# 1 Overview of MIMO communications

This chapter lays the foundations for the remainder of the book by presenting an overview of MIMO communications. Fundamental concepts and key terminology are introduced, and a summary of important matrix properties is provided, which will be referred to throughout the book. Some experimental results showing the benefits of MIMO are also presented.

## 1.1 What is MIMO?

Multiple Input Multiple Output communications, abbreviated MIMO, and normally pronounced like "My-Moe," refers to a collection of signal processing techniques that have been developed to enhance the performance of wireless communication systems using multiple antennas at the transmitter, receiver, or both. MIMO techniques improve communications performance by either *combating* or *exploiting* multipath scattering in the communications channel between a transmitter and receiver. MIMO techniques in the first category combat multipath by creating what is called *spatial diversity*, and those techniques that exploit multipath do so by performing *spatial multiplexing*. These two concepts are introduced in this chapter, and we will have much more to say about them throughout the remainder of the book. The subject of MIMO communications is the study of spatial diversity and spatial multiplexing techniques.

Figures 1.1 and 1.2 show block diagrams of generic MIMO communication systems. As indicated, the characteristics of the system depend on whether the focus of the MIMO processing is on creating spatial diversity, which improves reliability by combating fading, or if the purpose is to maximize throughput by performing spatial multiplexing. If the focus is on spatial diversity, information bits are normally encoded and modulated using conventional error correction coding and modulation techniques prior to undergoing some form of *space-time coding* (STC). At the receiver, space-time decoding is performed followed by demodulation and error decoding. If the focus is on spatial multiplexing, as illustrated in Figure 1.2, the information error encoded bits are passed through a serial-to-parallel converter and the individual output streams are modulated before being transmitted over separate antennas. At the receiver, each antenna receives a signal that consists of the sum of the signals from all of the transmit antennas; therefore, it is necessary to strip off each of the transmitted streams $\{s_i\}$ before demodulating them. The block that strips off each of the transmitted streams is often referred to as

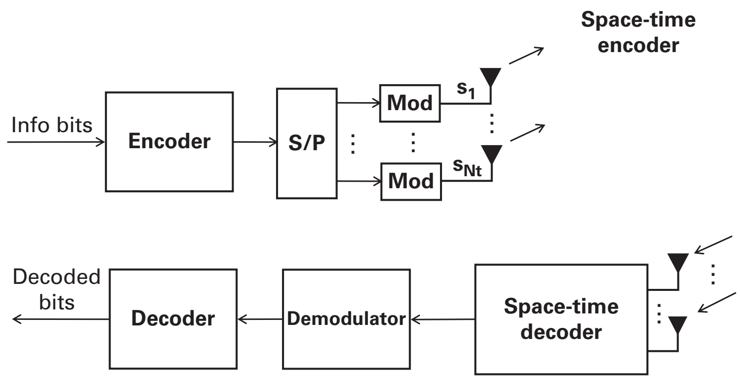


**Figure 1.1** A MIMO system for *spatial diversity*.

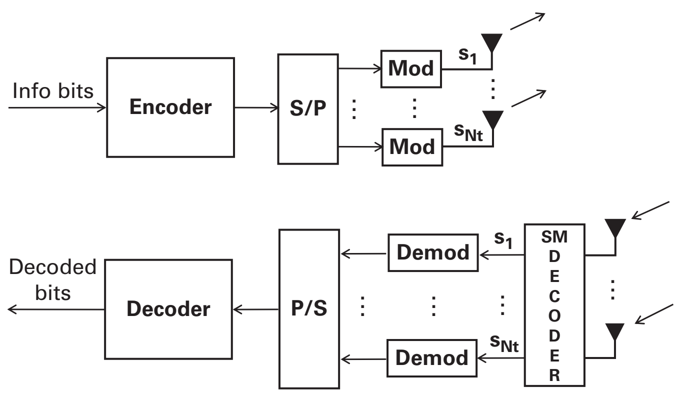


**Figure 1.2** A MIMO system for *spatial multiplexing*.

an *SM decoder* or demultiplexer. There are different types of spatial demultiplexing schemes based on zero-forcing or linear minimum mean square error based methods, which we discuss in detail in Chapter 8.

Table 1.1 summarizes the relationship between a variety of different terms that are used in the MIMO literature. At this point, there are two main concepts to be clear about. The first concept is that *spatial diversity refers to techniques that are used to improve the reliability* on a communications link by combating fading and that space-time coding is the means by which this is accomplished. The second concept is that *spatial multiplexing refers to techniques that are used to increase throughput* without increasing the required bandwidth by exploiting multipath. This is done by transmitting separate data streams on each of the transmit antennas and by separating those streams at the receiver using some form of spatial demultiplexing. The details of space-time coding are the subject of Chapters 6 and 7, and spatial multiplexing is covered in Chapters 3 and 8.

Strictly speaking, MIMO refers to communication systems that have multiple antennas at both the transmitter and receiver; however, the nomenclature can be a bit

**Table 1.1** Relationships of key MIMO concepts.

| MIMO technique | Purpose | Approach | Method |
|---|---|---|---|
| Spatial diversity | improve *reliability* | *combat* fading | space-time coding |
| Spatial multiplexing | increase *throughput* | *exploit* fading | spatial demultiplexing |

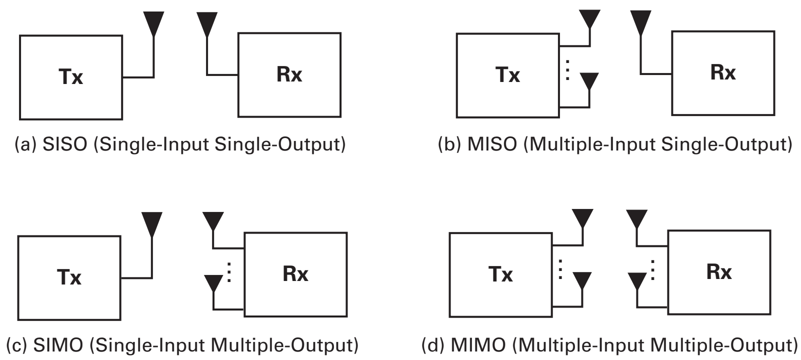


**Figure 1.3** Antenna configurations and their nomenclatures used in this book.

confusing on this point and there is not always agreement on the use of terminology. In this book, we use the term MIMO in two different ways: in a broad sense to refer to a communication system that has multiple antennas at either the transmitter, the receiver, or both, and in a particular way when referring to systems that have multiple antennas at both ends of the link. When there are multiple antennas at the transmitter and only one receiver, as may occur, for example, on a cellular forward link between the base station and a single mobile user, we call that type of system a Multiple Input Single Output (MISO) system. When the opposite is true and there are multiple receive antennas but only one transmit antenna, that system is called a Single Input Multiple Output (SIMO) system. When using the term in the broad sense, we often refer to MISO and SIMO systems as particular types of MIMO configurations. Conventional communication systems that only have a single transmit antenna and a single receive antenna are called Single Input Single Output (SISO) communication systems. Figure 1.3 illustrates the four types of antenna configurations and the nomenclature used in this book.

MIMO systems with $N_t$ transmit antennas and $N_r$ receive antennas are referred to as $N_t \times N_r$ MIMO systems. Thus, for example, a $2 \times 4$ MIMO system implies that there are two transmit antennas and four receive antennas.

## 1.2 History of MIMO

The phrase "Multiple Input Multiple Output" has an interesting history. Although now it is used to describe the communication techniques that are the subject of this book, it was originally used in electric circuit and filter theory as far back as the 1950s [23].

In that original context, "MIMO" referred to circuits that had multiple input and multiple output ports. In the 1990s, however, information theorists and communication system researchers adopted this term to refer to new signal processing techniques that they were developing for communication systems having multiple antennas. In this newer use of the term, the communications channel was the reference point, and the term *multiple input* referred to the signals from multiple transmit antennas that were"entering" or "being input" to the communications channel. Similarly, the term *multiple output* referred to signals arriving at multiple receiver antennas, which were viewed as "exiting" or "being output" from the channel. The first reference to the term MIMO in this newer communications sense was in a paper by Peter Driessen and Gerry Foschini in 1999 where they published an analysis on the theoretical communications capacity of a communication system with multiple transmit and multiple receive antennas [20].

Although MIMO communications requires the use of multiple antennas, it is not the first multi-antenna technique to be developed. So what's new or unique about MIMO? To help answer that question, it is useful to place MIMO in its proper historical context. We begin by recognizing that the idea of using multiple antennas to improve aspects of communications and radar performance goes back to the beginning of the 1900s. The first use of multiple antennas was for the purpose of creating phased array antennas, which were first proposed and then demonstrated in 1905 by Karl Braun [12]. During WW II, phased array technology was used to enable rapidly-steerable radar [7], and later, phased arrays were used in AM broadcast radio to switch from groundwave propagation during the day to skywave propagation at night. This was accomplished by switching the phase and power levels supplied to the individual antenna elements daily at sunrise and sunset so that the elevation angle of the radiation pattern was towards the horizon during daylight hours and pointed slightly upward at night. This had the obvious advantage of enabling the transmitter to change the direction that it emitted energy without having to mechanically point the antenna, a challenging feat with large antennas such as those used in AM radio. Phased array technology has also long been used to perform adaptive nulling for interference and jamming avoidance.

In addition to phased array applications, multi-antenna technology has been used for more than 70 years to reduce the impact of fading on communication systems through the use of receive diversity. An early paper on the concept of receive diversity was published by H. Beverage and H. Peterson [11] in 1931. In the 1950s, receive diversity combining found extensive application on troposcatter links for military applications in which radio waves are scattered within the troposphere layer of the atmosphere [2], [84]. The scattering that occurs on troposcatter links enables communications beyond the horizon, which, other than HF, was the only way to communicate beyond the horizon prior to the advent of satellite communications. Troposcatter links were found to suffer from significant fading effects, so multiple antennas at the receiver were used to create receive diversity, which was helpful in reducing the impact of the fading.

Beginning in the 1990s, two new types of multi-antenna techniques were developed, which are the subject of this book. One of these techniques uses multiple antennas to achieve *transmit diversity*, which, like receive diversity, reduces the effect of fading. Two early papers on this technique were published in 1991 and 1993 by A. Wittneben [81] and N. Seshadri, C. Sundberg, and V. Weerackody [68], respectively. Later, Alamouti [6] published a landmark paper that described another way to achieve transmit diversity that required less processing at the receiver. Alamouti's technique has since become one of the most popular MIMO schemes in use today by nearly all wireless systems. His paper described a simple space-time coding technique for achieving transmit diversity and spurred research into other space-time coding techniques.

At about the same time that research was being conducted on transmit diversity, another class of multi-antenna techniques was being developed. Unlike those who were researching ways to use multiple antennas to combat the effects of fading, this second group of researchers was interested in developing ways of exploiting fading to support increased throughput capacity. In 1996, Gerry Foschini at AT&T Research Labs published his landmark paper on layered space-time communications, which described the underlying concept for the class of spatial multiplexing techniques that would eventually be called the Bell-Labs Layered Space-Time (BLAST) schemes [30]. In 1998 Foschini and a team from AT&T Research Labs were the first to demonstrate a laboratory prototype system that implemented a particular type of BLAST technique called vertical BLAST (i.e., V-BLAST) [31].

Since these initial breakthroughs in spatial diversity and spatial multiplexing in the late 1990s, a large body of a research has been conducted, and MIMO techniques using the spatial diversity and spatial multiplexing methods emerging from this research have been adopted in an increasing number of commercial wireless standards. The first commercial MIMO technology was introduced by Iospan Wireless Inc. in 2001. Since 2005, when the WiMAX standard first included MIMO technology, most wireless standards now include MIMO.

Figure 1.4 shows a time line of some key breakthroughs in multi-antenna technology over the past century. This diagram and the discussion above indicate that MIMO can be viewed as the latest in a long line of advances in multi-antenna technology.

## 1.3 Smart antennas vs MIMO

In recent years, another multi-antenna term, *smart antennas*, has become popular in the literature. What are smart antennas and what is the difference between MIMO and smart antenna technology? There is not unanimous agreement on the answer to this question. One of the first researchers to use the term smart antennas was Jack Winters at AT&T Labs [80]. In his 1998 paper, he focuses on describing ways to dynamically generate beams at a cellular base station that point in the desired direction of mobile users, and ways to create nulls that point in directions of interference. In another paper by Angeliki Alexiou and Martin Haardt in 2004 [4], however, the term smart antennas is used in a much broader sense to include not only dynamic beamforming and antenna nulling, but also spatial multiplexing and spatial diversity techniques such as Alamouti's scheme. In their use of the term, MIMO is a subset of smart antenna technology.

In this book, we use the older original concept to delineate between MIMO and smart antennas. For our purposes, smart antennas are defined as systems that employ

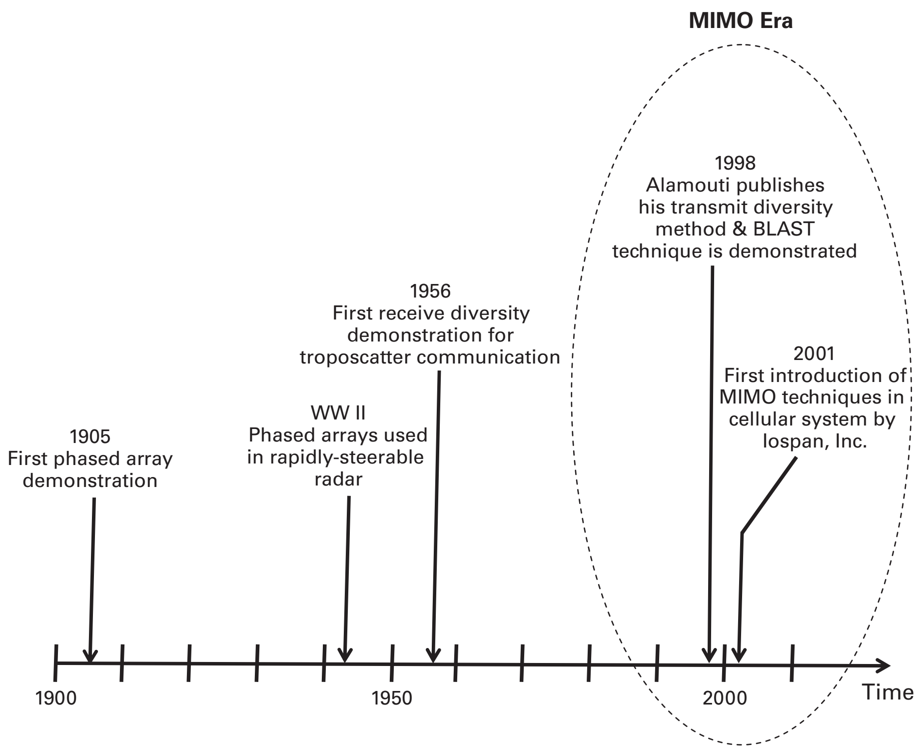


**Figure 1.4** Time line of key multi-antenna advances.

techniques that are primarily designed to form beams and nulls in desired directions based on feedback from the environment. MIMO techniques, in contrast, are defined as communication systems that involve baseband signal processing techniques such as space-time coding and spatial multiplexing schemes that are not focused on pointing beams or creating nulls in space. In summary, we distinguish between smart antennas and MIMO as follows:

**Smart antennas focus on:**
- Conventional beamforming – directing energy in a desired physical direction;
- Adaptive nulling – creating nulls in desired directions to reduce interference.

**MIMO focuses on:**
- Spatial diversity – combating fading effects by creating spatial diversity through the use of baseband space-time coding techniques;
- Spatial multiplexing – using spatial multiplexing techniques to exploit multipath in order to achieve higher data rates than are possible with conventional systems having the same bandwidth.

## 1.4 Single-user and multi-user MIMO

Before proceeding further, a few comments should be made regarding the terms *single-user* MIMO and *multi-user* MIMO, which have been coined to describe two classes of

MIMO communications that are used in wireless systems, such as LTE and WiMAX. Single-user MIMO (SU-MIMO) refers to conventional MIMO where there is a one transmitting node and one receiving node, and the transmitter node has multiple antennas, as illustrated in Figure 1.3 (b) and (d).

In multi-user MIMO (MU-MIMO), mobile cellular users, each with a single antenna, transmit to a base station, and the base station processes the signals from each of the individual mobiles as if they were coming from multiple transmit antennas on a single node. In this case, the base station performs the same operations as the receiver in Figure 1.2, so multiple mobile users can transmit data over the same bandwidth, and the base station is able to decouple the individual data streams using spatial decoding techniques. In MU-MIMO, the individual users will not experience increased throughput; however, the overall system will. That is, MU-MIMO allows more cellular users to transmit simultaneously on the uplink path over the *same bandwidth* than would otherwise be possible.

The focus of this book is on SU-MIMO; however, with the exception of eigenbeamforming, which is a spatial multiplexing technique described in Chapter 3, the spatial multiplexing techniques we describe can be used with both SU- and MU-MIMO.

## 1.5 Introduction to spatial diversity

As we have just explained, one of the key purposes of MIMO communications is to improve communications reliability by combating multipath fading, which is achieved through the creation of spatial diversity. In this section, we review the concept of diversity, describe the difference between receive and transmit spatial diversity, and define three important performance metrics: *diversity order*, *diversity gain*, and *array gain*.

### 1.5.1 The concept of diversity

In most environments where wireless communication systems operate, the strength of the received signal varies with time, which is called *fading*. Unfortunately, fading significantly degrades communications performance by causing the probability of bit error to increase compared to what it would be if only white noise were present. Figure 1.5 shows the probability of bit error as a function of bit-energy-to-noise power spectral density, $E_b/N_0$, for different types of modulation in both fading and non-fading environments. The results in this figure demonstrate two important characteristics. The first is simply that fading causes the error probability to increase dramatically for a given value of $E_b/N_0$. The second observation is that for Rayleigh fading, which is the type of fading assumed in this figure and that often occurs in practice, the error probability decreases linearly when plotted on a logarithmic scale against $E_b/N_0$ plotted in dB. This is an important observation, and we use it later in this chapter.

In order to reduce the impact of fading, the concept of diversity is often employed. Diversity refers to transmitting replicas of the same signal over a fading channel in such a way that each replica fades independently of the others. When this happens, each replica tends to fade at a different time, so the probability that all the replicas

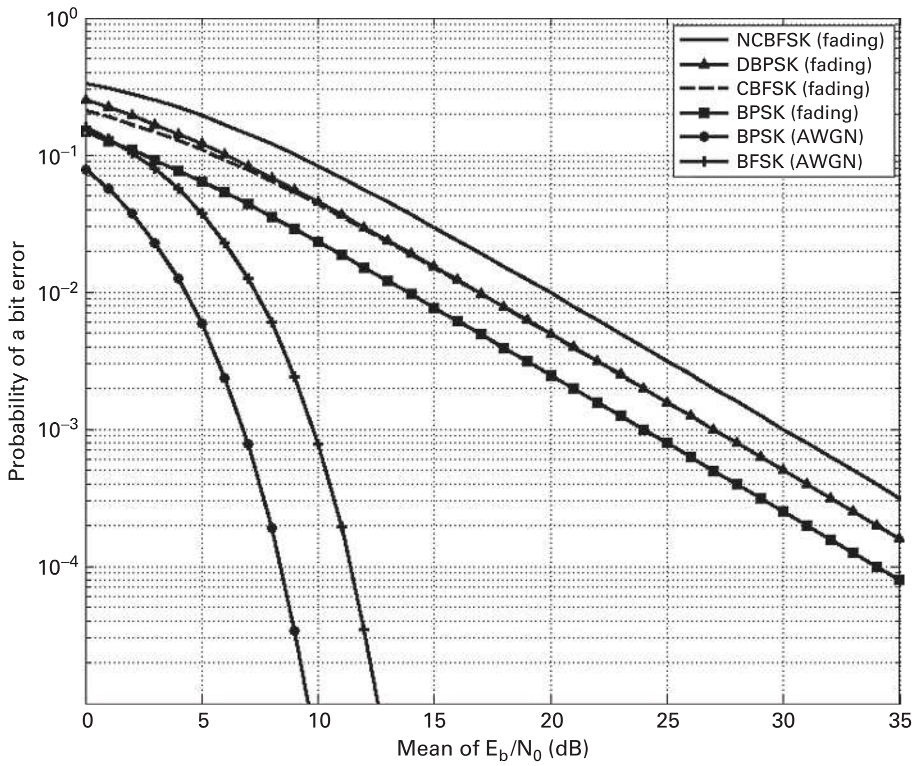


**Figure 1.5** Performance of binary signaling on a Rayleigh fading channel.

fade simultaneously decreases as the number of replicas gets larger. By combining the replicas, however, the depths of the fades, and, so too, their adverse effects, can be significantly reduced because the fades do not tend to occur at the same time.

Reducing the impact of fading through diversity, therefore, involves two steps:

a) creating independent replicas of the signal; and
b) combining the replicas.

There are various ways to generate replicas of a signal for diversity purposes. One is to transmit the signal on different RF frequencies that are spaced far enough apart that the fading occurs independently on each carrier. This is called *frequency diversity*. Another diversity technique, called *time diversity*, involves transmitting the same signal at different times. In a multipath environment, this occurs naturally because the same signal arrives at the receiver by traveling over multiple physical paths, which tend to experience independent fading. Rake receivers are used to process such signals. A third way to create diversity is to transmit the same information on signals having different polarizations, called *polarization diversity*. Normally, fading is independent of signals having different polarizations. A fourth type of diversity is called *spatial diversity*, which refers to transmitting the same information over different physical paths between the transmitter and receiver. One way to create spatial diversity is to transmit a signal from one transmit antenna and receive it using multiple receive antennas. If the receive antennas are far enough apart, the fading on each path will be independent. This is the type of diversity that was originally used in the 1950s to reduce the impact of fading on troposcatter links discussed earlier.

Just as there are multiple ways to generate independent replicas of a signal, there are also different ways to combine the replicas at the receiver. The simplest type of combining is called *selective combining*, which involves comparing the replicas at each sample time and choosing the largest value for the output of the combiner. A second combining technique, called *equal gain combining*, involves adding the replicas together. The third, and most common type of combining scheme, is called *maximal ratio combining* (MRC). In MRC, the replicas are added together in the same way as they are in equal gain combining, but prior to being added they are first scaled in proportion to the signal-to-noise ratio of each replica. In Chapter 6, we discuss MRC in greater detail.

Figure 1.6 illustrates the benefits of diversity combining by plotting the output amplitude of a selective combiner in the presence of Rayleigh fading for two cases: when there is no combining (i.e., the number of signals being combined is 1), and when there are five replicas being combined. The curve associated with no combining has the deepest fades and the curve associated with five combined signals has noticeably less fading. Similar improvements occur with the other combining techniques.

### 1.5.2 Receive and transmit diversity

As we discussed earlier, troposcatter was one of the first types of communications techniques to use diversity combining. From that time until the 1990s, diversity techniques involved transmitting a single version of a signal and extracting replicas of the transmitted signal at the receiver and then combining those replicas. Diversity of this type is called *receive diversity* because extraction of the replicas is performed at the receiver. Figure 1.7 illustrates the architecture of a communication system that implements spatial receive diversity. As shown in this figure, the transmitted signal is denoted by $s$, and the communications channel has the effect of multiplying the transmitted signal by a complex value, which we call the channel response, and denote by $h_i, i = 1, \ldots, N_r$, where $N_r$ represents the number of receive antennas. The inputs to the combiner, therefore, consist of the set of signals $\{r_i = h_i s\}$. If the receive antennas are spaced far enough apart, the random variables $\{h_i\}$ are independent, so the receiver is able to reduce the effect of fading by combining multiple independently fading signals.

In the late 1980s and early 1990s with the growing use of cellular communications, a desire for a different type of diversity architecture arose, called *transmit diversity*. The motivation for developing transmit diversity was the fact that the mobile unit in most cellular systems is small and, as a result, is often not capable of having multiple antennas. As a result, receive diversity on the forward link of cellular systems may not be possible. This led to the desire to find a spatially-based method of creating replicas of the transmitted signal at a receiver having only one antenna.

A little thought shows that this is not trivial. For example, if the base station is assumed to have multiple antennas, and if a signal, $s$, is simply transmitted from each of

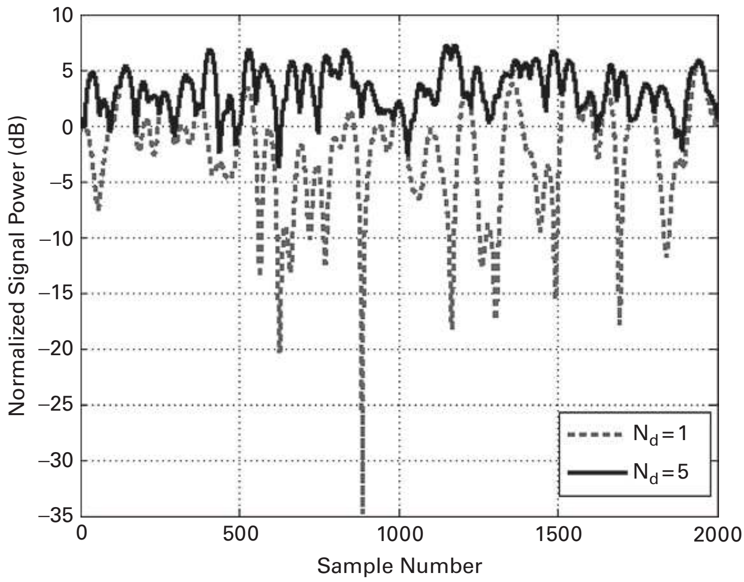


**Figure 1.6** Simulated output from an equal gain combiner in Rayleigh fading for different numbers of combined signals.

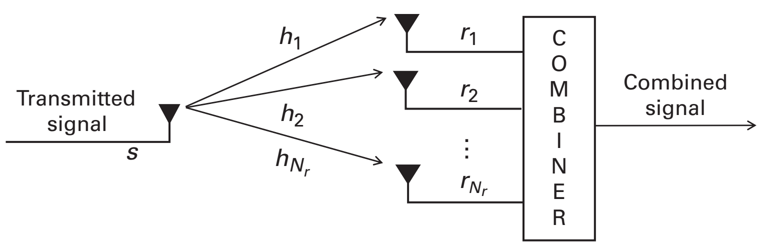


**Figure 1.7** Architecture of a communication system with receive diversity combining.

these antennas, then the received signal, $r$, at the single receive antenna on the mobile unit is given by

$$r = s \left[ \sum_{i=1}^{N_r} h_i \right]. \tag{1.1}$$

Unfortunately, this shows that the received signal is simply a scaled version of the transmitted signal, so the combiner at the mobile unit would not have access to multiple replicas of the transmitted signal if this technique were used. To correct this problem, it is necessary to perform some type of space-time coding at the transmitter. The architecture of a system with transmit diversity is depicted in Figure 1.8. One simple space-time code is the Alamouti code, which is used in most MIMO systems today. We discuss the Alamouti and other space-time codes in Chapters 6 and 7 and show how these schemes