1 Introduction

1.1 MIMO wireless communication

The use of multiple antennas at the transmitter and receiver in wireless systems, popularly known as MIMO (multiple-input multiple-output) technology, has rapidly gained in popularity over the past decade due to its powerful performance-enhancing capabilities. Communication in wireless channels is impaired predominantly by multi-path fading. Multi-path is the arrival of the transmitted signal at an intended receiver through differing angles and/or differing time delays and/or differing frequency (i.e., Doppler) shifts due to the scattering of electromagnetic waves in the environment. Consequently, the received signal power fluctuates in space (due to angle spread) and/or frequency (due to delay spread) and/or time (due to Doppler spread) through the random superposition of the impinging multi-path components. This random fluctuation in signal level, known as fading, can severely affect the quality and reliability of wireless communication. Additionally, the constraints posed by limited power and scarce frequency bandwidth make the task of designing high data rate, high reliability wireless communication systems extremely challenging.

MIMO technology constitutes a breakthrough in wireless communication system design. The technology offers a number of benefits that help meet the challenges posed by both the impairments in the wireless channel as well as resource constraints. In addition to the time and frequency dimensions that are exploited in conventional single-antenna (single-input single-output) wireless systems, the leverages of MIMO are realized by exploiting the spatial dimension (provided by the multiple antennas at the transmitter and the receiver).

We indicate the kind of performance gains that are expected from the use of MIMO technology by plotting in Figure 1.1 the data rate versus the receive signal-to-noise ratio (SNR) in a 100 kHz channel for an $M \times M$ (i.e., M receive and M transmit antennas) fading link with $M = 1, 2, 4$. The channel response is assumed constant over the bandwidth of interest for this simple example. Assuming a target receive SNR of 25 decibels (dB), a conventional single-input single-output (i.e., $M = 1$) system can deliver a data rate of 0.7 Mbps (where Mbps denotes Mbits per second). With $M = 2$ and 4 we can realize data rates of 1.4 and 2.8 Mbps respectively. This increase in data rate is realized for no additional power or bandwidth expenditure compared to the single-input single-output system. In principle, the single-input single-output system can achieve the data rate of 2.8 Mbps with a receive SNR of 25 dB if the bandwidth is increased to 400 kHz, or

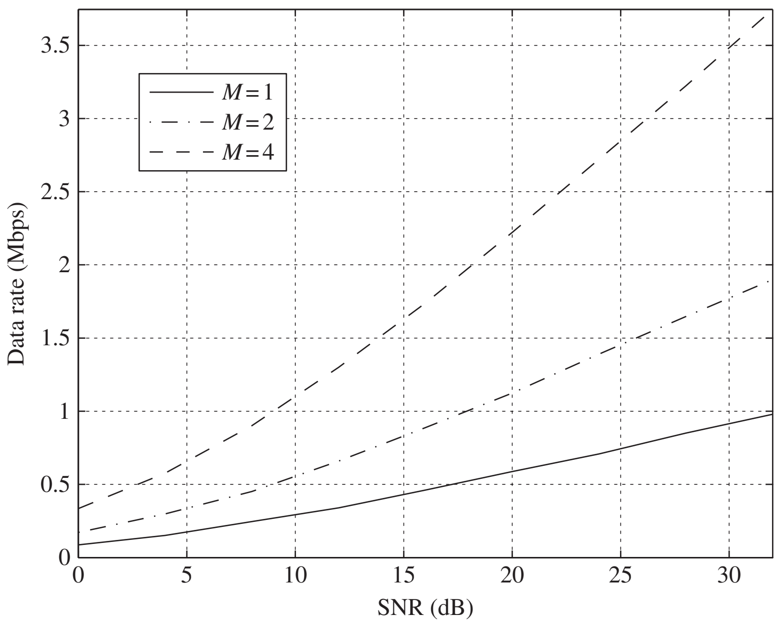

Fig. 1.1. Average data rate versus SNR for different antenna configurations. The channel bandwidth is 100 kHz.

alternatively, with a bandwidth of 100 kHz if the receive SNR is increased to 88 dB! The result presented in this example is based on optimal transceiver design. In practice, the modulation and impairments-constrained data rate delivered will be less but the general trend will still hold.

1.1.1 Benefits of MIMO technology

The benefits of MIMO technology that help achieve such significant performance gains are *array gain*, *spatial diversity gain*, *spatial multiplexing gain* and *interference reduction*. These gains are described in brief below.

Array gain

Array gain is the increase in receive SNR that results from a coherent combining effect of the wireless signals at a receiver. The coherent combining may be realized through spatial processing at the receive antenna array and/or spatial pre-processing at the transmit antenna array. Array gain improves resistance to noise, thereby improving the coverage and the range of a wireless network.

Spatial diversity gain

As mentioned earlier, the signal level at a receiver in a wireless system fluctuates or fades. Spatial diversity gain mitigates fading and is realized by providing the receiver with multiple (ideally independent) copies of the transmitted signal in space, frequency or time. With an increasing number of independent copies (the number of copies is often referred to as the diversity order), the probability that at least one of the copies is not experiencing a deep fade increases, thereby improving the quality and reliability of reception. A MIMO channel with M_T transmit antennas and M_R receive antennas potentially offers $M_T M_R$ independently fading links, and hence a spatial diversity order of $M_T M_R$.

Spatial multiplexing gain

MIMO systems offer a linear increase in data rate through spatial multiplexing [5, 9, 22, 35], i.e., transmitting multiple, independent data streams within the bandwidth of operation. Under suitable channel conditions, such as rich scattering in the environment, the receiver can separate the data streams. Furthermore, each data stream experiences at least the same channel quality that would be experienced by a single-input single-output system, effectively enhancing the capacity by a multiplicative factor equal to the number of streams. In general, the number of data streams that can be reliably supported by a MIMO channel equals the minimum of the number of transmit antennas and the number of receive antennas, i.e., $\min\{M_T, M_R\}$. The spatial multiplexing gain increases the capacity of a wireless network.

Interference reduction and avoidance

Interference in wireless networks results from multiple users sharing time and frequency resources. Interference may be mitigated in MIMO systems by exploiting the spatial dimension to increase the separation between users. For instance, in the presence of interference, array gain increases the tolerance to noise as well as the interference power, hence improving the signal-to-noise-plus-interference ratio (SINR). Additionally, the spatial dimension may be leveraged for the purposes of interference avoidance, i.e., directing signal energy towards the intended user and minimizing interference to other users. Interference reduction and avoidance improve the coverage and range of a wireless network.

In general, it may not be possible to exploit simultaneously all the benefits described above due to conflicting demands on the spatial degrees of freedom. However, using some combination of the benefits across a wireless network will result in improved capacity, coverage and reliability.

1.1.2 *Basic building blocks*

Figure 1.2 shows the basic building blocks that comprise a MIMO communication system. The information bits to be transmitted are encoded (using, for instance, a convolutional encoder) and interleaved. The interleaved codeword is mapped to data symbols (such as quadrature amplitude modulation or QAM symbols) by the symbol mapper. These data symbols are input to a space–time encoder that outputs one or more spatial data streams. The spatial data streams are mapped to the transmit antennas by the space–time precoding block. The signals launched from the transmit antennas propagate through the channel and arrive at the receive antenna array. The receiver collects the signals at the

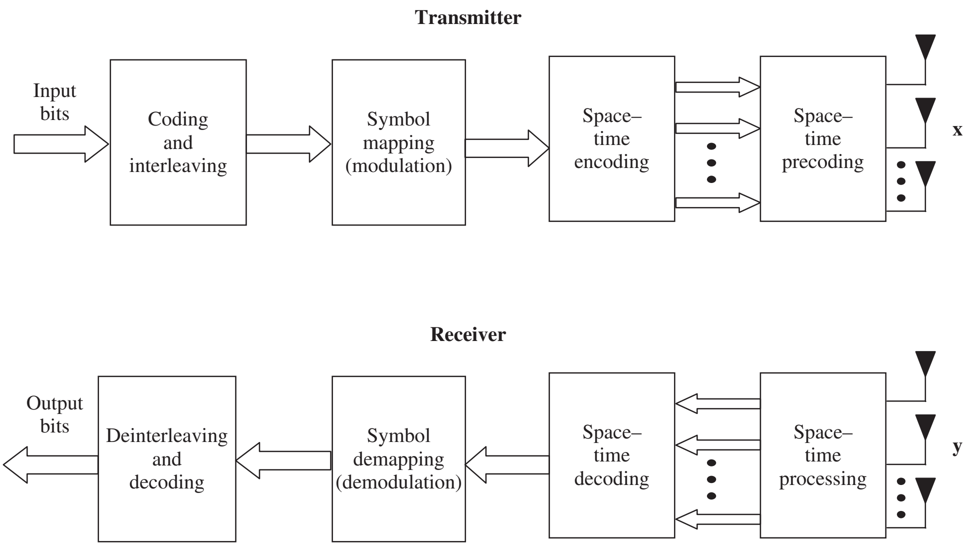

Fig. 1.2. Diagram of a complex equivalent baseband MIMO communication system. **x** and **y** stand for the transmitted and received signal vectors respectively.

output of each receive antenna element and reverses the transmitter operations in order to decode the data: receive space–time processing, followed by space–time decoding, symbol demapping, deinterleaving and decoding. Each of the building blocks offers the opportunity for significant design challenges and complexity–performance trade-offs. Furthermore, a number of variations can exist in the relative placement of the blocks, the functionality and the interactions between the blocks.

This book addresses key concepts and challenges in designing and understanding the performance limits of a MIMO communication system.

1.2 MIMO channel and signal model

In order to design efficient communication algorithms for MIMO systems and to understand the performance limits, it is important to understand the nature of the MIMO channel. For a system with M_T transmit antennas and M_R receive antennas, assuming frequency-flat fading[1] over the bandwidth of interest, the MIMO channel at a given time instant may be represented as an $M_R \times M_T$ matrix

$$\mathbf{H} = \begin{bmatrix} H_{1,1} & H_{1,2} & \cdots & H_{1,M_T} \\ H_{2,1} & H_{2,2} & \cdots & H_{2,M_T} \\ \vdots & \vdots & \ddots & \vdots \\ H_{M_R,1} & H_{M_R,2} & \cdots & H_{M_R,M_T} \end{bmatrix}, \tag{1.1}$$

[1] The delay spread in the channel is negligible compared to the inverse bandwidth.

where $H_{m,n}$ is the (single-input single-output) channel gain between the mth receive and nth transmit antenna pair. The nth column of $\mathbf{H}$ is often referred to as the spatial signature of the nth transmit antenna across the receive antenna array. The relative geometry of the M_T spatial signatures determines the distinguishability of the signals launched from the transmit antennas at a receiver. This is particularly important when independent data streams are launched from the transmit antennas, as is done in the case of spatial multiplexing.

As for the case of single-input single-output channels, the individual channel gains comprising the MIMO channel are commonly modeled as zero-mean circularly symmetric complex Gaussian random variables. Consequently, the amplitudes $|H_{m,n}|$ are Rayleigh-distributed random variables and the corresponding powers $|H_{m,n}|^2$ are exponentially distributed.

1.2.1 Classical independent, identically distributed (i.i.d.) Rayleigh fading channel model

The degree of correlation between the individual $M_T M_R$ channel gains comprising the MIMO channel is a complicated function of the scattering in the environment and antenna spacing at the transmitter and the receiver. Consider an extreme condition were all antenna elements at the transmitter are collocated and likewise at the receiver. In this case, all the elements of $\mathbf{H}$ will be fully correlated (in fact identical) and the spatial diversity order of the channel is one.

Decorrelation between the channel elements will increase with antenna spacing. However, antenna spacing alone is not sufficient to ensure decorrelation. Rich (i.e., omni-directional and isotropic) scattering in the environment in combination with adequate antenna spacing ensures decorrelation of the MIMO channel elements. With rich scattering, the typical antenna spacing required for decorrelation is approximately $\lambda/2$, where λ is the wavelength corresponding to the frequency of operation. Under ideal conditions, when the channel elements are perfectly decorrelated, we have[2] $H_{m,n}(m = 1, 2, \ldots, M_R, n = 1, 2, \ldots, M_T) \sim$ i.i.d. $\mathcal{CN}(0, 1)$. Summarizing, we get $\mathbf{H} = \mathbf{H}_w$, the classical i.i.d. frequency-flat Rayleigh fading MIMO channel. The spatial diversity order of $\mathbf{H}_w$ is $M_T M_R$.

1.2.2 Frequency-selective and time-selective fading

The channel model above assumes that the product of the bandwidth and the delay spread is very small. With increasing bandwidth and/or delay spread, this product is no longer negligible (0.1 is often considered the threshold for voice communication [11]), resulting in channel realizations that are frequency-dependent, i.e., $\mathbf{H}(f)$. While fading at a given frequency may be decorrelated in the spatial domain (resulting in $\mathbf{H}_w(f)$), correlation may exist across channel elements in the frequency domain. The correlation properties in the frequency domain are a function of the power delay profile. The coherence bandwidth B_c

[2] A complex-valued random variable $Z = X + jY$ is $\mathcal{CN}(0, 1)$ if X and Y are independent and normally distributed with zero mean and variance $\frac{1}{2}$.

is defined as the minimum separation in bandwidth required to achieve decorrelation. For two frequencies f_1 and f_2 with $|f_1 - f_2| > B_c$, we have $E[\mathrm{vec}(\mathbf{H}(f_1))\mathrm{vec}^H(\mathbf{H}(f_2))] = 0$. The coherence bandwidth is inversely proportional to the delay spread of the channel.

Furthermore, due to the motion of scatterers in the environment or of the transmitter or receiver, the channel realizations will vary with time. As with the case of frequency-selective fading, we can define a coherence time T_c, defined as the minimum separation in time required for decorrelation of the time-varying channel realizations. For two time instances t_1 and t_2 with $|t_1 - t_2| > T_c$, we have $E[\mathrm{vec}(\mathbf{H}(t_1))\mathrm{vec}^H(\mathbf{H}(t_2))] = 0$. The coherence time is inversely proportional to the Doppler spread of the channel.

1.2.3 *Real-world MIMO channels*

In practice, the behavior of $\mathbf{H}$ can significantly deviate from $\mathbf{H}_w$ due to a combination of inadequate antenna spacing and/or inadequate scattering leading to spatial fading correlation. Furthermore, the presence of a fixed (possibly line-of-sight or LOS) component in the channel will result in Ricean fading. Extensive measurements of real-world MIMO channels [3, 8, 15, 17, 30, 32] have been carried out by researchers across the world to develop accurate models [5, 10, 24]. Figure 1.3 shows the measured

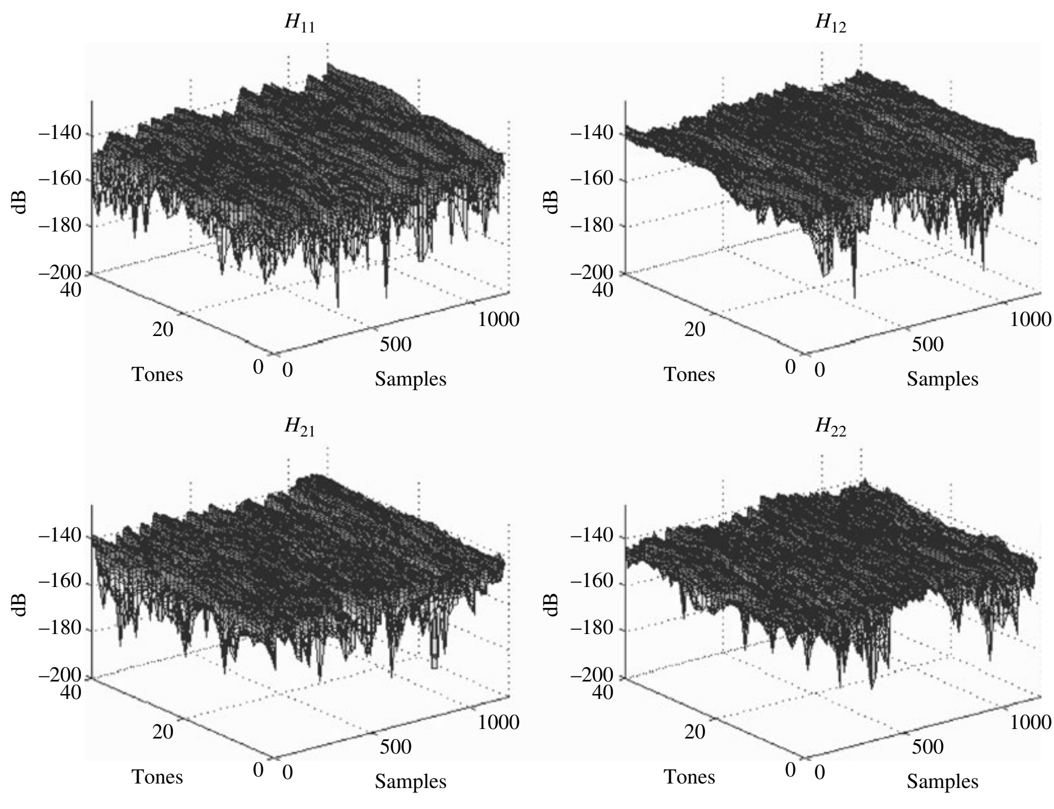

Fig. 1.3. Measured real-world MIMO channel. $H_{i,j}$ denotes the channel gain between the jth transmit antenna and ith receive antenna.

time–frequency response for an $M_T = M_R = 2$ MIMO channel in a fixed broadband wireless access system at 2.5 GHz. It is clear from the figure that real-world MIMO channels are triply selective, i.e., they exhibit fading across space, time and frequency.

In the presence of an LOS component between the transmitter and the receiver, the MIMO channel may be modeled as the sum of a fixed component and a fading component

$$\mathbf{H} = \sqrt{\frac{K}{1+K}}\overline{\mathbf{H}} + \sqrt{\frac{1}{1+K}}\mathbf{H}_w, \tag{1.2}$$

where $\sqrt{\frac{K}{1+K}}\overline{\mathbf{H}} = E[\mathbf{H}]$ is the LOS component of the channel and $\sqrt{\frac{1}{1+K}}\mathbf{H}_w$ is the fading component, assuming uncorrelated fading. $K \geq 0$ in (1.2) is the Ricean K-factor of the channel and is defined as the ratio of the power in the LOS component of the channel to the power in the fading component. When $K = 0$ we have pure Rayleigh fading. At the other extreme $K = \infty$ corresponds to a non-fading channel.

In general, real-world MIMO channels will exhibit some combination of Ricean fading and spatial fading correlation. Spatial correlation models will be discussed in Chapter 2. Furthermore, the use of polarized antennas will necessitate additional modifications to the channel model. These factors collectively will impact (probably adversely) the performance of a given MIMO signaling scheme. With appropriate knowledge of the MIMO channel at the transmitter, the signaling strategy can be appropriately adapted to meet performance requirements. The channel state information could be complete (i.e., the precise channel realization) or partial (i.e., knowledge of the spatial correlation, K-factor, etc.). Space–time precoding techniques to exploit channel knowledge at the transmitter are detailed in Chapter 3.

1.2.4 Discrete-time signal model

For a frequency-flat fading MIMO channel, the commonly used discrete-time input–output relation over a symbol period is given by

$$\mathbf{y} = \sqrt{\frac{E_x}{M_T}}\mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1.3}$$

where $\mathbf{y}$ is the $M_R \times 1$ received signal vector, $\mathbf{x}$ is the $M_T \times 1$ transmitted signal vector, $\mathbf{n}$ is additive temporally white complex Gaussian noise with $E[\mathbf{n}\mathbf{n}^H] = N_o\mathbf{I}_{M_R}$ and E_x is the total average energy available at the transmitter over a symbol period having removed losses due to propagation and shadowing. We constrain the total average transmitted power over a symbol period by assuming that the covariance matrix of $\mathbf{x}$, $\mathbf{R}_{\mathbf{xx}} = E[\mathbf{x}\mathbf{x}^H]$, satisfies $\mathrm{Tr}(\mathbf{R}_{\mathbf{xx}}) = M_T$. The ratio $\rho = E_x/N_o$ equals the SNR per receive antenna (simply referred to as SNR henceforth). Furthermore, it is commonly assumed that the channel is block fading [4], i.e., the channel remains constant over N consecutive symbol periods (determined by the coherence time) and then changes in an independent fashion to a new realization. Frequency-selective fading can be incorporated into the channel model by using a matrix tapped-delay line. The appropriate changes to the channel model will be described when required in subsequent chapters.

1.3 A fundamental trade-off

Two key performance metrics associated with any communication system are the transmission rate and the frame-error rate (FER). In the following, the transmission rate, R, is defined as the data rate transmitted per unit bandwidth. The FER, $P_e(\rho, R)$, is defined as the probability with which the transmitted frame (i.e., packet) is incorrectly decoded at the receiver, and is a function of the SNR and transmission rate.

Intuitively, for a fixed transmission rate an increase in SNR will result in reduced FER. Similarly, at a fixed target FER, an increase in SNR may be leveraged to increase the transmission rate. Hence, a fundamental trade-off exists in any communication system between the transmission rate and FER. In the context of MIMO systems, this trade-off is often referred to as the diversity–multiplexing trade-off [41] with diversity signifying the FER reduction and multiplexing signifying an increase in transmission rate. The diversity–multiplexing trade-off is central to MIMO communication theory and is described in brief in this section.

1.3.1 Outage capacity

The capacity of a communication channel is the maximum, asymptotic (in block length) error-free transmission rate that can be achieved. The capacity of a MIMO channel is a complicated function of the channel conditions and transmit/receive processing constraints [4, 9, 13, 19, 21, 35, 40]. A detailed discussion on the capacity of MIMO channels is provided in Chapter 2. The development below focuses on the outage capacity of MIMO channels, which gains operational significance when the fading channel holds constant over the entire duration of the transmitted frame.

The p percentage outage capacity at SNR ρ, $C_{out,p}(\rho)$, is defined as the transmission rate that can be supported by $(100-p)\%$ of the fading realizations of the channel [35]. Hence at SNR ρ, if a frame is transmitted with rate $C_{out,p}(\rho)$, the probability that the frame will be decoded correctly is $(100-p)\%$. Equivalently the FER associated with transmission rate $C_{out,p}(\rho)$ is $p\%$, i.e.,

$$P_e(\rho, C_{out,p}(\rho)) = p\%. \tag{1.4}$$

1.3.2 Multiplexing gain

The maximum multiplexing gain r_{max} that can be achieved over a MIMO channel is given by the asymptotic (in SNR) slope of the outage capacity (for fixed FER) plotted as a function of the SNR on a linear–log scale, i.e.,

$$r_{max} = \lim_{\rho\to\infty} \frac{C_{out,p}(\rho)}{\log_2 \rho}. \tag{1.5}$$

For the $\mathbf{H}_w$ MIMO channel with optimal transceiver design (i.e., Gaussian code books, asymptotically large frame length, maximum-likelihood detection, etc.) $r_{max} = \min\{M_R, M_T\}$ indicating that for a fixed FER, the transmission rate may be increased by $\min\{M_R, M_T\}$bps/Hz for every 3 dB increase in SNR.

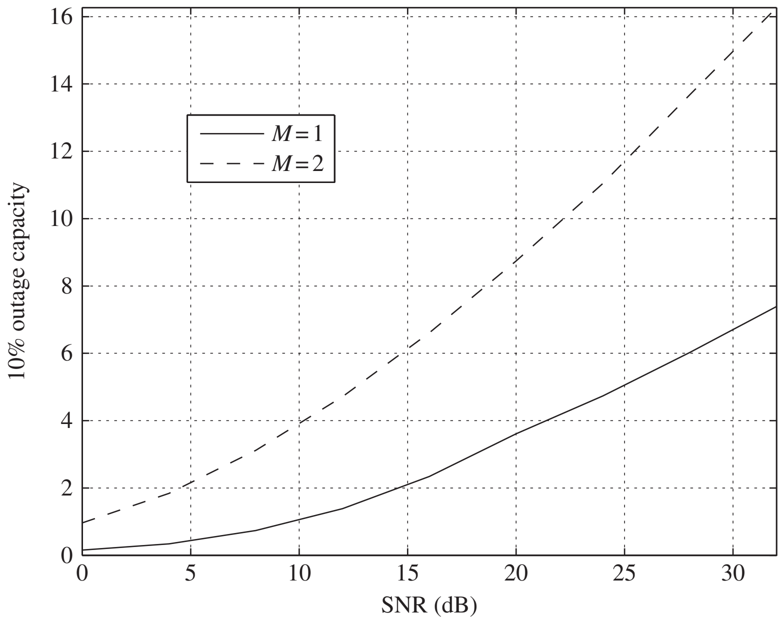

Fig. 1.4. 10% outage capacity for an $M \times M$ $\mathbf{H}_w$ channel plotted as a function of SNR.

Figure 1.4 shows a comparison of the 10% outage capacity of a 2×2 $\mathbf{H}_w$ MIMO channel to the 10% outage capacity of a single-input single-output channel, plotted as a function of SNR. At high SNR, the outage capacity for the MIMO channel grows with 2 bps/Hz/3 dB slope compared to 1 bps/Hz/3 dB slope for the single-input single-output channel.

1.3.3 Diversity gain

The maximum diversity gain d_{max} that can be achieved over a MIMO channel is given by the negative of the asymptotic (in SNR) slope of FER for a fixed transmission rate, plotted as a function of SNR on a log–log scale, i.e.,

$$d_{max} = -\lim_{\rho \to \infty} \frac{\log_2 P_e(\rho, R)}{\log_2 \rho}. \tag{1.6}$$

For the $\mathbf{H}_w$ MIMO channel with optimal transceiver design (i.e., Gaussian code books, asymptotically large frame length, maximum-likelihood detection, etc.) $d = M_R M_T$, indicating that for a fixed transmission rate, with every 3 dB increase in SNR, the FER decreases by a factor of $2^{-M_R M_T}$.

For $R = 2$ bps/Hz, Figure 1.5 compares the FER in a 2×2 $\mathbf{H}_w$ MIMO channel to the FER in a single-input single-output channel, plotted as a function of SNR. The fourth-order diversity provided by the MIMO channel is clearly reflected by the slope of the FER curve – at high SNR, the FER decreases by a factor of 2^{-4} in the MIMO channel (compared to 2^{-1} in the single-input single-output channel) for a 3 dB increase in SNR.

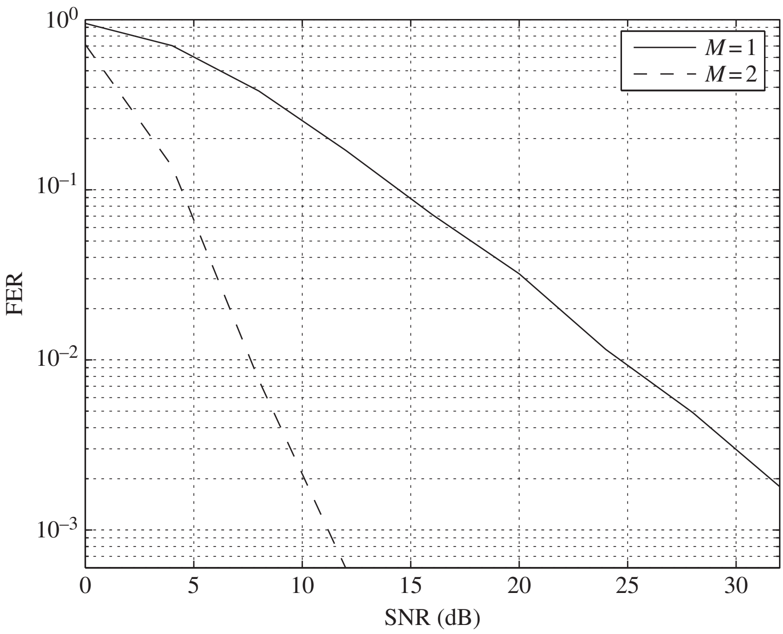

Fig. 1.5. FER at $R = 2$ bps/Hz for an $M \times M$ $\mathbf{H}_w$ channel plotted as a function of SNR.

1.3.4 Flexible trade-off

Individually, (1.5) and (1.6) represent the extremities of the diversity–multiplexing trade-off for MIMO channels. In (1.5) an increase in SNR is completely utilized to linearly (in $\min\{M_R, M_T\}$) increase the transmission rate, keeping the FER fixed. At the other extreme, in (1.6), an increase in SNR yields an exponential (the exponent is $-M_R M_T$) reduction in FER at a fixed transmission rate. Furthermore, (1.5) and (1.6) represent the gains of MIMO communication over single-input single-output systems as demonstrated in Figures 1.5 and 1.4.

In certain scenarios, we may desire to utilize an increase in SNR for some combination of transmission rate increase and FER reduction. It has been shown that a flexible trade-off between diversity and multiplexing can be achieved – the optimal trade-off curve for the $\mathbf{H}_w$ MIMO channel, $d(r)$, is piecewise linear (see Figure 1.6) connecting $(r, d(r))$, $r = 0, 1, \ldots, r_{max}$, where

$$d(r) = (M_R - r)(M_T - r). \tag{1.7}$$

The trade-off curve implies that if the transmission rate is increased by r bps/Hz over a 3 dB increase in SNR, the corresponding reduction in FER will equal $2^{-d(r)}$. Hence it is not possible to increase the transmission rate and decrease the FER simultaneously to the fullest extent (represented by r_{max} and d_{max} respectively) possible.