Cambridge University Press
978-0-521-06533-7 - Space-Time Block Coding for Wireless Communications
Erik G. Larsson and Petre Stoica
Excerpt
More information

Chapter 1

INTRODUCTION

The demand for capacity in cellular and wireless local area networks has grown in a literally explosive manner during the last decade. In particular, the need for wireless Internet access and multimedia applications require an increase in information throughput with orders of magnitude compared to the data rates made available by today's technology. One major technological breakthrough that will make this increase in data rate possible is the use of *multiple antennas* at the transmitters and receivers in the system. A system with multiple transmit and receive antennas is often called a multiple-input multiple-output (MIMO) system. The feasibility of implementing MIMO systems and the associated signal processing algorithms is enabled by the corresponding increase of computational power of integrated circuits, which is generally believed to grow with time in an exponential fashion.

1.1 Why Space-Time Diversity?

Depending on the surrounding environment, a transmitted radio signal usually propagates through several different paths before it reaches the receiver antenna. This phenomenon is often referred to as *multipath* propagation. The radio signal received by the receiver antenna consists of the superposition of the various multipaths. If there is no line-of-sight between the transmitter and the receiver, the attenuation coefficients corresponding to different paths are often assumed to be independent and identically distributed, in which case the central limit theorem [Papoulis, 2002, Ch. 7] applies and the resulting path gain can be modelled as a complex Gaussian random variable (which has a uniformly distributed phase and a Rayleigh distributed magnitude). In such a situation, the channel is said to be *Rayleigh fading*.

Since the propagation environment usually varies with time, the fading is time-variant and owing to the Rayleigh distribution of the received amplitude, the channel gain can sometimes be so small that the channel becomes useless. One way to mitigate this problem is to employ *diversity*, which amounts to transmitting the same information over multiple channels which fade independently of each other. Some common diversity techniques include time diversity and frequency diversity, where the same information is transmitted at different time instants or in different frequency bands, as well as antenna diversity, where one exploits the fact that the fading is (at least partly) independent between different points in space.

One way of exploiting antenna diversity is to equip a communication system with *multiple antennas at the receiver*. Doing so usually leads to a considerable performance gain, both in terms of a better link budget and in terms of tolerance to co-channel interference. The signals from the multiple receive antennas are typically combined in digital hardware, and the so-obtained performance gain is related to the diversity effect obtained from the independence of the fading of the signal paths corresponding to the different antennas. Many established communication systems today use receive diversity at the base station. For instance, a base station in the Global System for Mobile communications (GSM) [Mouly and Pautet, 1992] typically has two receive antennas. Clearly, a base station that employs receive diversity can improve the quality of the *uplink* (from the mobile to the base station) without adding any cost, size or power consumption to the mobile. See, for example, [Winters et al., 1994] for a general discussion on the use of receive diversity in cellular systems and its impacts on the system capacity.

In recent years it has been realized that many of the benefits as well as a substantial amount of the performance gain of receive diversity can be reproduced by using *multiple antennas at the transmitter* to achieve *transmit diversity*. The development of transmit diversity techniques started in the early 1990's and since then the interest in the topic has grown in a rapid fashion. In fact, the potential increase in data rates and performance of wireless links offered by transmit diversity and MIMO technology has proven to be so promising that we can expect MIMO technology to be a cornerstone of many future wireless communication systems. The use of transmit diversity at the base stations in a cellular or wireless local area network has attracted a special interest; this is so primarily because a performance increase is possible without adding extra antennas, power consumption or significant complexity to the mobile. Also, the cost of the extra transmit antenna at the base station can be shared among all users.

1.2 Space-Time Coding

Perhaps one of the first forms of transmit diversity was antenna hopping. In a system using antenna hopping, two or more transmit antennas are used interchangeably to achieve a diversity effect. For instance, in a burst or packet-based system with coding across the bursts, every other burst can be transmitted via the first antenna and the remaining bursts through the second antenna. Antenna hopping attracted some attention during the early 1990's as a comparatively inexpensive way of achieving a transmit diversity gain in systems such as GSM. More recently there has been a strong interest in *systematic* transmission techniques that can use multiple transmit antennas in an *optimal* manner. See [Paulraj and Kailath, 1993], [Wittneben, 1991], [Alamouti, 1998], [Foschini, Jr., 1996], [Yang and Roy, 1993], [Telatar, 1999], [Raleigh and Cioffi, 1998], [Tarokh et al., 1998], [Guey et al., 1999] for some articles that are often cited as pioneering work or that present fundamental contributions. The review papers [Ottersten, 1996], [Paulraj and Papadias, 1997], [Naguib et al., 2000], [Liu et al., 2001b], [Liew and Hanzo, 2002] also contain a large number of relevant references to earlier work (both on space-time coding and antenna array processing for wireless communications in general). However, despite the rather large body of literature on space-time coding, the current knowledge on optimal signal processing and coding for MIMO systems is probably still only the tip of the iceberg.

Space-time coding finds its applications in cellular communications as well as in wireless local area networks. Some of the work on space-time coding focuses on explicitly improving the performance of existing systems (in terms of the probability of incorrectly detected data packets) by employing extra transmit antennas, and other research capitalizes on the promises of information theory to use the extra antennas for increasing the throughput. Speaking in very general terms, the design of space-time codes amounts to finding a constellation of matrices that satisfy certain optimality criteria. In particular, the construction of space-time coding schemes is to a large extent a trade-off between the three conflicting goals of maintaining a simple decoding (i.e., limit the complexity of the receiver), maximizing the error performance, and maximizing the information rate.

1.3 An Introductory Example

The purpose of this book is to explain the concepts of antenna diversity and space-time coding in a systematic way. However, before we introduce the necessary formalism and notation for doing so, we will illustrate the fundamentals of receive and transmit diversity by studying a simple example.

1.3.1 One Transmit Antenna and Two Receive Antennas

Let us consider a communication system with one transmit antenna and two receive antennas (see Figure 1.1), and suppose that a complex symbol s is transmitted. If the fading is frequency flat, the two received samples can then be written:

$$\begin{aligned} y_1 &= h_1 s + e_1 \\ y_2 &= h_2 s + e_2 \end{aligned} \tag{1.3.1}$$

where h_1 and h_2 are the channel gains between the transmit antenna and the two receive antennas, and e_1, e_2 are mutually uncorrelated noise terms. Suppose that given y_1 and y_2, we attempt to recover s by the following *linear combination*:

$$\hat{s} = w_1^* y_1 + w_2^* y_2 = (w_1^* h_1 + w_2^* h_2) s + w_1^* e_1 + w_2^* e_2 \tag{1.3.2}$$

where w_1 and w_2 are weights (to be chosen appropriately). The SNR in $\hat{s}$ is given by:

$$\text{SNR} = \frac{|w_1^* h_1 + w_2^* h_2|^2}{(|w_1|^2 + |w_2|^2) \cdot \sigma^2} \cdot E\left[|s|^2\right] \tag{1.3.3}$$

where σ^2 is the power of the noise. We can choose w_1 and w_2 that maximize this SNR. A useful tool towards this end is the Cauchy-Schwarz inequality [Horn and Johnson, 1985, Th. 5.1.4], the application of which yields:

$$\text{SNR} = \frac{|w_1^* h_1 + w_2^* h_2|^2}{(|w_1|^2 + |w_2|^2) \cdot \sigma^2} \cdot E\left[|s|^2\right] \leq \frac{|h_1|^2 + |h_2|^2}{\sigma^2} \cdot E\left[|s|^2\right] \tag{1.3.4}$$

where equality holds whenever w_1 and w_2 are chosen proportional to h_1 and h_2:

$$\begin{aligned} w_1 &= \alpha \cdot h_1 \\ w_2 &= \alpha \cdot h_2 \end{aligned} \tag{1.3.5}$$

for some (complex) scalar α. The resulting SNR in (1.3.4) is proportional to $|h_1|^2 + |h_2|^2$. Therefore, loosely speaking, even if one of h_1 or h_2 is equal to zero, s can still be detected from $\hat{s}$. More precisely, if the fading is Rayleigh, then $|h_1|^2 + |h_2|^2$ is χ^2-distributed, and we can show that the error probability of detecting s decays as SNR_a^{-2} when $\text{SNR}_a \to \infty$ (by SNR_a here we mean the average channel SNR). This must be contrasted to the error rate for transmission

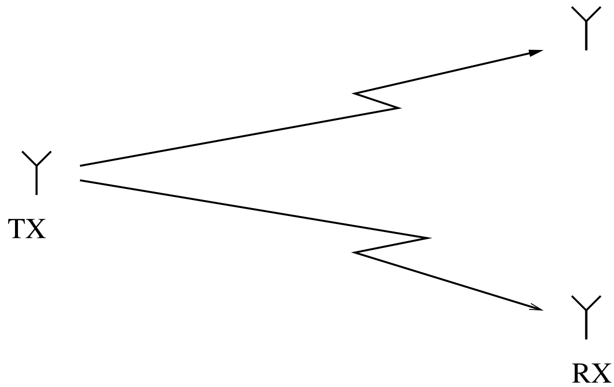

Figure 1.1. A system with one transmit antenna and two receive antennas.

and reception with a single antenna in Rayleigh fading, which typically behaves as SNR_a^{-1}.

In loose words, the *diversity order* of a system is the slope of the BER curve if plotted versus the average SNR on a log-log scale (a more formal definition is given in Chapter 4). Hence, we can say that the above considered system, provided that w_1 and w_2 are chosen optimally, achieves a diversity of order two.

1.3.2 Two Transmit Antennas and One Receive Antenna

Let us now study the "dual" case, namely a system with two transmit antennas and one receive antenna (see Figure 1.2). At a given time instant, let us transmit a symbol s, that is pre-weighted with two weights w_1 and w_2. The received sample can be written:

$$y = h_1 w_1 s + h_2 w_2 s + e \tag{1.3.6}$$

where e is a noise sample and h_1, h_2 are the channel gains. The SNR in y is:

$$\mathrm{SNR} = \frac{|h_1 w_1 + h_2 w_2|^2}{\sigma^2} \cdot E\left[|s|^2\right] \tag{1.3.7}$$

If w_1 and w_2 are fixed, this SNR has the same statistical distribution (to within a scaling factor) as $|h_1|^2$ (or $|h_2|^2$). Therefore, if the weights w_1 and w_2 are not allowed to depend on h_1 and h_2 it is impossible to achieve a diversity of order two. However, it turns out that if we assume that the transmitter knows the channel, and w_1 and w_2 are chosen to be functions of h_1 and h_2, it is possible to achieve

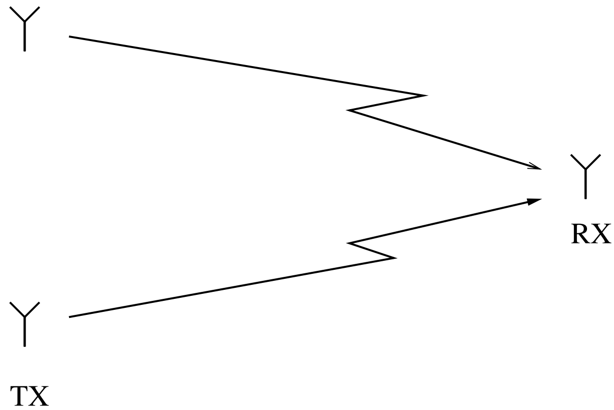

Figure 1.2. A system with two transmit antennas and one receive antenna.

an error probability that behaves as SNR_a^{-2}. We defer a deeper discussion of this aspect to Section 6.1.

We have seen that without channel knowledge at the transmitter, diversity cannot be achieved. However, if we are allowed to use more than one time interval for the transmission, we can achieve a diversity of order two rather easily. To illustrate this, suppose that we use *two* time intervals to transmit a single symbol s, where in the first interval only the first antenna is used and where during the second time interval only the second antenna is used. We get the following two received samples:

$$\begin{aligned} y_1 &= h_1 s + e_1 \\ y_2 &= h_2 s + e_2 \end{aligned} \tag{1.3.8}$$

Equation (1.3.8) is of the same form as (1.3.1) and hence the error rate associated with this method is equal to that for the case where we had one transmit and two receive antennas. However, *the data rate* is halved.

This simple example shows that transmit diversity is easy to achieve, if a sacrifice in information rate is acceptable. Space-time coding is concerned with the harder and more interesting topic: how can we maximize the transmitted information rate, at the same time as the error probability is minimized? This book will present some of the major ideas and results from the last decade's research on this topic.

1.4 Outline of the Book

Our book is organized as follows. We begin in Chapter 2 by introducing a formal model for the MIMO channel, along with appropriate notation. In Chapter 3, we study the promises of the MIMO channels from an information theoretical point of view. Chapter 4 is devoted to the analysis of error probabilities for transmission over a fading MIMO channel. In Chapter 5, we study a "classical" receive diversity system with an arbitrary number of receive antennas. This discussion sets, in some sense, the goal for transmit diversity techniques. In Chapter 6, we go on to discuss how transmit diversity can be achieved and also review some space-time coding methods that achieve such diversity. Chapter 7 studies a large and interesting class of space-time coding methods, namely linear space-time block coding (STBC) for the case of frequency flat fading. The case of frequency selective fading is treated in the subsequent Chapter 8. In Chapter 9 we discuss receiver structures for linear STBC, both for the coherent and the noncoherent case. Finally, Chapters 10 and 11 treat two special topics: space-time coding for transmitters with partial channel knowledge, and space-time coding in a multiuser environment.

1.5 Problems

1. Prove (1.3.3).

2. In (1.3.5), find the value of α such that

$$\hat{s} = s + e \tag{1.5.1}$$

where e is a zero-mean noise term. What is the variance of e? Can you interpret the quantity $\hat{s}$?

3. In Section 1.3.2, suppose that the transmitter *knows the channel* and that it can use this knowledge to choose w_1 and w_2 in an "adaptive" fashion. What is the SNR-optimal choice of w_1 and w_2 (as a function of h_1 and h_2)? Prove that by a proper choice of w_1 and w_2, we can achieve an error rate that behaves as SNR_a^{-2}.

4. In Section 1.3.2, can you suggest a method to transmit *two* symbols during two time intervals such that transmit diversity is achieved, without knowledge of h_1 and h_2 at the transmitter, and without sacrificing the transmission rate?

Cambridge University Press
978-0-521-06533-7 - Space-Time Block Coding for Wireless Communications
Erik G. Larsson and Petre Stoica
Excerpt
More information

Chapter 2

THE TIME-INVARIANT LINEAR MIMO CHANNEL

If we adopt the standard complex baseband representation of narrowband signals (see Appendix A), the input-output relation associated with a linear and time-invariant MIMO communication channel can easily be expressed in a matrix-algebraic framework. In this chapter, we will discuss both the frequency flat and the frequency-selective case. Models for single-input single-output (SISO), single-input multiple-output (SIMO) and multiple-input single-output (MISO) channels follow as special cases.

2.1 The Frequency Flat MIMO Channel

We consider a system where the transmitter has n_t antennas, and the receiver has n_r antennas (see Figure 2.1). In the current section we also assume that the bandwidth of the transmitted signal is so small that no intersymbol interference (ISI) occurs, or equivalently, that each signal path can be represented by a complex gain factor. For practical purposes, it is common to model the channel as frequency flat whenever the bandwidth of the system is smaller than the inverse of the delay spread of the channel; hence a wideband system operating where the delay spread is fairly small (for instance, indoors) may sometimes also be considered as frequency flat. Models for frequency-selective multiantenna channels, i.e., MIMO channels with non-negligible ISI, will be presented in Section 2.2.

Let $h_{m,n}$ be a complex number corresponding to the channel gain between transmit antenna n and receive antenna m. If at a certain time instant the complex signals $\{x_1, \ldots, x_{n_t}\}$ are transmitted via the n_t antennas, respectively, the received signal at antenna m can be expressed as

$$y_m = \sum_{n=1}^{n_t} h_{m,n} x_n + e_m \qquad (2.1.1)$$

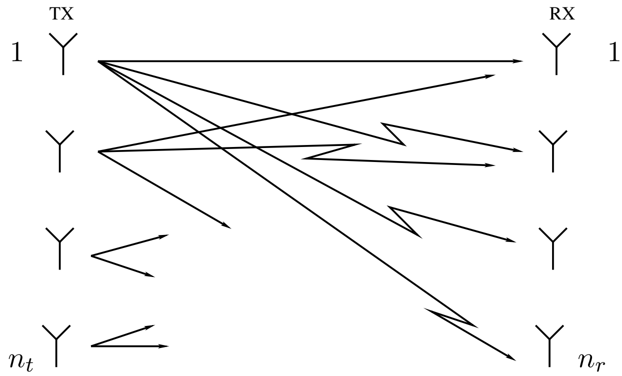

Figure 2.1. A MIMO channel with n_t transmit and n_r receive antennas.

where e_m is a noise term (to be discussed later). The relation (2.1.1) is easily expressed in a matrix framework. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be n_t and n_r vectors containing the transmitted and received data, respectively. Define the following $n_r \times n_t$ channel gain matrix:

$$\boldsymbol{H} = \begin{bmatrix} h_{1,1} & \cdots & h_{1,n_t} \\ \vdots & & \vdots \\ h_{n_r,1} & \cdots & h_{n_r,n_t} \end{bmatrix} \tag{2.1.2}$$

Then we have

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{e} \tag{2.1.3}$$

where $\boldsymbol{e} = [e_1 \ \cdots \ e_{n_r}]^T$ is a vector of noise samples. If several consecutive vectors $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ are transmitted, the corresponding received data can be arranged in a matrix

$$\boldsymbol{Y} = [\boldsymbol{y}_1 \ \cdots \ \boldsymbol{y}_N] \tag{2.1.4}$$

and written as follows:

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{E} \tag{2.1.5}$$

where

$$\boldsymbol{X} = [\boldsymbol{x}_1 \ \cdots \ \boldsymbol{x}_N] \tag{2.1.6}$$

and

$$\boldsymbol{E} = [\boldsymbol{e}_1 \ \cdots \ \boldsymbol{e}_N] \tag{2.1.7}$$

Note that vectorization of (2.1.5) yields the following equivalent model:

$$\boldsymbol{y} = \left(\boldsymbol{X}^T \otimes \boldsymbol{I}\right)\boldsymbol{h} + \boldsymbol{e} \tag{2.1.8}$$

where $\boldsymbol{y} = \text{vec}(\boldsymbol{Y})$, $\boldsymbol{h} = \text{vec}(\boldsymbol{H})$ and $\boldsymbol{e} = \text{vec}(\boldsymbol{E})$. This expression will be useful for performance analysis purposes.

2.1.1 The Noise Term

In this text, the noise vectors $\{\boldsymbol{e}_n\}$ will, unless otherwise stated, be assumed to be spatially white circular Gaussian random variables with zero mean and variance σ^2:

$$\boldsymbol{e}_n \sim N_C(\boldsymbol{0}, \sigma^2 \boldsymbol{I}) \tag{2.1.9}$$

Such noise is called additive white Gaussian noise (AWGN).

The Gaussian assumption is customary, as there are at least two strong reasons for making it. First, Gaussian distributions tend to yield mathematical expressions that are relatively easy to deal with. Second, a Gaussian distribution of a disturbance term can often be motivated via the central limit theorem.

Unless otherwise stated we will also assume throughout this book that the noise is temporally white. Although such an assumption is customary, it is clearly an approximation. In particular, the $\boldsymbol{E}$ term may contain interference consisting of modulated signals that are not perfectly white.

To summarize, the set of complex Gaussian vectors $\{\boldsymbol{e}_n\}$ has the following statistical properties:

$$\begin{aligned} E[\boldsymbol{e}_n \boldsymbol{e}_n^H] &= \sigma^2 \boldsymbol{I} \\ E[\boldsymbol{e}_n \boldsymbol{e}_k^H] &= \boldsymbol{0}, \quad n \neq k \\ E[\boldsymbol{e}_n \boldsymbol{e}_k^T] &= \boldsymbol{0}, \quad \text{for all } n, k \end{aligned} \tag{2.1.10}$$

2.1.2 Fading Assumptions

The elements of the matrix $\boldsymbol{H}$ correspond to the (complex) channel gains between the transmit and receive antennas. For the purpose of assessing and predicting the performance of a communication system, it is necessary to postulate a statistical distribution of these elements. This is true to some degree also for receiver design, in the sense that knowledge of the statistical behavior of $\boldsymbol{H}$ could potentially be used to improve the performance of the receiver.