

1.1 The Crowded Spectrum

It is all about spectrum. Marconi pioneered the wireless industry 100 years ago. Today life does not seem possible without wireless in some form or the other. In fact wireless permeates every aspect of our lives. The demands on bandwidth and spectral availability are endless. Currently wireless finds its widest expression in fixed and mobile roles. In the fixed role, wireless is used extensively for data transfer, especially from desktop computers and laptops. In the mobile role, wireless networks provide mobility for use from fast vehicles for both voice and data. Consequently wireless designers face an uphill task of limited availability of radio frequency spectrum and complex time varying problems in the wireless channel, such as fading and multipath, as well as meeting the demand for high data rates.

Simultaneously, there is an urgent need for better quality of service (QoS), compared with that obtainable from DSL and cable.

1.2 Need for High Data Rates

The gradual evolution of mobile communication systems follows the quest for high data rates, measured in bits/sec (bps) and with a high spectral efficiency, measured in bps/Hz. The first mobile communications systems were analog and are today referred to as systems of the first generation. In the beginning of the 1990s, the first digital systems emerged, denoted as second generation (2G) systems. In Europe, the most popular 2G system introduced was the global system for mobile communications (GSM), which operated in the 900 MHz or the 1,800-MHz band and supported data rates up to 22.8 kbit/s. In many parts of the world today, GSM is still in vogue. Basically GSM is a cellular system [i.e., it typically uses a single base transceiver station (BTS), which marks the center of a cell and which serves several mobile stations (MS), meaning the users].

In the United States, the most popular 2G system is the TDMA/136, which is also a digital cellular system. TDMA stands for time-division multiple access. To accomplish higher data rates, two add-ons were developed for GSM, namely high-speed circuit switched data (HSCSD) and the general packet radio service (GPRS), providing data rates up to 38.4 kbit/s and 172.2 kbit/s, respectively.

The demand for yet higher data rates forced the development of a new generation of wireless systems, the so-called third generation (3G). 3G systems are characterized by a maximum data rate of at least 384 kbit/s for mobile and 2 mbit/s for indoors.

One of the leading technologies for 3G systems is the now well known universal mobile telephone system (UMTS) [also referred to as wideband code-division multiplex (WCDMA) or UTRA FDD/TDD]. UMTS represents an evolution in terms of services and data speeds from today's "second generation" mobile networks. As a key member of the global family of 3G mobile technologies identified by the International Telecommunication Union (ITU), UMTS is the natural evolutionary choice for operators of GSM networks, currently representing a customer base of more than 850 million end users in 195 countries and representing over 70% of today's digital wireless market. UMTS is already a reality.

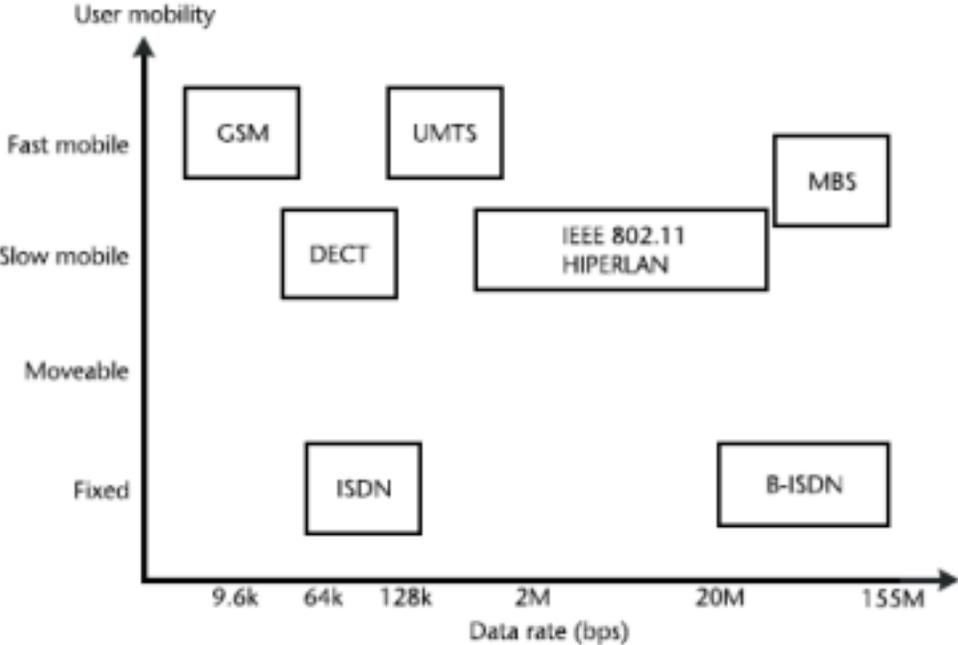
Japan launched the world's first commercial WCDMA network in 2001, and WCDMA networks are now operating commercially in Austria, Italy, Sweden and the United Kingdom, with more launches anticipated during 2004. Several other pilot and precommercial trials are operational in the Isle of Man, Monaco, and other European territories. UMTS is also a cellular system and operates in the 2-GHz band. Compared with the 2G systems, UMTS is based on a novel technology. To yield the 3G data rates, an alternative approach was made with the enhanced data rates for GSM evolution (EDGE) concept.

The EDGE system is based on GSM and operates in the same frequency bands. The significantly enhanced data rates are obtained by means of a new modulation scheme, which is more efficient than the GSM modulation scheme. As for GSM, two add-ons were developed for EDGE, namely enhanced circuit switched data (ECS-D) and the enhanced general packet radio service (EGPRS). The maximum data rate of the EDGE system is 473.6 kbit/s, which is accomplished by means of EGPRS. EDGE was introduced in the United States as a generic air interface to enhance the TDMA/136 system. Some 200 operators worldwide are also giving their customers a taste of faster data services with so called 2.5G systems based on GPRS technology, a natural evolutionary steppingstone toward UMTS.

The new IEEE and High Performance Radio Local Area Network (HIPERLAN) standards specify bit rates of upto 54 mbit/s, although 24 mbit/s will be the typical rate used in most applications. Such high data rates impose large bandwidths, thus pushing carrier frequencies for values higher than the UHF band. HIPERLAN has frequencies allocated in the 5- and 17-GHz bands; multimedia broadcasting systems (MBS) will occupy the 40- and 60-GHz bands; and even the infrared band is being considered for broadband wired local area networks (WLANs).

A comparison of several systems, based on two of the key features (mobility and data rate) is shown in Figure 1.1, where it is clear that no competition exists between the different approaches. The applications and services of the various systems are also different. IEEE 802.11 is

mainly intended for communications between computers (thus being an extension of WLANs).



Figure

1.1 Comparison of mobility and data rates for several systems.

Future wireless broadband applications are likely to require data rates that are hundreds of megabits per second—up to 250 times the maximum data rate promised for UMTS. Such a broadband service could, for example, be wireless high-quality video conferencing (upto 100 mbit/s) or wireless virtual reality (up to 500 mbit/s, when allowing free body movements).

Therefore, the goal of the next generation of wireless systems—the fourth generation (4G)—is to provide data rates yet higher than the ones of 3G while granting the same degree of user mobility. 4G is the short term for fourth generation wireless, the stage of broadband mobile communications that will follow the still burgeoning 3G that is expected to reach maturity between 2003–2005. 4G services are expected to be introduced first in Japan, as early as 2006—four years ahead of the previous target date. The major distinction of 4G over 3G communications is increased data transmission rates, just as it is for 3G over 2G and 2.5G (the current state of wireless services, hovering somewhere between 2G and 3G).

According to NTT-DoCoMo, the leading Japanese wireless company, the current download speed for i-Mode (mobile Internet service) data is, theoretically, 9.6 kbit/s, although in practice the rates tend to be slower. 3G rates are expected to reach speeds 200 times higher, and 4G to yield further increases, reaching 20–40 mbit/s (about 10–20 times the current rates of ADSL service). 4G is expected to deliver more advanced versions of the same improvements promised by 3G, such as enhanced

multimedia, smooth streaming video, universal access, and portability across all types of devices.

Industry insiders are reluctant to predict the direction that less than-immediate future technology might take, but 4G enhancements are expected to include worldwide “roaming” capability. As was projected for the ultimate 3G system, 4G might actually connect the entire globe and be operable from any location on—or above—the surface of the earth. This aspect makes it distinctly different from the technologies developed until now.

The first 4G systems are likely to be an integration of 3G systems and WLAN systems. By this means, considerable data rates can be granted at hot spots. On the other hand, the interworking of WLAN and 3G systems will provide a good degree of mobility, given that seamless handover is accomplished between several heterogeneous systems. Figure 1.2 reveals the possible candidates for 4G systems. The figure is self explanatory. There are, however, a few interesting points. The “hottest” candidates for 4G appear to be:

BWIF:

The Broadband Wireless Internet Forum (BWIF) [3] is the principal organization chartered with creating and developing next generation fixed wireless standards. The broadband wireless specifications are based on vector orthogonal frequency division multiplexing (VOFDM) technology and data over cable service interface specification (DOCSIS). BWIF was formed to address the needs of the quickly emerging wireless broadband market. Further, through BWIF, members establish product road maps that lower product costs, simplify deployment of advanced services, and ensure the availability of interoperable solutions. BWIF extends the partnership model to all companies offering expanded broadband wireless technology to multiple markets. At the core of the partnerships, membership includes:

- ASIC semiconductor companies, which develop new ASICs based on VOFDM technology.
- Customer premise equipment (CPE) companies, which use the chips to build new subscriber equipment.
- Systems integrators, which design and deploy the networks based on these products.
- Service providers, which incorporate VOFDM products and technology into their network infrastructure to offer to new services customers.
- RF/ODU manufacturers, which supply subsystems to the total wireless solution offering.

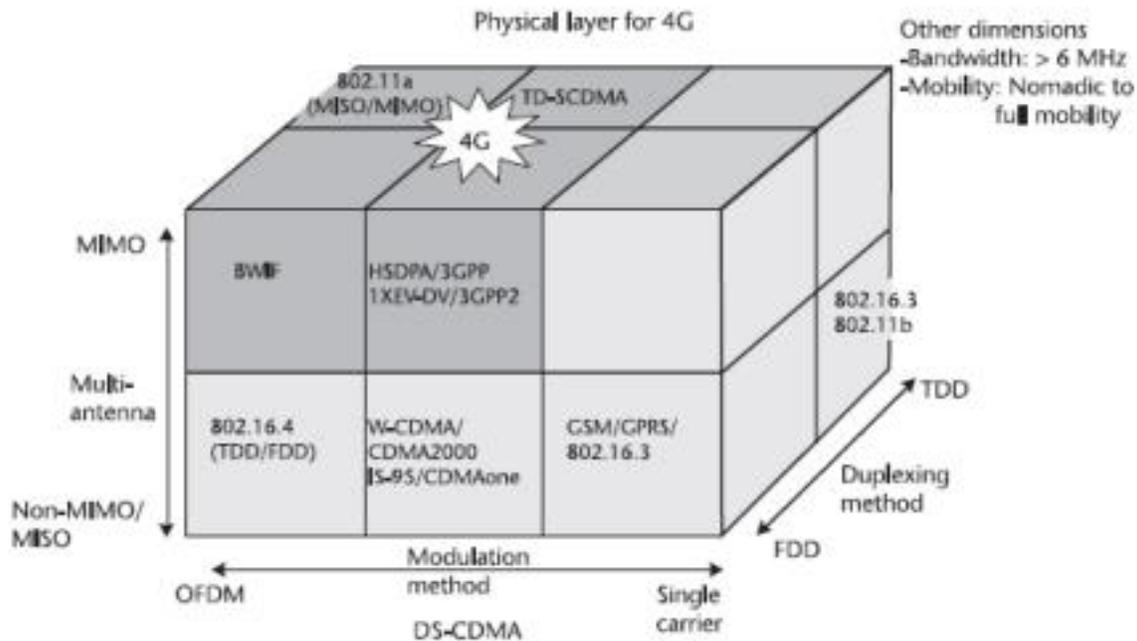


Figure 1.2 Possible candidates for 4G systems.

Many companies associated with wireless technology see the market potential for new applications, products, and services. BWIF members have committed to the VOFDM specification for optimized and open broadband fixed wireless access. Founding companies and promoting members do not collect royalties for the intellectual property they contribute to support VOFDM technology. In addition, BWIF is organized as a program of the IEEE Industry Standards and Technology Organization (IEEE-ISTO), which acts as the managing body for this forum.

TD-SCDMA:

Time-division synchronous CDMA (TD-SCDMA) is the Chinese Contribution to the ITU's IMT-2000 specification for 3G wireless mobile services. It endeavors to integrate with the existing GSM system. It is designed to manage both symmetric circuit-switched services, such as speech or video, as well as asymmetric packet-switched services, such as mobile Internet data flows. TD-SCDMA combines two leading technologies—an advanced TDMA system with an adaptive CDMA component—to overcome this challenge.

HSDPA:

High-speed downlink packet access (HSDPA) is a packet-based data service in WCDMA downlink with data transmission up to 8–10 mbit/s [and 20 mbit/s for multiple-input multiple-output (MIMO) systems] over a 5-MHz bandwidth in WCDMA downlink. HSDPA implementations include adaptive modulation and coding (AMC), MIMO, hybrid automatic request (HARQ), fast cell search, and advanced receiver design. In third generation partnership project (3GPP) standards, Release 4 specifications provide efficient IP support, enabling provision of services through an all-IP core network, and Release 5 specifications focus on HSDPA to provide data rates up to

approximately 10 mbit/s to support packet-based multimedia services.

MIMO systems are the work item in Release 6 specifications, which will support even higher data transmission rates up to 20 mbit/s. HSDPA is evolved from and backward compatible with Release 99 WCDMA systems.

1.3 Multiple-Input Multiple-Output Systems

This technique is mainly based on the theoretical work developed by Teletar and Foschini. The core of this idea is to use multiple antennas both for transmission and reception. This increases the capacity of the wireless channel. Capacity is expressed as the maximum achievable data rate for an arbitrarily low probability of error. Hence, the thrust has been toward the development of codes and schemes that would enable systems to approach their Shannon capacity limit.

This technology received a fillip when Tarokh introduced their space time trellis coding techniques and Alamouti introduced his space time block coding techniques to improve link-level performance based on diversity. It received another boost when Bell Laboratories introduced its Bell Laboratories Layered Space-Time (BLAST) coding technique, demonstrating spectral efficiencies as high as 42 bit/s/Hz. This represents a tremendous boost in spectral efficiency compared with the current 2–3 bit/s/Hz achieved in cellular mobile and wirelessLAN systems. There is, therefore, a need for communication engineers to understand this remarkable technology. This book has been expressly written to fulfill such a need.

We will discuss a MIMO system pioneered by AT&T Labs-Research in Middletown, New Jersey. It conducted field tests to characterize the mobile MIMO radio channel. The company measured the capacity of a system with four antennas on a laptop computer and four antennas on a rooftop base station. The field tests showed that close to the theoretical fourfold increase in capacity over a single antenna system can be supported in a 30-KHz channel with dual polarized spatially separated base station and mobile terminal antennas. Figure 1.3 shows the arrangement. Note the mounting of the four antennas on the laptop computer and the rooftop antennas.

The base station rooftop antenna array used dual-polarized antennas separated by 11.3 feet, which is approximately 20 wavelengths apart and a multibeam antenna. The laptop-mounted antennas included a vertically polarized array and a dual-polarized array with elements spaced half a wavelength apart. Different signals were transmitted out of each antenna simultaneously in the same bandwidth and then separated at the receiver.



Figure 1.3 Upper: Transmitter with four antennas on a laptop and the 1,900-MHz coherent transmitters. Lower: The four receivers with real-time baseband processing and rooftop antennas.

With four antennas at the transmitter and receiver, this has the potential to provide four times the data rate of a single antenna system without an increase in transmit power or bandwidth, provided the multipath environment is rich enough. This means that high capacities are theoretically possible unless there is a direct line of sight between transmitter and receiver.

1.4 MIMO systems – Array Gain, Diversity Gain, Data Pipes, Spatial MUX,

Multiantenna Systems

Figure 1.4 illustrates different antenna configurations used in defining space-time systems. Single-input single-output (SISO) is the well-known wireless configuration, single-input multiple-output (SIMO) uses a single transmitting antenna and multiple (MR) receive antennas, multiple-input single-output (MISO) has multiple (MT) transmitting antennas and one receive antenna, MIMO has multiple (MT) transmitting antennas and multiple (MR) receive antennas and, finally, MIMO-multiuser (MIMO-MU), which refers to a configuration that comprises a base station with multiple transmit/receive antennas interacting with multiple users, each with one or more antennas. We now examine the meaning of certain terms.

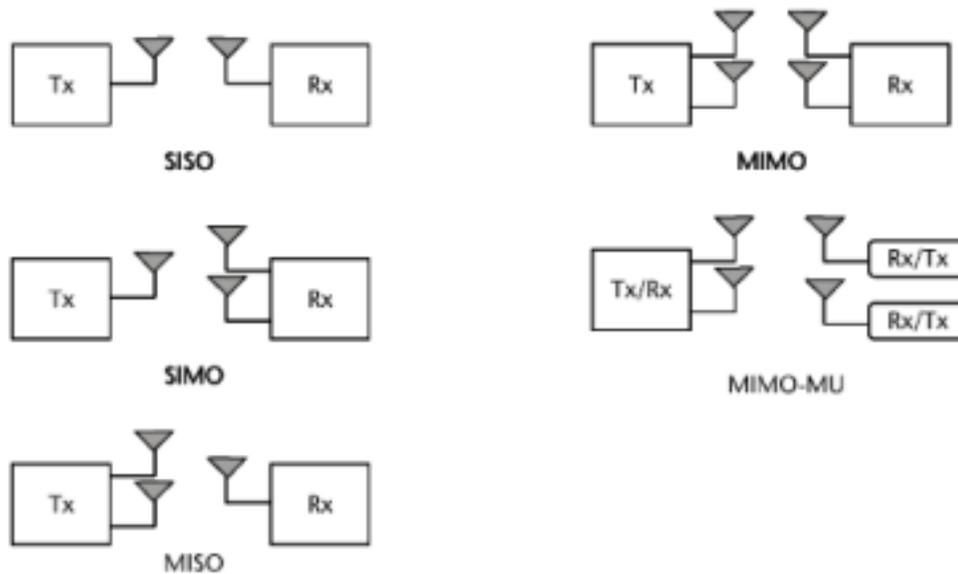


Figure 1.4 Different antenna configurations in space-time systems.

Array Gain

Array gain is the average increase in the signal-to-noise ratio (SNR) at the receiver that arises from the coherent combining effect of multiple antennas at the receiver or transmitter or both. If the channel is known to the multiple antenna transmitter, the transmitter will weight the transmission with weights, depending on the channel coefficients, so that there is coherent combining at the single antenna receiver (MISO case). The array gain in this case is called transmitter array gain. Alternately, if we have only one antenna at the transmitter and no knowledge of the channel and a multiple antenna receiver, which has perfect knowledge of the channel, then the receiver can suitably weight the incoming signals so that they coherently add up at the output (combining), thereby enhancing the signal.

This is the SIMO case. This is called receiver array gain. Basically, multiple antenna systems require perfect channel knowledge either at the transmitter or receiver or both to achieve this array gain.

Diversity Gain

Multipath fading is a significant problem in communications. In a fading channel, signals experience fades (i.e., they fluctuate in their strength). When the signal power drops significantly, the channel is said to be in a fade. This gives rise to high bit error rates (BER). We resort to diversity to combat fading. This involves providing replicas of the transmitted signal over time, frequency, or space. There are three types of diversity schemes in wireless communications.

Temporal diversity:

In this case replicas of the transmitted signal are provided across time by a combination of channel coding and time interleaving

strategies. The key requirement here for this form of diversity to be effective is that the channel must provide sufficient variations in time. It is applicable in cases where the coherence time of the channel is small compared with the desired interleaving symbol duration. In such an event, we are assured that the interleaved symbol is independent of the previous symbol. This makes it a completely new replica of the original symbol.

Frequency diversity:

This type of diversity provides replicas of the original signal in the frequency domain. This is applicable in cases where the coherence bandwidth of the channel is small compared with the bandwidth of the signal. This assures us that different part of the relevant spectrum will suffer independent fades.

Spatial diversity:

This is also called antenna diversity and is an effective method for combating multipath fading. In this case, replicas of the same transmitted signal are provided across different antennas of the receiver. This is applicable in cases where the antenna spacing is larger than the coherent distance to ensure independent fades across different antennas.

The traditional types of diversity schemes are selection diversity, maximal ratio diversity, and equal gain diversity. Space-time codes exploit diversity across space and time. Basically the effectiveness of any diversity scheme lies in the fact that at the receiver we must provide independent samples of the basic signal that was transmitted. In such an event we are assured that the probability of two or more relevant parts of the signal undergoing deep fades will be very small.

The constraints on coherence time, coherence bandwidth, and coherence distance ensure this. The diversity scheme must then optimally combine the received diversified waveforms so as to maximize the resulting signal quality. We can also categorize diversity under the subheading of spatial diversity, based on whether diversity is applied to the transmitter or to the receiver.

Receive diversity:

Maximum ratio combining is a frequently applied diversity scheme in receivers to improve signal quality. In cell phones it becomes costly and cumbersome to deploy. This is one of the main reasons transmit diversity became popular, since transmit diversity is easier to implement at the base station.

Transmit diversity:

In this case we introduce controlled redundancies at the transmitter, which can be then exploited by appropriate signal processing techniques at the receiver. Generally this technique requires

complete channel information at the transmitter to make this possible. It became possible to implement transmit diversity without knowledge of the channel. This was one of the fundamental reasons why the MIMO industry began to rise. Space-time codes for MIMO exploit both transmit as well as receive diversity schemes, yielding a high quality of reception.

Therefore, in MIMO we talk a lot about receive antenna diversity or transmit antenna diversity. In receive antenna diversity, the receiver that has multiple antennas receives multiple replicas of the same transmitted signal, assuming that the transmission came from the same source. This holds true for SIMO channels. If the signal path between each antenna pair fades independently, then when one path is in a fade, it is extremely unlikely that all the other paths are also in deep fade. Therefore, the loss of signal power due to fade in one path is countered by the same signal but received through a different path (route).

This is like a line of soldiers. When one soldier falls in battle, another is ready to take his place. Hence, extending this analogy further, the more the soldiers, the stronger the line. The same is the argument in diversity. The more the diversity, the easier we can combat fades in a channel. Diversity is characterized by the number of independent fading branches, or paths (routes). These paths are also known as diversity order and are equal to the number of receive antennas in SIMO channels.

Logically, the higher the diversity order (independent fading paths, or receive antennas), the better we combat fading. If the number of receive antennas tends to infinity, the diversity order tends to infinity and the channel tends to additive white Gaussian noise (AWGN).

In the category of spatial diversity there are two more types of diversity that we need to consider. These are:

Polarization diversity:

In this type of diversity horizontal and vertical polarization signals are transmitted by two different polarized antennas and received correspondingly by two different polarized antennas at the receiver.

Different polarizations ensure that there is no correlation between the data streams, without having to worry about coherent distance of separation between the antennas.

Angle diversity:

This applies at carrier frequencies in excess of 10 GHz. At such frequencies, the transmitted signals are highly scattered in space. In such an event the receiver can have two highly directional antennas

facing in totally different directions. This enables the receiver to collect two samples of the same signal, which are totally independent of each other.

Data Pipes

The term data pipe is derived from fluid mechanics. Pipes are used to transfer water to a tank/reservoir. The more the number of pipes, the greater the quantum of flow of water into a tank/reservoir. This is similar to data pipes, but the analogy of communications with fluid mechanics ends there. We consider a case of two data pipes between the transmitter and receiver. In this situation there are two cases; either the data in the data pipes are identical to each other or they are independent samples, completely different from each other.

In the former case, effectively the data going through is as if it is going through one data pipe, with the other pipe merely being a replica of the first one. This is a case of full correlation and because of this correlation, we do not get any throughput (bits per second)

advantage. However, we do get a diversity advantage of two. The latter case deals with a situation where there is absolutely no correlation between the data carried by the two pipes. The data streams are independent. Hence, there is no diversity, but the throughput (output in bit/s) is definitely higher than in the first case.

Therefore, the more the data pipes, the higher the throughput provided the signals in the data pipes are not replicas of each other or correlated. In such an event the same signal is going through both pipes, so no new information is getting transferred.

Remember that transmit diversity comes at the cost of throughput and vice versa. If we wish to eat the cake and still have it, then one way out is to sacrifice transmit diversity at the cost of throughput and incorporate diversity in the receiver (receive diversity). This way we at least have receive diversity, rather than no diversity at all in the system. This is what is done in spatial multiplexing.

Spatial Multiplexing

Spatial multiplexing offers a linear (in the number of transmit receive antenna pairs or $\min(M_R, M_T)$) increase in the transmission rate (or capacity) for the same bandwidth and with no additional power expenditure. It is only possible in MIMO channels. Consider the case of two transmit and two receive antennas. This can be extended to more general MIMO channels. The bit stream is split into two half-rate bit streams, modulated and transmitted simultaneously from both the antennas.

The receiver, having complete knowledge of the channel, recovers these individual bit streams and combines them so as to recover the original bit stream. Since the receiver has knowledge of the channel it provides receive diversity, but the system has no transmit diversity since the bit

streams are completely different from each other in that they carry totally different data. Thus spatial multiplexing increases the transmission rates proportionally with the number of transmit receive antenna pairs.

This concept can be extended to MIMO-MU. In such a case, two users transmit their respective information simultaneously to the base station equipped with two antennas. The base station can separate the two signals and can likewise transmit two signals with spatial filtering so that each user can decode his or her own signal correctly. This allows capacity to increase proportionally to the number of antennas at the base station and the number of users.

1.5 MIMO System Model

We consider a MIMO system with a transmit array of M_T antennas and a receive array of M_R antennas. The block diagram of such a system is shown in Figure 1.5. The transmitted matrix is a $M_T \times 1$ column matrix \mathbf{s} where s_i is the i th component, transmitted from antenna i . We consider the channel to be a Gaussian channel such that the elements of \mathbf{s} are considered to be independent identically

Distributed Gaussian variables. If the channel is unknown at the transmitter, we assume that the signals transmitted from each antenna have equal powers of E_s / M_T . The covariance matrix for this transmitted signal is given by

$$R_{ss} = E_s * I_{M_T} / M_T$$

where E_s is the power across the transmitter irrespective of the number of antennas M_T and I_{M_T} is an $M_T \times M_T$ identity matrix. The transmitted signal bandwidth is so narrow that its frequency response can be considered flat (i.e., the channel is memoryless). The channel matrix \mathbf{H} is a $M_R \times M_T$ complex matrix. The component $h_{i,j}$ of the matrix is the fading coefficient from the j th transmit antenna to the i th receive antenna. We assume that the received power for each of the receive antennas is equal to the total transmitted power E_s . This implies we ignore signal attenuation, antenna gains, and so on.

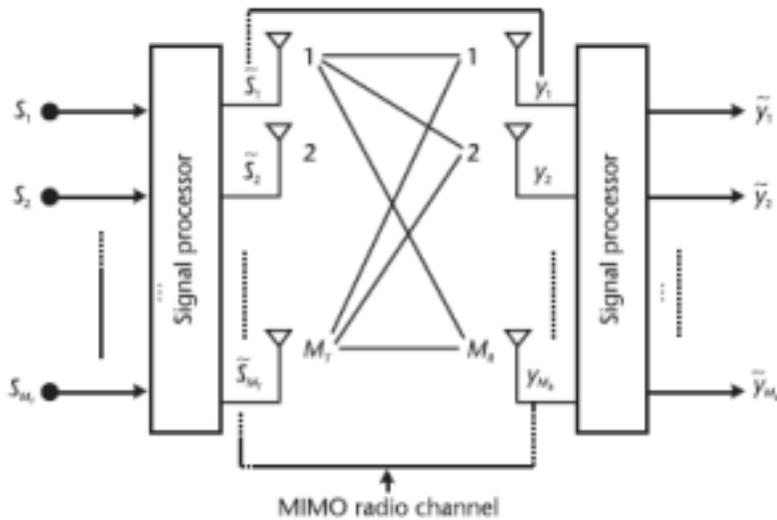


Figure 1.5

Block diagram of a MIMO system.

Thus we obtain the normalization constraint for the elements of \mathbf{H} , for a deterministic channel as

If the channel elements are not deterministic but random, the

$$\sum_{j=1}^{M_T} |b_{i,j}|^2 = M_T, i = 1, 2, \dots, M_R$$

normalization will apply to the expected value. We assume that the channel matrix is known at the receiver but unknown at the transmitter. The channel matrix can be estimated at the receiver by transmitting a training sequence. If we require the transmitter to know this channel, then we need to communicate this information to the transmitter via a feedback channel. The elements of \mathbf{H} can be deterministic or random.

The noise at the receiver is another column matrix of size $M_R \times 1$, denoted by \mathbf{n} . The components of \mathbf{n} are zero mean circularly symmetrical complex Gaussian (ZMCSCG) variables. The covariance matrix of the receiver noise is

If there is no correlation between components of the covariance matrix is obtained as

$$\mathbf{R}_{nn} = E\{\mathbf{nn}^H\} \quad \mathbf{n},$$

$$\mathbf{R}_{nn} = N_0 \mathbf{I}_{M_R}$$

Each of the M_R receive branches has identical noise power of N_0 .

The receiver operates on the maximum likelihood detection principle over M_R receive antennas. The received signals constitute a $M_R \times 1$ column matrix denoted by \mathbf{r} , where each complex component refers to a receive antenna. Since we assumed that the total received power per antenna is

equal to the total transmitted power, the SNR can be written as

Therefore, the received vector can be expressed as $\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{n}$ The

received signal

$$\gamma = \frac{E_s}{N_0}$$

covariance

matrix defined as $E\{\mathbf{r}\mathbf{r}^H\}$, is given by

$$\mathbf{R}_{rr} = \mathbf{H}\mathbf{R}_{ss}\mathbf{H}^H$$

while the total signal power can be expressed as $\text{tr}(\mathbf{R}_{rr})$.

1.6 MIMO System Capacity

The system capacity is defined as the maximum possible transmission rate such that the probability of error is arbitrarily small. We assume that the channel knowledge is unavailable at the transmitter and known only at the receiver.

The capacity of MIMO channel is defined as

Where $f(\mathbf{s})$ is the probability distribution of the vector \mathbf{s} and

$$C = \max_{f(\mathbf{s})} I(\mathbf{s}; \mathbf{y})$$

$I(\mathbf{s}; \mathbf{y})$ is the mutual information between vectors \mathbf{s} and \mathbf{y} . We note that

$$I(\mathbf{s}; \mathbf{y}) = H(\mathbf{y}) - H(\mathbf{y} | \mathbf{s})$$

where $H(\mathbf{y})$ is the differential entropy of the vector \mathbf{y} , while $H(\mathbf{y} | \mathbf{s})$ is the conditional differential entropy of the vector \mathbf{y} , given knowledge of the vector \mathbf{s} . Since the vectors \mathbf{s} and \mathbf{n} are independent, $H(\mathbf{y} | \mathbf{s}) = H(\mathbf{n})$.

$$I(\mathbf{s}; \mathbf{y}) = H(\mathbf{y}) - H(\mathbf{n})$$

If we maximize the mutual information $I(\mathbf{s}; \mathbf{y})$ reduces to maximizing $H(\mathbf{y})$. The covariance matrix of \mathbf{y} ,

$\mathbf{R}_{yy} = E\{\mathbf{y}\mathbf{y}^H\}$, satisfies

$$\mathbf{R}_{yy} = \frac{E_s}{M_T} \mathbf{H}\mathbf{R}_{ss}\mathbf{H}^H + N_0 \mathbf{I}_{M_R}$$

where $\mathbf{R}_{ss} = \varepsilon \{ss^H\}$, is the covariance matrix of \mathbf{s} .

Among all vectors \mathbf{y} with a given covariance matrix \mathbf{R}_{yy} , the differential entropy $H(\mathbf{y})$ is maximized when \mathbf{y} is ZMCSCG. This implies that \mathbf{s} must also be ZMCSCG vector, the distribution of which is completely characterized by \mathbf{R}_{ss} . The differential entropies of the vectors \mathbf{y} and \mathbf{n} are given by

$$H(\mathbf{y}) = \log_2 (\det(\pi e \mathbf{R}_{yy})) \text{ bps/Hz}$$

$$H(\mathbf{n}) = \log_2 (\det(\pi e \sigma^2 \mathbf{I}_{M_R})) \text{ bps/Hz}$$

Therefore, $I(\mathbf{s}; \mathbf{y})$

$$I(\mathbf{s}; \mathbf{y}) = \log_2 \det \left(\mathbf{I}_{M_R} + \frac{E_s}{M_T N_0} \mathbf{H} \mathbf{R}_{ss} \mathbf{H}^H \right) \text{ bps/Hz}$$

The capacity of the MIMO channel is given by

$$C = \max_{\text{Tr}(\mathbf{R}_{ss})=M_T} \log_2 \det \left(\mathbf{I}_{M_R} + \frac{E_s}{M_T N_0} \mathbf{H} \mathbf{R}_{ss} \mathbf{H}^H \right) \text{ bps/Hz}$$

The capacity C is also called error-free spectral efficiency or data rate per unit bandwidth that can be sustained reliably over the MIMO link. Thus if our bandwidth is W Hz, the maximum achievable data rate over this bandwidth using MIMO techniques is WC bit/s.

1.7 Channel Unknown to the Transmitter

If the channel is unknown to the transmitter, then the vector \mathbf{s} is statistically independent (i.e., $\mathbf{R}_{ss} = \mathbf{I}_{M_T}$). This implies that the signals are independent and the power is equally divided among the transmit antennas. The capacity in such a case is,

$$C = \log_2 \det \left(\mathbf{I}_{M_R} + \frac{E_s}{M_T N_0} \mathbf{H} \mathbf{H}^H \right)$$

The reader is cautioned that this is not Shannon capacity since it is possible to outperform $\mathbf{R}_{ss} = \mathbf{I}_{M_T}$, if one has the channel knowledge. Nevertheless we shall refer capacity. Now $\mathbf{H} \mathbf{H}^H$ is an $M_R * M_R$ positive semidefinite Hermitian matrix. The eigen decomposition of such a matrix is given by $\mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H$, where \mathbf{Q} is an $M_R * M_R$ matrix satisfying $\mathbf{Q}^H \mathbf{Q} = \mathbf{Q} \mathbf{Q}^H = \mathbf{I}_{M_R}$ and

$\mathbf{\Lambda} = \text{diag} \{ \lambda_1 \lambda_2 \dots \lambda_{M_R} \}$ with $\lambda_i \geq 0$. We assume that the eigen values are ordered so that $\lambda_i \geq \lambda_{i+1}$. Then

where r is the rank of the channel and λ_i ($i = 1, 2, \dots, r$) are the positive eigen values of HH^H . Equation expresses the capacity of the MIMO channel as a sum of the capacities of r SISO channels, each having a power gain of λ_i ($i = 1, 2, \dots, r$) and transmit power E_s / M_T .

This means that the technique of multiple antennas at the transmitter and receiver opens up multiple scalar spatial data pipes between the transmitter and receiver. Furthermore, equal transmit energy is allocated to each spatial data pipe. This is for the case when the channel is unknown at the transmitter.

Then if the channel matrix is of full rank such that $M_T = M_R = M$, the capacity C is maximized when $\lambda_i = \lambda_j = \beta / M$ ($i, j = 1, 2, \dots, M$) (Remember, the channel is unknown, so equal power distribution). To achieve this, $HH^H = H^H H = (\beta / M) I_M$, (i.e., the channel matrix H should be orthogonal). This gives

The capacity of an orthogonal MIMO channel is therefore M times the scalar channel capacity.

1.8 Channel Known to the Transmitter

It is possible by various means, which will be discussed in Chapter 4, to learn the channel state information (CSI) at the transmitter. In such an event the capacity can be increased by resorting to the so called “water filling principle”, by assigning various levels of transmitted power to various transmitting antennas. This power is assigned on the basis that the better the channel gets, the more power it gets and vice versa. This is an optimal energy allocation algorithm.

Water-Pouring Principle

Consider a MIMO channel where the channel parameters are known at the transmitter. The “water-pouring principle” or “water filling principle” can be derived by maximizing the MIMO channel capacity under the rule that more power is allocated to the channel that is in good condition and less or none at all to the bad channels.

Consider a ZMCSCG signal vector \hat{s} of dimension $r \times 1$ where r is the rank of the channel H to be transmitted. We note from Figure 1.6 that the vector is multiplied by a matrix V prior to transmission (based on the fact that $H = U\Sigma V^H$ through singular value decomposition). At the receiver, the received signal vector y is multiplied by the matrix U^H .

The input-output relationship for this operation is given by $\hat{y} = U^H y$. Where \hat{y} is the transformed received signal vector of size $r \times 1$ and

\tilde{n} is the ZMCSCG transformed noise vector of size $r \times 1$ with the covariance matrix $\epsilon \{\tilde{n}\tilde{n}^H\} = N_0 I_r$.

The vector \hat{s} satisfies $\epsilon \{\hat{s}\hat{s}^H\} = M_T$ to constrain the total transmit energy. Equation shows us that with channel knowledge at the transmitter, H can be explicitly decomposed into r parallel SISO channels satisfying

The capacity of the MIMO channel is the sum of the individual parallel SISO channel capacities and is given by

To maximize mutual information, the transmitter can access the individual subchannels and allocate variable power levels to them. Hence, the mutual information maximization problem becomes,

Figure 1.6 Decomposition of H when the channel is known to the transmitter and receiver.

We determine this optimal energy allocation iteratively through the “water pouring algorithm”. We set the iteration count p to 1 and calculate the constant μ

If the power allotted to the channel with the lowest gain is negative (i.e., $\lambda_{r-p+1} < 0$), we discard this channel by setting $\gamma_{r-p+1}^{\text{opt}} = 0$ and rerun the algorithm with the iteration count p incremented by 2. The optimal power allocation strategy, therefore, allocates power to those spatial subchannels that are non-negative. Figure illustrates the water-pouring algorithm. Obviously, since this algorithm only concentrates on good-quality channels and rejects the bad ones during each channel realization, it is to be expected that this method yields a capacity that is equal or better than the situation when the channel is unknown to the transmitter.

1.9 Capacity Of Deterministic Channels

In a SIMO channel, $M_T = 1$ and there are M_R receive antennas. In such a case the channel matrix is a column matrix

$$H = (h_1 \ h_2 \ \dots \ h_{M_R})^T$$

where $(\cdot)^T$ denotes matrix transpose. Since $M_R > M_T$, is modified as

achieves a diversity gain of M_R relative to the SISO case. For $M_R = 4$ and $\text{SNR} = 10$ dB, the SIMO capacity is 5.258 bit/s/Hz. The addition of receive antennas yields a logarithmic increase in capacity in SIMO channels. Knowledge of the channel at the transmitter in this case provides no additional benefit.

MISO Channel Capacity

In MISO channels, $M_R = 1$ and there are M_T transmit antennas. In this case, since $M_T > M_R$. The channel is represented by the row matrix

We note that the equation is the same as for a SISO case (i.e., the capacity did not increase with the number of antennas). This is for the case when the channel is unknown at the transmitter. The reason for this result is that there is no array gain at the transmitter because the transmitter has no knowledge of the channel

parameters. Array gain is the average increase in the SNR at the receiver that arises from the coherent combining effect of multiple antennas at the receiver or transmitter or both. If the channel is known to the transmitter, the transmitter will weight the transmission with weights depending on the channel coefficients, so that there is coherent combining at the receiver (MISO case).

If we take the case when the channel is known at the transmitter, we apply the equation. Since the channel matrix has rank 1, there is only one term in the sum in and only one nonzero eigen value given by

For $M_T = 4$ and $\text{SNR} = 10$ dB, the MISO capacity is 5.258 bit/s/Hz. This is with channel knowledge at the transmitter. In both cases of SIMO and MISO there is only one spatial data pipe (i.e., the rank of the channel matrix is one). Basically, the channel matrix is a $M_R \times M_T$ matrix. In a MISO case, $M_R = 1$ and in a SIMO case, $M_T = 1$. In either case, the channel matrix has only one eigen value and its rank is 1.

Physically, this means that there is only one route from transmitter to receiver for the signals to pass through. Hence, we have one data pipe. If we had $M_T = M_R = 2$, then we would have a MIMO case with a channel matrix of rank 2 and having two eigen values, hence, two routes from transmitter to receiver (i.e., we have two data pipes and so on).

1.10 Random Channels And Frequency Selective Channels

We have until now discussed MIMO capacity when the channel is a deterministic channel. We now consider the case when \mathbf{H} is chosen randomly according to a Rayleigh distribution in a quasi-static channel. This is a real-life situation encountered, for example, in wireless LANs with high data rates and low fade rates. We assume that the receiver has perfect knowledge of the channel and the transmitter has no knowledge of the channel. Since the channel is random, the information rate associated with it is also random. The cumulative distribution function (CDF) of the information rate of a flat fading MIMO channel is shown in Figure for a 2×2 system. The SNR is 10 dB and the channel is unknown to the transmitter.

Figure

1.7 CDF of information rate for channel matrix with a 2×2 system and SNR = 10 dB.

The ergodic capacity of a MIMO channel is the ensemble average of the information rate over the distribution of the elements of the channel matrix \mathbf{H} [7]. It is the capacity of the channel when every channel matrix \mathbf{H} is an independent realization [i.e., it has no relationship to the previous matrix but is typically representative of its class (ergodic)]. This implies that it is a result of infinitely long measurements.

Since the process model is ergodic, this implies that the coding is performed over an infinitely long interval. Hence, it is the Shannon capacity of the channel. Based on equation the ergodic capacity is expressed as

where $\rho = E_s/N_0$. The expectation operator applies in this case because the channel is random. Since \mathbf{H} is random, the information rate associated with it is also random. The CDF of the information rate is depicted in Figure 1.7. The ergodic capacity is the median of the CDF curve. In this case it is 7.0467 bit/s/Hz. Figure 1.8 shows the ergodic capacity over different system configurations as a function of ρ . We note that ergodic capacity increases with increasing ρ and with increasing M_T and M_R .

Ergodic capacity when the channel is known to the transmitter is based

on the water-filling algorithm and is given from

Figure 1.8 Ergodic capacity for different antenna configurations with

= M.

$$M_T = M_R$$

Equation is the ensemble average of the capacity achieved when the Water-filling optimization is performed for each realization of \mathbf{H} . Figure shows the performance comparison of ergodic capacity of a MIMO channel with $M_T = M_R = 4$ when the channel is unknown to the transmitter and also when known and the channel is Rayleigh i.i.d. The ergodic capacity when the channel is known to the transmitter is always higher than when it is unknown. This advantage reduces at high SNRs. Another way of looking at this situation is to appreciate the fact that at high SNRs, all eigen channels perform equally well (i.e., there is no difference in quality between them). Hence, all the channels will perform to their capacities, making both cases nearly identical.

Capacity of Frequency Selective MIMO Channels

We now consider a real-life situation wherein the channel is not narrowband but frequency selective. Intuitively, subdividing the wideband channel into N narrowband ones, and then summing the capacities of these N frequency flat channels can achieve this. The bandwidth of each of these subchannels will be B/N Hz where B is the overall channel bandwidth. This is provided the coherent bandwidth of the channel permits this (i.e., it is more than or equal to B/N Hz), as

otherwise the subchannels will not be frequency flat.
We take the i th subchannel.

The input-output relationship is defined as

$$r_i = H_i s_i + n_i$$

where r_i is the $M_R \times 1$ received signal vector, s_i is the $M_T \times 1$ transmitted signal vector and n_i is the $M_R \times 1$ noise vector for the i th subchannel. Hence, for the overall wideband channel we deal with block matrixes as

of S , constrained so that $\text{Tr}(R_{ss}) = NMT$. This constrains the total average transmit power to E_s . Capacity of such a channel is given by

We now examine the two usual cases of when the channel is

unknown to the transmitter and when it is known to the transmitter.

Channel Unknown to the Transmitter

In this case, we should choose $R_{ss} = I_{MTN}$, which implies that the covariance matrix is of full rank (no correlation) and this in turn means that transmit power is allocated evenly across space (transmit antennas) and frequency (subchannels). This yields a deterministic capacity of

The outage capacity is similarly defined. However, this outage capacity will be much better (higher) than for the earlier examined cases of frequency flat channels (at low outage rates). This is due to the high amount of frequency diversity present in the frequency selective channel. This is manifest in Figure 1.9.

In this Figure, as the number of narrowband channels increases, with increasing frequency selectivity, the outage capacity also rises proportionately because of rising frequency diversity. Hence, the more the frequency selectivity, the higher the outage capacity. Note also the tendency of the curve to flatten with rising frequency

selectivity and rising N . This bears out the statement that as $N \rightarrow \infty$, the capacity tends to a fixed value. This means that asymptotically (in N), the outage capacity of a sample realization of the frequency selective MIMO channel equals its ergodic capacity (because $N \rightarrow \infty$).

Figure 1.9 Performance of frequency selectivity versus 10% outage capacity.

Channel Known to the Transmitter

The treatment regarding this case is similar as was done earlier for frequency flat channels. In this case, we need to distribute the energy or power across space (antennas) and frequency (subchannels) so as to maximize spectral efficiency. This is called space-frequency water-filling. Since water-filling is applicable only to purely orthogonal channels, it becomes necessary to achieve orthogonal channels by using OFDM techniques to convert a frequency select channel into a set of parallel frequency flat channels, which are orthogonal to each other.

