,n)$ to be the minimum number p such that there exists a $p \times n$ generalized complex orthogonal design with rate R. Our main focus in this subsection will be to establish the limitations on $A_c(R,n)$.

Theorem 3.3.3 ([19]). *The following inequalities hold:*

(i) For any R, $A_c(R,n) \geq \dfrac{A(R,2n)}{2}$.

(ii) For $R \leq \dfrac{1}{2}$, $A_c(R,n) \leq 2A(2R,n)$.

Proof. For part i), note first that there is nothing to prove if $A_c(R,n) = \infty$, i.e. if there does not exist a $p \times n$ orthogonal design for some given R and n. If $A_c(R,n) < \infty$, we consider an $A_c(R,n) \times n$ generalized complex orthogonal design $\mathcal{G}_c$ with rate at least R. From $\mathcal{G}_c$ we can obtain a $2A_c(R,n) \times 2n$ generalized real orthogonal design $\mathcal{G}$ by applying the construction proved in Lemma 3.3.1. Furthermore $\mathcal{G}$ has the same rate as $\mathcal{G}_c$. This shows that $A(R,2n) \leq 2A_c(R,n)$.

For the second part, assume that we have a $p \times n$ generalized real orthogonal design $\mathcal{G}$ of rate at least $2R$ where $p = A(2R,n)$. Since $2R \leq 1$, such a design always exists. From $\mathcal{G}$ we form another matrix $\tilde{\mathcal{G}}$ by replacing each entry x_i by the symbolic conjugate x_i^*. Then we define a $2p \times n$ matrix $\mathcal{G}_c$ by appending $\tilde{\mathcal{G}}$ below $\mathcal{G}$, i.e. the i^{th} row of $\mathcal{G}_c$ is the i^{th} row of $\mathcal{G}$ and the $(p+i)^{\text{th}}$ row of $\mathcal{G}_c$ is the i^{th} row of $\tilde{\mathcal{G}}$ for $1 \leq i \leq p$. Our claim is that $\mathcal{G}_c$ is a generalized complex orthogonal design. To show this first we note that for all $i = 1, 2, \ldots, n$

$$\begin{aligned}(\mathcal{G}_c^*\mathcal{G}_c)_{ii} &= (\mathcal{G}_c)_i^*(\mathcal{G}_c)_i = \sum_{j=1}^{2p}(\mathcal{G}_c)_{ji}^*(\mathcal{G}_c)_{ji} = \sum_{j=1}^{p}|\mathcal{G}_{ji}|^2 + \sum_{j=1}^{p}|(\mathcal{G}_{ji})^*|^2 \\ &= (|x_1|^2 + |x_2|^2 + \ldots + |x_k|^2) + (|x_1^*|^2 + |x_2^*|^2 + \ldots + |x_k^*|^2) \\ &= 2(|x_1|^2 + |x_2|^2 + \ldots + |x_k|^2). \qquad (3.37)\end{aligned}$$

where k is the number of indeterminates in $\mathcal{G}$ and single subscripts denote the columns of matrices as usual. Now let us assume $i \neq j$. We will use the notation introduced

in the proof of Lemma 3.3.1. Again ϵ^2 is the identity permutation and $\epsilon_{ij}(r) = s$ will mean

$$\mathcal{G}_{ri}\mathcal{G}_{rj} + \mathcal{G}_{si}\mathcal{G}_{sj} = 0. \tag{3.38}$$

Since the $(p+l)^{\text{th}}$ row of $\mathcal{G}_c$ is equal to the matrix conjugate of the l^{th} row of $\mathcal{G}$, we have $(\mathcal{G}_c)_{l+p,q} = (\mathcal{G}_{lq})^*$ for all $q = 1, 2, \ldots, n$. It follows that

$$\begin{aligned}
(\mathcal{S}_c)_r^{ij} &= (\mathcal{G}_c^*)_{ir}(\mathcal{G}_c)_{rj} + (\mathcal{G}_c^*)_{is}(\mathcal{G}_c)_{sj} + (\mathcal{G}_c^*)_{i,r+p}(\mathcal{G}_c)_{r+p,j} + (\mathcal{G}_c^*)_{j,s+p}(\mathcal{G}_c)_{s+p,j} \\
&= (\mathcal{G}_{ri})^*\mathcal{G}_{rj} + (\mathcal{G}_{si})^*\mathcal{G}_{sj} + \mathcal{G}_{ri}(\mathcal{G}_{rj})^* + \mathcal{G}_{si}(\mathcal{G}_{sj})^* \\
&= \left[(\mathcal{G}_{ri})^*\mathcal{G}_{rj} + \mathcal{G}_{si}(\mathcal{G}_{sj})^*\right] + \left[\mathcal{G}_{ri}(\mathcal{G}_{rj})^* + (\mathcal{G}_{si})^*\mathcal{G}_{sj}\right] \\
&= 0 + 0 = 0
\end{aligned} \tag{3.39}$$

in view of (3.38), taking into account that all the involved terms are indeterminates. We can now compute the inner product of i^{th} and j^{th} columns of $\mathcal{G}_c$. This is just equal to

$$\begin{aligned}
(\mathcal{G}_c^*\mathcal{G}_c)_{ij} &= \sum_{l=1}^{2p} (\mathcal{G}_c^*)_{il}(\mathcal{G}_c)_{lj} \\
&= \frac{1}{2}\sum_{r=1}^{p} (\mathcal{S}_c)_r^{ij} = 0.
\end{aligned} \tag{3.40}$$

From (3.37) and (3.40), it follows that $\mathcal{G}_c$ is a $2p \times n$ generalized complex orthogonal design with rate at least R. Therefore, we have $A_c(R, n) \leq 2p = 2A(2R, n)$. □

Part ii) of the above theorem yields an explicit construction of generalized complex orthogonal design up to rates $\frac{1}{2}$. One example for three transmit antennas, which was constructed in [19], is the following:

$$\mathcal{G}_c^3 = \begin{bmatrix}
x_1 & x_2 & x_3 \\
-x_2 & x_1 & -x_4 \\
-x_3 & x_4 & x_1 \\
-x_4 & -x_3 & x_2 \\
x_1^* & x_2^* & x_3^* \\
-x_2^* & x_1^* & -x_4^* \\
-x_3^* & x_4^* & x_1^* \\
-x_4^* & -x_3^* & x_2^*
\end{bmatrix}. \tag{3.41}$$

The space-time block code constructed from this design takes 8 time slots to transmit 4 complex symbols. Therefore, its rate is equal to $R = \frac{4}{8} = \frac{1}{2}$. In fact, this is not the best that can be achieved for three transmit antennas. The following design

$$\mathcal{G}_c^{3,3/4} = \begin{bmatrix} x_1 & x_2 & x_3 \\ -x_2^* & x_1^* & 0 \\ -x_3^* & 0 & x_1 \\ 0 & -x_3^* & x_2^* \end{bmatrix}, \tag{3.42}$$

which was derived in [24], achieves rate $\frac{3}{4}$ for three transmit antennas. For two designs of the same rate which have the additional property that linear combination of indeterminates are allowed as the entries, one can see [19]. These two designs are for three and four transmit antennas. The following shows in particular that this is the best possible for these many antennas.

Theorem 3.3.4 ([25])**.** *Let n_T be an integer greater than* 1*. Then for R greater than $\frac{1}{2} + \frac{1}{2\left\lfloor \frac{n_T+1}{2} \right\rfloor}$, we have $A_c(R, n_T) = \infty$.*

The nonexistence of a full rate complex orthogonal space-time block code for more than two transmit antennas can be inferred from this theorem. As the expression suggests, the difference of the maximum possible rate from 1/2 loses its significance as the number of transmit antennas becomes large. Furthermore, codes achieving the maximum possible rate for large number of transmit antennas have very long decoding delays when compared to codes having their rate equal to 1/2. For example, for $n_T = 16$, the maximum possible rate is 9/16 and such a design constructed in [26] has $p = 22880$, i.e. takes 22880 time slots to transmit the information symbols. On the other hand, since $A(1, 16) = 128$ by Corollary 3.2.5, the construction described in the proof of the second part of Theorem 3.3.3 yields a 256×16 design with rate 1/2. This suggests that, in terms of decoding delay, it is not advisable to employ codes with the highest possible rate for large number of transmit antennas.

3.3.3 An Example: Alamouti's Code

Alamouti's code, which was first proposed in [18], is of special importance since it is the only complex orthogonal space-time block code which achieves full rate. Despite

its simplicity, its uniqueness has earned it the distinction of being the code which has been proposed in several third-generation cellular standards.

Alamouti's scheme is based on the 2×2 complex design

$$\mathcal{O}_c^2 = \begin{bmatrix} x_1 & x_2 \\ -x_2^* & x_1^* \end{bmatrix}, \tag{3.43}$$

which is orthogonal since $(\mathcal{O}_c^2)^* \mathcal{O}_c^2 = (|x_1|^2 + |x_2|^2) I_2$. The encoding scheme is obvious. Suppose that the complex signal constellation $\mathcal{S}$ to be used has 2^b elements. Then, $2b$ bits are transmitted in each data frame. Namely, if a binary data sequence of $2b$ bits arrives at the encoder, this block is divided into two parts of the same length and these parts select complex symbols s_1 and s_2 from $\mathcal{S}$. At the first time slot, s_1 is transmitted from the first transmit antenna and s_2 is transmitted from the second transmit antenna. At the second time slot, $-s_2^*$ is transmitted from the first transmit antenna and s_1^* is transmitted from the second transmit antenna. By this way, the encoder takes two time slots to transmit $2b$ bits. Thus, the transmission rate is b bits/s/Hz, which shows that the code is optimal in terms of rate.

Although the above coding scheme can be used with any number of receive antennas, for simplicity we will take $n_R = 1$. The path gains from transmit antennas 1 and 2 to the receive antenna are α_1 and α_2 respectively. The received signals at time slot 1 and 2 are given by

$$r_1 = \alpha_1 s_1 + \alpha_2 s_2 + \eta_1, r_2 = -\alpha_1 s_2^* + \alpha_2 s_1^* + \eta_2 \tag{3.44}$$

respectively, where η_1 and η_2 are the noise values associated with the receive antenna at time slots 1 and 2 respectively. Decoding is carried out by the method described in Section 3.1.3. Applying (3.9) to the present scheme, we see that the metric

$$|r_1 - \alpha_1 s_1 - \alpha_2 s_2|^2 + |r_2 + \alpha_1 s_2^* - \alpha_2 s_1^*|^2 \tag{3.45}$$

should be minimized over all possible values of s_1 and s_2. As before, some simplification can be made over this and it turns out that we should minimize

$$S_1 = |\alpha_1^* r_1 + \alpha_2 r_2^* - s|^2 + (|\alpha_1|^2 + |\alpha_2|^2 - 1)|s|^2 \tag{3.46}$$

over all $s \in \mathcal{S}$ to find s_1 and

$$S_2 = |\alpha_2^* r_1 - \alpha_1 r_2^* - s|^2 + (|\alpha_1|^2 + |\alpha_2|^2 - 1)|s|^2 \tag{3.47}$$

over all $s \in \mathcal{S}$ to find s_2. This process can be further simplified if all the constellation symbols have equal energies, i.e. if $|s|$ is constant for all $s \in \mathcal{S}$. If this is the case, the right-hand sides of (3.46) and (3.47) are constant. Then we form two decision variables, which are

$$\tilde{s}_1 = \alpha_1^* r_1 + \alpha_2 r_2^*, \tilde{s}_2 = \alpha_2^* r_1 - \alpha_1 r_2^*. \tag{3.48}$$

In this case, s_1 is equal to $\arg\min_{s \in \mathcal{S}} |s - \tilde{s}_1|$, the constellation symbol which is closest to $\tilde{s}_1$. Likewise, s_2 is the constellation symbol closest to $\tilde{s}_2$. This is a simple decoding scheme which requires only linear processing at the receiver.

Alamouti's scheme can be analyzed in terms of various aspects some of which we have not touched so far. For instance, the change in the error rate performance of the code with the energy of the constellation symbols can be considered. For this and other details, the reader is referred to [18] and [23, Section 4.2].

CHAPTER 4

SPACE-TIME BLOCK CODES FROM RANK DISTANCE CODES

We have seen that the performance of a space-time code in terms of pairwise error probability is quantified by the diversity criterion. Namely, in order to achieve the maximum possible diversity advantage, all of the difference matrices constructed from distinct pairs of codewords should have full rank. In view of this criterion, it is very natural to treat the codewords of a space-time code as matrices whose distance is equal to the rank of their difference. Such a view connotes *rank distance codes*, a class of matricial codes from coding theory with the rank metric as their associated distance metric.

Rank distance codes were introduced by E. M. Gabidulin in [27]. Their original purpose was to correct rank errors occurring in information transmission. Apart from being used as error-correcting codes, they serve as building blocks of several cryptosystems(see eg. [28]). More recently, several authors have focused on the possibility of utilizing rank distance codes to construct space-time codes. Among such studies we can count [29, 30].

The focus of this chapter is on rank distance codes and how they are used to construct space-time block codes. In Section 4.1, we introduce rank distance codes and distinguish a special class of codes among all rank distance codes, called the *MRD codes*. A specific subclass of MRD codes is given in the same section. Then, in Section 4.2, how MRD codes can be used to construct full rank space-time block codes is described in some detail.

4.1 Rank Distance Codes

This section will present a brief introduction to rank distance codes. In the first part, basic definitions and notations about rank distance codes are introduced and maximal rank distance codes are defined. Secondly, a specific class of maximal rank distance codes are introduced.

4.1.1 Basic Definitions and Properties

Let $GF(q)$ denote the field with q elements and let $GF(q^m)$ be its extension field with q^m elements where $m \geq 1$. Note that $GF(q^m)$ can be viewed as a vector space of dimension m over $GF(q)$. We begin by defining the rank norm.

Definition 4.1.1 (Rank of a vector, [27]). *Let* $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in GF(q^m)^n$ *and let* $\mathbf{b} = \{\beta_1, \beta_2, \ldots, \beta_m\}$ *be a basis of* $GF(q^m)$ *over* $GF(q)$*. The rank of* $\mathbf{x}$ *over* $GF(q)$*, denoted by* $\mathrm{Rk}(\mathbf{x}|GF(q))$*, or simply by* $\mathrm{Rk}(\mathbf{x})$ *where there is no ambiguity, is defined as the rank of the* $m \times n$ *matrix* (x_{ij}) *where* $x_j = \sum_{i=1}^{m} \beta_i x_{ij}$ *for* $j = 1, 2, \ldots, n$*.*

Thus, in order to find the rank of a vector $\mathbf{x} \in GF(q^m)^n$ over $GF(q)$, we express each of its coordinates with respect to a fixed basis of $GF(q^m)/GF(q)$. By this way each coordinate is mapped to an $m \times 1$ column matrix, resulting in an $m \times n$ matrix having entries in $GF(q)$. The reader may easily verify that the rank of this matrix is independent of the choice of the basis. This shows that the rank of $\mathbf{x}$ is well defined. Rank norm induces a metric over $GF(q^m)^n$ as follows:

Definition 4.1.2 (Rank norm). *Let* $\mathbf{e}_1, \mathbf{e}_2 \in GF(q^m)^n$*. The rank distance of* $\mathbf{e}_1$ *and* $\mathbf{e}_2$ *is defined by*

$$d_R(\mathbf{e}_1, \mathbf{e}_2) = \mathrm{Rk}(\mathbf{e}_1 - \mathbf{e}_2). \tag{4.1}$$

The following can be considered as the analog of minimum Hamming distance from the theory of block codes:

Definition 4.1.3. *Let* $\mathcal{C} \subseteq GF(q^m)^n$ *be a code of block length* n*. The minimum rank distance of* $\mathcal{C}$ *is defined by*

$$d_R(\mathcal{C}) = \min_{\mathbf{c}_1 \neq \mathbf{c}_2 \in \mathcal{C}} d_R(\mathbf{c}_1, \mathbf{c}_2). \tag{4.2}$$

Given a code $\mathcal{C} \subseteq GF(q^m)^n$ with $|\mathcal{C}| = M$, we call it a $(n, M, d)_r - code$ if its minimum rank distance is equal to d. If, in addition, it is linear with dimension k, it is called a $[n, k, d]_r - code$.

The following reveals the existence of a Singleton-like bound between the parameters of a rank distance code:

Theorem 4.1.4 ([27]). *Let $\mathcal{C} \subseteq GF(q^m)^n$ be a rank distance code with $|\mathcal{C}| = M$ and $d_R(\mathcal{C}) = d$. Then the following inequality holds:*

$$M \leq q^{\min\{m(n-d+1), n(m-d+1)\}}. \tag{4.3}$$

Proof. Suppose that there exist two codewords $\mathbf{c}_1$ and $\mathbf{c}_2$ which have all their first $n-d+1$ coordinates the same. In this case, the first $n-d+1$ coordinates of $\mathbf{c}_1 - \mathbf{c}_2$ becomes 0 and consequently $d_R(\mathbf{c}_1, \mathbf{c}_2) < d$, which contradicts the assumption that $\mathcal{C}$ has minimum rank distance d. This shows that no two codewords from $\mathcal{C}$ can agree in all of the first $n-d+1$ coordinates. Therefore, the size of $\mathcal{C}$ is upper bounded by the number of all $(n-d+1)$-tuples over $GF(q^m)$. This shows that

$$M \leq |GF(q^m)^{n-d+1}| = q^{m(n-d+1)}. \tag{4.4}$$

To complete the proof, we fix any basis $\mathbf{b}$ of $GF(q^m)$ over $GF(q)$ and replace all coordinates of each codeword in $\mathcal{C}$ by their representation with respect to $\mathbf{b}$, as explained in Definition 4.1.1. By this way each codeword $\mathbf{c}$ is mapped to an $m \times n$ matrix having entries in $GF(q)$, which we denote by $\mathbf{B}(\mathbf{c})$. It can easily be verified that, for any two codewords $\mathbf{c}_1$ and $\mathbf{c}_2$, $\mathrm{Rk}(\mathbf{c}_1, \mathbf{c}_2)$ is equal to the rank of the matrix $\mathbf{B}(\mathbf{c}_1) - \mathbf{B}(\mathbf{c}_2)$. Since rank of a matrix is equal to its row rank, no two matrices from the set $\mathcal{B}(\mathcal{C}) = \{\mathbf{B}(\mathbf{c}) : \mathbf{c} \in \mathcal{C}\}$ can agree in all of the first $m-d+1$ rows. Thus, the number of elements in $\mathcal{B}(\mathcal{C})$ cannot exceed the size of the set of all $(m-d+1) \times n$ matrices over $GF(q)$, which is $q^{n(m-d+1)}$. But since the mapping from $\mathcal{C}$ to $\mathcal{B}(\mathcal{C})$ is one-to-one, we have $M \leq q^{n(m-d+1)}$. Combining this with (4.4) completes the proof. □

Corollary 4.1.5. *Let $\mathcal{C} \subseteq GF(q^m)^n$ be an $[n, k, d]_r$-code. Then*

$$mk \leq \min\{m(n-d+1), n(m-d+1)\}. \tag{4.5}$$

Given a rank distance code $\mathcal{C} \subseteq GF(q^m)^n$, (4.3) can also be expressed as a tradeoff between the rate and the minimum rank distance of $\mathcal{C}$. The rate R of $\mathcal{C}$ is defined in the same fashion as for the codes in Hamming metric. Namely, $R = (\log_{q^m} |\mathcal{C}|)/n = \log_q |\mathcal{C}| \ / \ mn$. Then, taking logarithm of both sides and dividing by q in (4.3) gives

$$R \leq 1 - \frac{d-1}{\min\{m,n\}}. \tag{4.6}$$

Definition 4.1.6. *A code $\mathcal{C} \subseteq GF(q^m)^n$ is called a maximal rank distance(MRD) code if it satisfies*

$$|\mathcal{C}| = q^{\min\{m(n-d+1),n(m-d+1)\}}. \tag{4.7}$$

Thus, MRD-codes constitute the class of codes whose ranks and rates achieve the tradeoff specified by the Singleton-like bound.

4.1.2 Gabidulin Codes

The proof of Theorem 4.1.4 gives no clue as to whether MRD-codes exist. In this section we summarize an explicit construction which was first given by Gabidulin in [27].

Let $n \leq m$ and $g_1, g_2, \ldots, g_n$ be n elements from $GF(q^m)$ which are linearly independent over $GF(q)$. We define the vector

$$\mathbf{g}[i] = (g_1^{q^i}, g_2^{q^i}, \ldots, g_n^{q^i}) \tag{4.8}$$

and consider the $k \times n$ matrix whose rows are $\mathbf{g}[i]$ for $i = 0, 1, \ldots, k-1$. More explicitly, we consider the matrix

$$\mathbf{G} = \begin{bmatrix} g_1 & g_2 & \cdots & g_n \\ g_1^q & g_2^q & \cdots & g_n^q \\ \vdots & \vdots & \ddots & \vdots \\ g_1^{q^{k-1}} & g_2^{q^{k-1}} & \cdots & g_n^{q^{k-1}} \end{bmatrix}. \tag{4.9}$$

The subspace of $GF(q^m)^n$ which is spanned by rows of $\mathbf{G}$ is called a *Gabidulin code*. In other words, a Gabidulin code is a rank distance code $\mathcal{C} \subseteq GF(q^m)^n$ which has a generator matrix of the above form. Thus, an arbitrary codeword from a Gabidulin code has the form

$$\mathbf{c} = (a_1 g_1 + a_2 g_1^q + \ldots + a_k g_1^{q^{k-1}}, \ldots, a_1 g_n + a_2 g_n^q + \ldots + a_k g_n^{q^{k-1}}) \tag{4.10}$$

where $a_1, a_2, \ldots, a_k \in GF(q^m)$ and $g_1, g_2, \ldots, g_n$ are linearly independent. A proof of the following can be found in [27]:

Proposition 4.1.7. *Gabidulin codes as described above are MRD-codes. Thus, they have minimum rank distance* $n - k + 1$.

Here we put an end to our discussion of rank distance codes. More detail can be found in [27, 31, 32, 33]. We proceed with how to use what we have seen to construct space-time block codes.

4.2 MRD-Codes as Space-Time Block Codes

We saw earlier that there is a tradeoff between the rate and diversity advantage of a space-time code. This tradeoff can be viewed as an analog of the Singleton-like bound between the parameters of a rank distance code, which we introduced in the previous section. Therefore, it is natural to suspect that rank distance codes can be used to construct good space-time codes. In the present section, this issue will be addressed.

4.2.1 Interpretation of Space-Time Codes as Rank Distance Codes

Let us recall the space-time coding scheme which we introduced in Section 2.2.1. Again, n_T and n_R are the number of transmit and receive antennas respectively. T is the frame length which we assume to satisfy $T \geq n_T$. Equation (2.5) can be expressed in the form of a matrix equation as

$$Y = \rho SH + W. \tag{4.11}$$

Here Y is the $T \times n_R$ received signal matrix where the (t, i) entry is the symbol received by receive antenna i at time slot t. H is the $n_T \times n_R$ channel matrix consisting of the path gains from transmit antennas to receive antennas. ρ is just a scaling factor and W is the $T \times n_R$ noise matrix. Finally, S is the $T \times n_T$ space-time code matrix where the (t, i) entry is the symbol transmitted by transmit antenna i at time slot t. To avoid confusion, we note that this time rows of S correspond to time slots whereas its columns correspond to transmit antennas, contrary to the convention we adopted in previous chapters. The reason will be clarified soon.

In this setting, let $\mathcal{S}$ be a space-time block code defined over a finite complex signal constellation $\mathcal{A} \subset \mathbb{C}$ of q elements where q is a power of some prime. Each codeword

$$\boldsymbol{s} = (s_1^1, s_1^2, \ldots, s_1^{n_T}, s_2^1, s_2^2, \ldots, s_2^{n_T}, \ldots, s_T^1, s_T^2, \ldots, s_T^{n_T}) \tag{4.12}$$

from $\mathcal{S}$ has a corresponding matrix representation as described above. Therefore, $\mathcal{S}$ can be considered as a matricial code which is a subset of $\mathrm{Mat}_{T\times n_T}(\mathcal{A})$, the set of all $T \times n_T$ matrices having entries in $\mathcal{A}$. The distance of two code matrices $S_1, S_2 \in \mathcal{S}$ is defined to be equal to $\mathrm{Rk}(S_1 - S_2)$. Then the minimum distance of $\mathcal{S}$ is seen to be equal to its diversity d.

In order to relate $\mathcal{S}$ to rank distance codes, we define the mapping $P : \mathrm{Mat}_{T\times n_T}(\mathcal{A}) \to (\mathcal{A}^T)^{n_T}$ where the i^{th} coordinate of $P(S)$ is the i^{th} row of S. P is clearly a bijection. This shows that each element of $\mathcal{S}$ can be represented by a unique element of $(\mathcal{A}^T)^{n_T}$. Since $\mathbb{C}$ does not have any finite subfield, $\mathcal{A}$ cannot be a field and hence space-time codes over complex signal constellations are not rank distance codes. On the other hand, the above explanations and the mapping P ensure that each space-time block code is in a one-to-one relation with a rank distance code. Furthermore, since we did not use the fact that the underlying code alphabet is a field in the proof of Theorem 4.1.4, it holds without any change for space-time block codes. Thus we have

$$|\mathcal{S}| \leq q^{T(n_T-d+1)}. \tag{4.13}$$

In the same manner as we expressed the Singleton-like bound for rank distance codes, we can express (4.13) in terms of the rate of $\mathcal{S}$. We previously defined the rate of a space-time code to be the number of constellation symbols transmitted per time slot. Therefore, the rate R of $\mathcal{S}$ is equal to $\frac{1}{T}\log_{|\mathcal{A}|}|\mathcal{S}|$. Since $|\mathcal{A}| = q$, taking logarithm of (4.13) and dividing by T yields

$$R \leq n_t - d + 1. \tag{4.14}$$

This is called the *rate-diversity tradeoff*. Notice that full rank codes have $d = n_T$ and the above implies $R \leq 1$ in this case. In a bit-rate sense we have $R \leq \log|\mathcal{A}|$ bits/s/Hz, which we had already proved in Corollary 2.4.2.

4.2.2 Space-Time Code Construction from Rank Distance Codes

As we have explained, rank distance codes cannot be used directly as space-time codes because wireless applications typically employs complex baseband symbols. The diversity criterion describes a rather simple relation between coding matrices consisting of complex constellation symbols; however, it is not clear what conditions it imposes on the unprocessed binary data before transmission. Several attempts have been made to settle this issue. In [34], the mapping from $GF(2)$ to the BPSK constellation was shown to preserve the rank distribution, which shows MRD-codes can be used in a straightforward manner to construct full rank space-time block codes for BPSK-modulation. Extension to PSK-modulation in general was also partially achieved by further results. A similar study was conducted for QAM-modulation in [35].

In this part, we summarize a space-time code construction method which first appeared in [29]. The underlying rank distance code is chosen to be a Gabidulin code of dimension 1. For this purpose, a primitive element α of $GF(q^{n_T})$ is chosen. Then $1, \alpha, \alpha^2, \ldots, \alpha^{n_T-1}$ are linearly independent elements of $GF(q^{n_T})$. If we set

$$\boldsymbol{c} = (1, \alpha, \ldots, \alpha^{n_T-1}), \tag{4.15}$$

then the one dimensional subspace $\mathcal{C}$ of $GF(q^{n_T})^{n_T}$ spanned by $\boldsymbol{c}$ is a Gabidulin code and hence a MRD-code of minimum rank distance n_T. This code has a more explicit representation. We begin with

$$\mathcal{C} = \{a \cdot \boldsymbol{c} : a \in GF(q^{n_T})\} \tag{4.16}$$

where "$\cdot$" denotes the scalar product and note that each nonzero element of $GF(q^{n_T})$ is equal to some power of α since α is primitive. In view of this, an even simpler description of $\mathcal{C}$ is given by

$$\mathcal{C} = \{\mathbf{0}\} \cup \{\alpha^i \cdot (1, \alpha, \ldots, \alpha^{n_T-1}) : i = 0, 1, \ldots, q^{n_T} - 2\}. \tag{4.17}$$

As explained before, codewords from $\mathcal{C}$ can be represented in the form of $n_T \times n_T$ matrices over $GF(q)$ through a map $Q : \mathcal{C} \to \mathrm{Mat}_{n_T \times n_T}(GF(q))$. For this purpose we consider the set $\mathbf{b} = \{1, \alpha, \ldots, \alpha^{n_T-1}\}$, which is linearly independent and hence forms a basis of $GF(q^{n_T})$ over $GF(q)$. Then coordinates of a codeword can be expanded

columnwise with respect to $\mathbf{b}$ to arrive at a matrix representation. For instance, the representation of the codeword $(\alpha, \alpha^2, \ldots, \alpha^{n_T})$ is equal to

$$C = \begin{bmatrix} 0 & 0 & \ldots & 0 & -b_0 \\ 1 & 0 & \ldots & 0 & -b_1 \\ 0 & 1 & \ldots & 0 & -b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -b_{n_T-1} \end{bmatrix}, \tag{4.18}$$

where the coefficients b_i come from $f(x) = x^{n_T} + b_{n_T-1}x^{n_T-1} + \ldots + b_1 x + b_0$, the irreducible polynomial of α. In the theory of finite fields, C is known as the *companion matrix* of f and is a representation of the field $GF(q^{n_T})$(for more explanation see [36, page 106]). Furthermore the mapping Q satisfies $Q(\alpha^i \cdot \boldsymbol{c}) = C^i$. Therefore, in view of (4.17), our original codeword $\mathcal{C}$ is mapped to

$$Q(\mathcal{C}) = \boldsymbol{\mathcal{C}} = \{0_{n_T \times n_T}, C, C^2, \ldots, C^{q^{n_T}-1}\}, \tag{4.19}$$

a matricial MRD code with cardinality q^{n_T} and minimum rank distance n_T. $\boldsymbol{\mathcal{C}}$ cannot be used directly as a space-time code since the transmitted symbols lie in a constellation $\mathcal{A}$ which consists of complex numbers. Instead, a one-to-one mapping from $GF(q)$ to $\mathcal{A}$ is used to replace code matrices from $\boldsymbol{\mathcal{C}}$ with space-time code matrices over $\mathcal{A}$. By this way we obtain a space-time block code of cardinality $GF(q^{n_T})$. If the mapping from $GF(q)$ to $\mathcal{A}$ preserves the rank distribution of matrices, the resulting space-time code is of full rank and hence satisfies the rank criterion. In general, there is no known method to construct a mapping which guarantees full rank over an arbitrary complex signal constellation. On the other hand, several methods which apply to a specific set of constellation alphabets has been developed in the literature. One such example from [29] follows from a result proved in [37]. Now we explain it shortly.

A *Gaussian integer* is a complex number with integer real and imaginary parts. A special type of Gaussian integer is when we consider a prime number $p \equiv 1 \pmod 4$. In this case, it is a known fact from number theory that p can be expressed as the sum of squares of two integers; i.e. $p = u^2 + v^2$ for integers u and v. In this case, the complex number $\Pi = u + \boldsymbol{i}v$ is known as a *Gaussian prime*. Let us consider the mapping from $GF(p)$ given by

$$\xi(k) = \zeta_k = k \ mod \ \Pi = k - \left[\frac{k\Pi^*}{u^2 + v^2}\right]\Pi \tag{4.20}$$

where * is complex conjugation as usual and [.] denotes the operation of rounding to the nearest Gaussian integer. It was shown in [37] that matrices over $GF(p)$ preserve their ranks after mapping them through ξ to matrices over Gaussian integers. More explicitly, if A is a matrix over $GF(p)$ and $\xi(A)$ denotes the matrix formed by replacing the (i,j) entry A_{ij} of A by $\xi(A_{ij})$, then A and $\xi(A)$ have the same rank. Therefore, ξ can be used to map any MRD-code to a space-time block code which is optimal with respect to the rate-diversity tradeoff. In particular, for the matricial code $\boldsymbol{\mathcal{C}}$ constructed above, $\xi(\boldsymbol{\mathcal{C}}) = \{0_{n_T \times n_T}, \xi(C), \xi(C^2), \ldots, \xi(C^{q^{n_T}-1})\}$ is a space-time block code of full rank over the constellation $\mathcal{A} = G_{\Pi}(p) = \{0, \zeta_1, \zeta_2, \ldots, \zeta_{p-1}\}$.

As an example, let us construct a full rank space-time block code for two transmit antennas. For this purpose we first construct the matricial counterpart of a $[2,1,2]_r$ MRD code over $GF(25)$ by the method explained above. We consider the primitive polynomial $f(x) = x^2 + 4x + 2$ over $GF(5)$. Its companion matrix is given by

$$C = \begin{bmatrix} 0 & 3 \\ 1 & 1 \end{bmatrix}. \tag{4.21}$$

Thus,the matricial code over $GF(5)$ defined by $\boldsymbol{\mathcal{C}} = \{0_{2\times 2}, C, C^2, \ldots, C^{24}\}$ is MRD with minimum rank distance 2. In fact, there is a more explicit way of defining $\boldsymbol{\mathcal{C}}$. If α is a root of f, just like $\{1, \alpha\}$ forms a basis of $GF(25)$ over $GF(5)$, $\boldsymbol{\mathcal{C}}$ is spanned by I_2 and C, where I_2 is the 2×2 identity matrix. Therefore,

$$\boldsymbol{\mathcal{C}} = \left\{ i_0 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + i_1 \begin{bmatrix} 0 & 3 \\ 1 & 1 \end{bmatrix} : i_0, i_1 \in GF(5) \right\} = \left\{ \begin{bmatrix} i_0 & 3i_1 \\ i_1 & i_0 + i_1 \end{bmatrix} : i_0, i_1 \in GF(5) \right\}$$

gives a more explicit description of $\boldsymbol{\mathcal{C}}$. It remains to map the elements of $GF(5)$ to Gaussian integers using the map ξ given in (4.20). Since $5 = 2^2 + 1^2$, $\Pi = 2 + \boldsymbol{i}$ is the Gaussian prime which defines ξ. Straightforward calculation shows that $\xi(0) = 0$, $\xi(1) = 1$, $\xi(2) = -\boldsymbol{i}$, $\xi(3) = \boldsymbol{i}$, $\xi(4) = -1$. Making these substitutions, we obtain a new set $\xi(\boldsymbol{\mathcal{C}})$ of 2×2 matrices over the signal constellation $\mathcal{A} = \{0, 1, -1, \boldsymbol{i}, -\boldsymbol{i}\}$. Since ξ preserves the ranks of matrices, $\xi(\boldsymbol{\mathcal{C}})$ is a full rank space-time block code for two transmit antennas having rate $R = 1$.

Construction of space-time codes that are optimal with respect to the rate-diversity tradeoff was studied more extensively in [30]. There, binary MRD codes was used to construct optimal space-time codes which applies to a wide set of signal constellations including QAM, PAM and PSK modulations. The reader is referred to [30] for details.

REFERENCES

[1] V. Tarokh, N. Seshadri and A. R. Calderbank, "Space-Time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction", *IEEE Trans. Inform. Theory*, Vol. 44(2), pp. 744-765, March 1998.

[2] C. E. Shannon, "A Mathematical Theory of Communication", *Bell System Technical Journal*, Vol. 27, pp. 379-423, 623-656, July, October, 1948.

[3] E. Telatar, "Capacity of Multi-antenna Gaussian Channels", *AT&T-Bell Labs Internal Tech. Memo.*, June 1995.

[4] G. J. Foschini, Jr. and M. J. Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas", *Wireless Personal Commun.*, Vol. 6(3), pp. 311-335, March 1998.

[5] A. J. Goldsmith, *Wireless Communications*, Cambridge University Press, New York, 2005.

[6] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*, Cambridge University Press, New York, 2005.

[7] A. Narula, "Information Theoretic Analysis of Multiple-Antenna Transmission Diversity", MIT PhD Dissertation, 1997.

[8] A. J. Goldsmith, "Design and Performance of High-Speed Communication Systems over Time-Varying Radio Channels", University of California at Berkeley PhD Dissertation, 1994.

[9] A. Narula, M. D. Trott, G. W. Wornell, "Information Theoretic Analysis of Multiple Antenna Transmission Diversity for Fading Channels", in *Proc. Int. Symp. Inform. Theory Appl.*, September 1996.

[10] M. Médard, "The Capacity of Time Varying Multiple User Channels in Wireless Communications", MIT PhD Dissertation, 1995.

[11] R. E. Blahut, *Theory and Practice of Error Control Codes*, Addison-Wesley, Massachusetts, 1983.

[12] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*, Prentice Hall, Englewood Cliffs, NJ, 1995.

[13] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambrdige University Press, New York, 1988.

[14] Z. Chen, J. Yuan and B. Vucetic, "An Improved Space-Time Trellis Coded Modulation Scheme on Slow Rayleigh Fading Channels", *Proc. IEEE ICC'01*, pp. 1110-1116, June 2001.

[15] R. J. McEliece, W. Lin, "The Trellis Complexity of Convolutional Codes", *IEEE Trans. Inform. Theory*, Vol. 42(6), pp. 1855-1864, November 1996.

[16] S. Bäro, G. Bauch and A. Hansmann, "Improved Codes for Space-Time Trellis Coded Modulation", *IEEE Commun. Lett.*, Vol. 4(1), pp. 20-22, January 2000.

[17] G. D. Forney and M. D. Trott, "The Dynamics of Group Codes: State Spaces, Trellis Diagrams, and Canonical Encoders", *IEEE Trans. Inform. Theory*, Vol. 39(9), pp. 1491-1513, September 1993.

[18] S. M. Alamouti, "A Simple Transmitter Diversity Scheme for Wireless Communications", *IEEE J. Select. Areas Commun.*, Vol. 16(8), pp. 1451-1458, October 1998.

[19] V. Tarokh, H. Jafarkhani and A. R. Calderbank, "Space-Time Block Codes from Orthogonal Designs", *IEEE Trans. Inform. Theory*, Vol. 44(2), pp. 744-765, March 1998.

[20] G. Abreu and R. Kohno, "Orthogonal Decoding of Space-Time Block Codes in Fast Fading", in *Proc. Int. Symp. Inform. Theory*, June-July 2003.

[21] X. Li, T. Luo, G. Yue, and C. Yin, "A Squaring Method to Simplify the Decoding of Orthogonal Space-Time Block Codes", *IEEE Trans. Commun.*, Vol. 49(10), pp. 1700-1703, October 2001.

[22] J. Radon, "Lineare Scharen Orthogonaler Matrizen", in *Abhandlungen aus dem Mathematischen Seminar der Hamburgishen Universität*, Vol. 1, pp. 1-14, 1922.

[23] H. Jafarkhani, *Space-Time Coding*, Cambridge University Press, New York, 2005.

[24] G. Ganesan and P. Stoica, "Space-Time Block Codes: A Maximum SNR Approach", *IEEE Trans. Inform. Theory*, Vol. 47(14), pp. 1650-1656, May 2001.

[25] Xue-Bin Liang, "Orthogonal Designs with Maximum Rates", *IEEE Trans. Inform. Theory*, Vol. 49(10), pp. 2468-2503, October 2003.

[26] W. Su, Xiang-Gen Xia, and K. J. Ray Lui, "A Systematic Design of High-Rate Complex Orthogonal Space-Time Block Codes", *IEEE Commun. Lett.*, Vol. 8(6), pp. 380-382, June 2004.

[27] E. M. Gabidulin, "Theory of Codes With Maximal Rank Distance", *Problems of Information Transmission*, Vol. 21, pp. 1-12, July 1985.

[28] A. V. Ourivski, E. M. Gabidulin, B. Honary, and B. Ammar, "Reducible Rank Codes and Their Applications to Cryptography", *IEEE Trans. Inform. Theory*, Vol. 49(12), pp. 3289-3293, December 2003.

[29] P. Lusina, E. M. Gabidulin, and M. Bossert, "Maximum Rank Distance Codes as Space-Time Codes", *IEEE Trans. Inform. Theory*, Vol. 49(10), pp. 2757-2760, October 2003.

[30] H. F. Lu and P. V. Kumar, "A Unified Construction of Space-Time Codes with Optimal Rate-Diversity Tradeoff", *IEEE Trans. Inform. Theory*, Vol. 51(5), pp. 1709-1730, May 2005.

[31] P. Loidreau, "Properties of Codes in Rank Metric", available at `http://arxiv.org/PS_cache/cs/pdf/0610/0610057v1.pdf`.

[32] M. Gadouleau and Z. Yan, "Properties of Rank Metric Codes", available at `http://arxiv.org/PS_cache/cs/pdf/0702/0702077v3.pdf`.

[33] E. M. Gabidulin and P. Loidreau, "Properties of Subspace Subcodes of Gabidulin Codes", *Adv. Math. Commun.*, Vol. 2(2), pp. 147-157, May 2008.

[34] A. R. Hammons and H. El Gamal, "On the Theory of Space-Time Codes for PSK Modulation", *IEEE Trans. Inform. Theory*, Vol. 46(2), pp. 524-542, March 2000.

[35] Y. Liu, M. P. Fitz and O. Y. Takeshita, "A Rank Criterion for QAM Space-Time Codes", *IEEE Trans. Inform. Theory*, Vol. 48(12), pp. 3062-3079, December 2002.

[36] F. MacWilliams and N. Sloane, *The Theory of Error Correcting Codes*, Amsterdam, The Netherlands: North Holland, 1993.

[37] M. Bossert, E. M. Gabidulin, and P. Lusina, "Space-Time Codes Based on Gaussian Integers", in *Proc IEEE Int. Symp. Inf. Theory*, p. 273, June 2002.