

Chapter 2

Digital Transmission System

Throughout the manuscript, we use as application the outer receiver, which is a part of the complete transmission chain as shown in Fig. 1.3. The model for the entire transceiver chain can be very complex. Thus it is essential to make further abstractions of this transmission chain model. As mentioned, the task of the inner transceiver is to create a good time discrete channel from the inner transmitter input to the inner receiver output. Assuming a perfect synchronization of the demodulator in time and frequency we can abstract the outer transmitter chain by a simple baseband transmission chain as shown in Fig. 2.1. The term baseband means that there is no frequency component present in the model. The final simple baseband model features a binary random source and a binary sink, which replaces the source compression/decompression of Fig. 1.3. This outer transceiver chain composed of channel encoder, interleaver, and modulator is called bit-interleaved coded modulation (BICM) and is often utilized in communication systems.

In the following we describe the nomenclature of the variables shown in Fig. 2.1. Most channel codes used in digital communications systems, such as the linear block codes that we explain in Chap. 3, operate on vectors. The vector input to the channel encoder is \mathbf{u} with the vector $\mathbf{u} = [u_0, u_1, \dots, u_{K-1}]$ of length K . \mathbf{u} is called *information word*, with $u_k \in \{0, 1\}$ being the *information bits*. The purpose of the channel encoder is to introduce, in a controlled manner, redundancy to the information word \mathbf{u} . The redundancy can be used at the receiver side to overcome the signal degradation encountered during the transmission through the discrete channel. The output of the channel encoder is the *codeword* $\mathbf{x} = [x_0, x_1, \dots, x_{N-1}]$ of length $N > K$. The interleaver permutes the sequence of the bits in the codeword, creating the interleaved codeword \mathbf{x}' . The interleaving is explained in more detail in Sect. 4.3.

This sequence is mapped by the digital modulator onto a sequence of symbols s . We can always map groups of Q bits on one modulation symbol. Thus, there exist 2^Q distinct symbols which we can transmit via the channel. One of the most common modulation schemes is the *Binary Phase Shift Keying* (BPSK) with $Q = 1$, where $x \in \{0, 1\}$ is mapped to $s \in \{-1, +1\}$. The final mapped transmit sequence s of length N_s is then disturbed by noise. For wireless transmission and the already

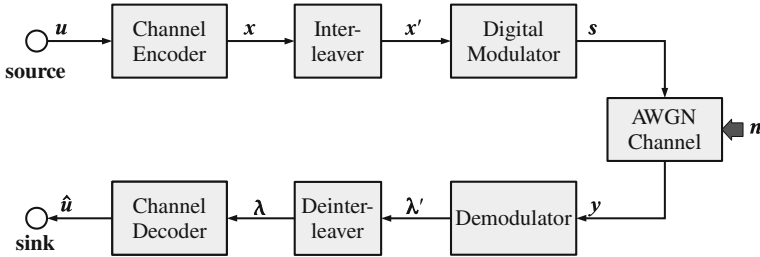


Fig. 2.1 Digital transmission system model of the outer transceiver

mentioned perfect synchronization, we model the noise as *additive white Gaussian noise* (AWGN) with zero mean and variance σ^2 . This channel model is used for all simulations in this manuscript. The received vector \mathbf{y} results in

$$\mathbf{y} = \mathbf{s} + \mathbf{n} \quad (2.1)$$

with \mathbf{n} the noise vector which is added to the sent transmit sequence.

The task of the demodulator is to transform the disturbed input sequence of symbols \mathbf{y} into a sequence of bit probabilities. This sequence λ' contains likelihood values, indicating the probability for each bit in the output of the channel encoder having been sent as a 1 bit or a 0 bit. This is done by calculating the conditional probability for every bit position $i \in 0, N - 1$ under the condition that \mathbf{y} was received. λ' is de-interleaved to λ which means we restore the original index positions of the channel decoder output. Note, that the prime to indicate an interleaved version of a sequence is omitted in the following. The channel decoder now has the task to detect and correct occurring errors. Thus it calculates an estimation of the sent information bits $\hat{\mathbf{u}}$. The AWGN channel and the demodulator calculations are explained in Sects. 2.1 and 2.2 respectively.

The modeling of different stages within the outer receiver chain requires a basic understanding of probability theory. For example, the model of a channel requires a probabilistic description of the effects of noise caused by the environment. Many more stages like the demodulator or the channel decoder deal with probabilities, as they calculate estimations of which information was most likely sent given the received disturbed input information. All decoding methods for LDPC codes, turbo codes, and convolutional codes were originally derived in the probability domain. The basic probability terms, which introduce terms like joint probability distribution, conditional probability, and marginalization are described in Appendix C. The probability terms are as well required to describe the information content which can be transmitted via an AWGN channel. This is denoted as channel capacity and is introduced in Sect. 2.3. The amount of information we can transmit via a channel can even be increased by using multiple antennas at the transmitter and/or receiver side. This so called MIMO transmission is introduced in Sect. 2.4. Especially the MIMO transmission with respect to a BICM transmission system will be further addressed in Chap. 8.

2.1 Channel Model

Typically, real world channels have complicated characteristics with all kind of interferences, e.g multi-path propagations, frequency selectivity and so on. However, as already mentioned, the task of the inner receiver is to provide a discrete channel to the outer receiver. The remaining discrete channel is modeled as additive white Gaussian noise (AWGN) channel, which is the most important channel model for the examination of channel codes. The AWGN channel model has no memory and adds to each random input variable a random noise variable N with density function:

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\} \quad (2.2)$$

The noise variable is Gaussian distributed with variance σ^2 and a mean value $\mu = 0$, see Sect. C: The output of the time discrete channel for each time step i is:

$$y_i = s_i + n_i. \quad (2.3)$$

In the following a *Binary Phase Shift Keying* (BPSK) modulation is assumed with $s \in \{-\sqrt{E_s}, \sqrt{E_s}\}$. $\sqrt{E_s}$ represents the magnitude of the signal such that the signal energy results in E_s . The received samples y_i are corrupted by noise. When we plot the distribution of the received samples two independent Gaussian distributions occur. The two distributions are centered at their respective mean value and are shown in Fig. 2.2. The three sub-figures show the different distributions for different values of the variance σ^2 and $\mu = \pm 1$ respectively. Depending on the variance of the additive Gaussian noise the overlapping of the two distributions, the gray shaded area in Fig. 2.2, changes.

One way to estimate the received bits (\hat{x}) is to apply a threshold detection at the 0 level. If the received information $y_i \geq 0$ then $\hat{x}_i = 1$, and vice versa. The resulting errors are indicated by the gray shaded areas in Fig. 2.2. Depending on the variance of the Gaussian distribution the possible error region gets smaller or larger. We define the ratio of incorrectly received bits to the total number of received bits as the *bit-error rate* (BER).

We can calculate the probability of errors for the presented binary transmission by using two shifted Gaussian distributions:

$$p(y|s = \sqrt{E_s}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (y - \sqrt{E_s})^2 \right\} \quad (2.4)$$

and

$$p(y|s = -\sqrt{E_s}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (y + \sqrt{E_s})^2 \right\}. \quad (2.5)$$

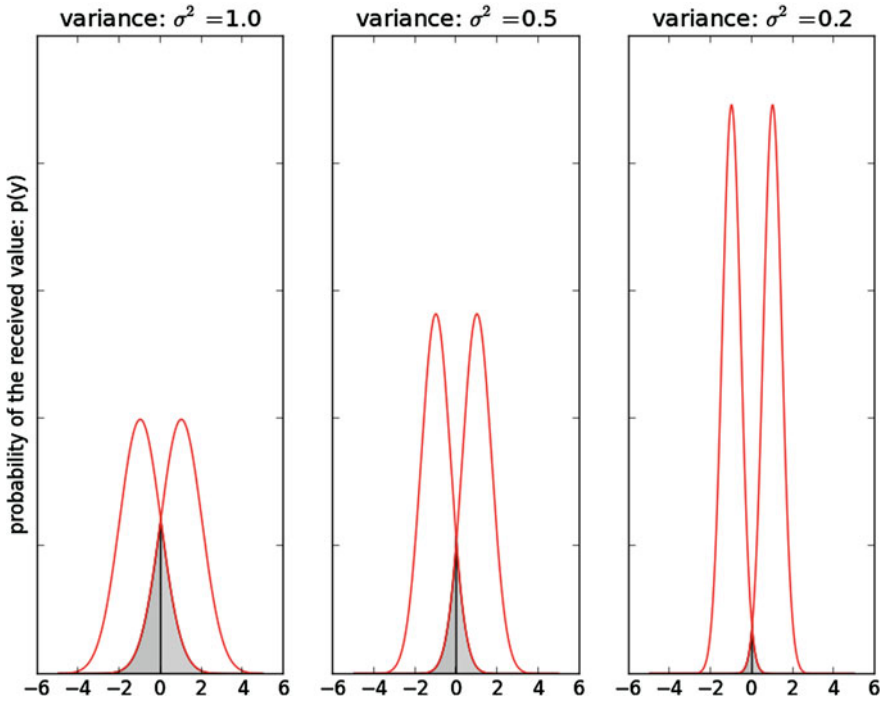


Fig. 2.2 Distributions of received values y with BPSK modulation for different variances of the noise

As mentioned, E_S is the transmission energy of one symbol, while the noise energy (N_0) is part of the channel properties. The noise energy is directly related to the variance of the Gaussian distribution, which is given by:

$$\frac{N_0}{2} = \sigma^2. \tag{2.6}$$

The quantity $\frac{N_0}{2}$ is the two-sided noise power spectral density. The probability of an error for the binary detection can be calculated using the Q -function.

$$Q(t) = Q(N > t) = \frac{1}{\sqrt{2\pi}} \int_t^\infty \exp\left\{-n^2/2\right\} dn. \tag{2.7}$$

The integral evaluates the probability that the random variable N is larger than a threshold t . A Gaussian distribution with zero mean and variance one is assumed in this function. However, for our transmission system we have to include the corresponding mean and variance values.

The resulting error rate when transmitting symbols with $s = -\sqrt{E_s}$ is

$$Q(Y > t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_t^\infty \exp\left\{-\frac{1}{2\sigma^2}(y + \sqrt{E_s})^2\right\} dn = Q\left(\frac{t + \sqrt{E_s}}{\sigma}\right). \quad (2.8)$$

The probability of a bit error assuming BPSK is the weighted probability of both possible transmissions ($s = \pm\sqrt{E_s}$), i.e.

$$P_b = Q\left(\frac{t + \sqrt{E_s}}{\sigma}\right) P(-\sqrt{E_s}) + Q\left(\frac{\sqrt{E_s} - t}{\sigma}\right) P(\sqrt{E_s}). \quad (2.9)$$

Both BPSK symbols are transmitted equally likely, $P(\sqrt{E_s}) = P(-\sqrt{E_s})$, while the threshold t for a detected bit error lies at $t = 0$ as shown in Fig. 2.2. Thus P_b evaluates to:

$$P_b = Q\left(\frac{\sqrt{E_s}}{\sigma}\right) = Q\left(\sqrt{\frac{2E_s}{N_0}}\right). \quad (2.10)$$

For the evaluation of the obtained results, typically graphs as shown in Fig. 2.3 are used. Shown is the bit error rate on the y-axis in logarithmic scale, the x-axis represents the signal-to-noise ratio in E_s/N_0 for an AWGN channel.

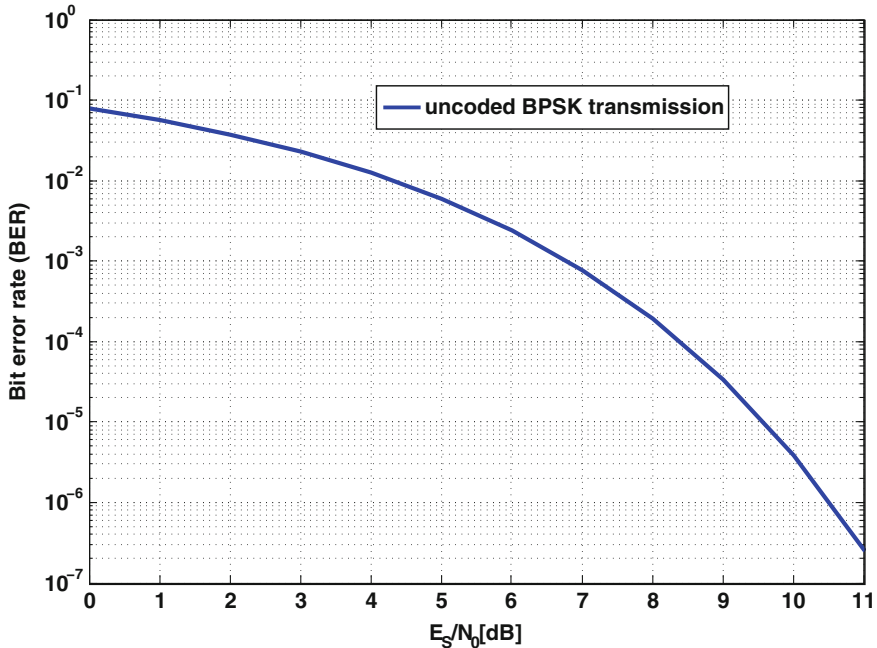


Fig. 2.3 Bit error rate for a BPSK transmission without channel coding (uncoded)

Quality of Service

For many services, data are packed into frames (blocks, packets, ...). For packet based transmission like LTE up to 25 codewords are grouped into one transmission time interval frame (TTI), which implies a higher-level grouping of bits with additional header information. However, in the context of channel coding in the baseband model of Fig. 2.1, the terms frame, block, and packet are used interchangeably and all denote one transmitted codeword. Often it is irrelevant how many bits could be received correctly within one frame. It only matters whether the frame contains errors or not, as in many services every erroneous frame is considered lost, including all the data therein. This leads to the ratio of the number of erroneous frames to the total number of frames, the *frame-error rate* (FER). Both BER and FER are usually expressed in the logarithmic domain:

$$BER|_{dB} = \log \left(\frac{\text{number of erroneous bits}}{\text{total number of bits}} \right), \quad (2.11)$$

and

$$FER|_{dB} = \log \left(\frac{\text{number of erroneous frames}}{\text{total number of frames}} \right). \quad (2.12)$$

BER and FER always have to be put in relation to the noise present in a channel. The relative amount of noise is expressed by the *signal-to-noise ratio* (SNR).

$$\text{SNR} = \frac{E_S}{N_0}. \quad (2.13)$$

The energy needed for the transmission of one information bit (E_b) is generally not equal to E_S . E_b incorporates also information about the *code rate* (R) and the number of modulation bits Q . The code rate is defined as the ratio of information bits per coded bit. A code rate $R = 1/3$ for example denotes that three bits are transmitted to communicate one information bit. The number of modulation bits determines how many bits are grouped to one modulation symbol. For the already mentioned BPSK modulation this is $Q = 1$. E.g. in the case of $Q = 2$ we use $2^2 = 4$ different symbols for transmission, see Sect. 2.2. The SNR can consequently also be expressed in relation to the information bit energy instead of the symbol energy. The relation between the two forms of expression is:

$$\frac{E_S}{N_0} = Q \cdot R \cdot \frac{E_b}{N_0}, \quad (2.14)$$

or in the logarithmic domain:

$$\left. \frac{E_S}{N_0} \right|_{dB} = 10 \cdot \log(Q) + 10 \cdot \log(R) + \left. \frac{E_b}{N_0} \right|_{dB}. \quad (2.15)$$

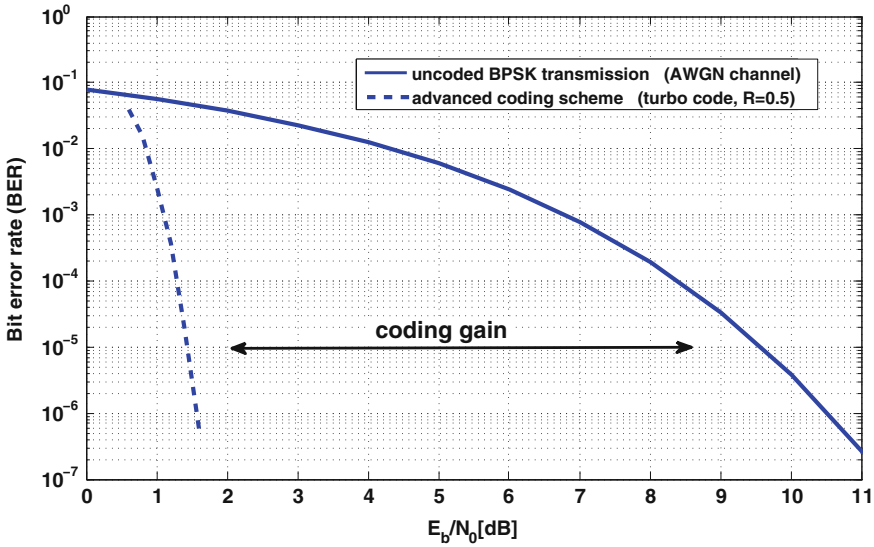


Fig. 2.4 Bit error rate for a BPSK transmission with and without channel coding

For comparing different channel coding schemes the required SNR level has to be evaluated with respect to a certain frame error rate. In order to obtain a statistic for the frame error rate at a given signal-to-noise ratio so called Monte Carlo simulations can be performed. According to the modeling of Fig. 2.1 a large number of frames are simulated, always at one fixed signal-to-noise ratio and one fixed channel coding setup (type, block length, code rate). Then, the simulations are repeated for different noise levels, by changing the variance of the Gaussian noise.

Figure 2.4 shows again the bit error rate on the y-axis in logarithmic scale, while the x-axis represents the signal-to-noise ratio in E_b/N_0 for an AWGN channel. Shown is the BER for an uncoded system and a system using a channel code, a so called turbo code. Turbo codes are introduced in Chap. 6 and are defined in the 3GPP communication standards. A code rate of $R = 0.5$ is used which means two bits are transmitted for each information bit. Even though transmitting more coded bits, the energy required to transmit an information bit reliable is lower than in the uncoded case. We can see a large gain in terms of signal-to-noise ratio to obtain a specific BER. In other words: With respect to a higher noise level we can obtain the same quality-of-service when using a channel coding scheme. This gain between coded and uncoded transmission is denoted as coding gain.

For turbo codes we transmit the information in blocks and we should use the resulting frame error rate to judge on the achieved communications performance. Figure 2.5 shows FER curves over SNR for the same kind of turbo code, however, applying different code rates for transmission.

All simulations have a fixed number of information bits of $K = 6144$, while the codeword size ranges from $N = 18432$ to $N = 6400$ bits, which corresponds to

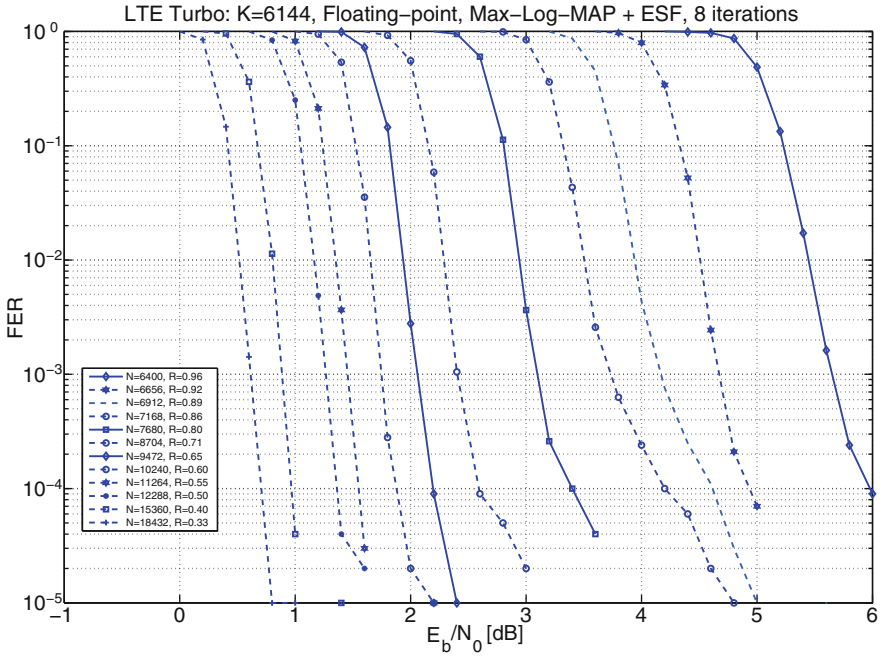


Fig. 2.5 FER vs. SNR for different code rates of a LTE turbo code

code rates of $R = 1/3$ to $R = 0.96$, respectively. For each E_b/N_0 and each code rate 100000 frames were simulated, in order to get statistically stable results.

All curves show a characteristic drop of the observed FER for an increased SNR. All code rates have a $FER = 1$ at a very low E_b/N_0 ratio. E.g. at $E_b/N_0 = 0\text{dB}$ all simulations have at least one bit error per frame after the channel decoder. It can be seen that the lower the code rate, the smaller the SNR level at which a certain FER can be obtained. It is really important that we compare the SNR in terms of E_b/N_0 , since only then a comparison of different code rates and the evaluation of the final achieved coding gain is possible. When comparing the coding gain of two coding schemes, here two different code rates, we always need a certain FER as reference. For practical wireless transmission the reference frame error rate is often $FER = 10^{-2}$ or $FER = 10^{-3}$.

2.2 Digital Modulator and Demodulator

The modulator maps groups of Q bit to one symbol, before transmitting these via the channel. Task of this modulation is the increase of sent information per channel usage.

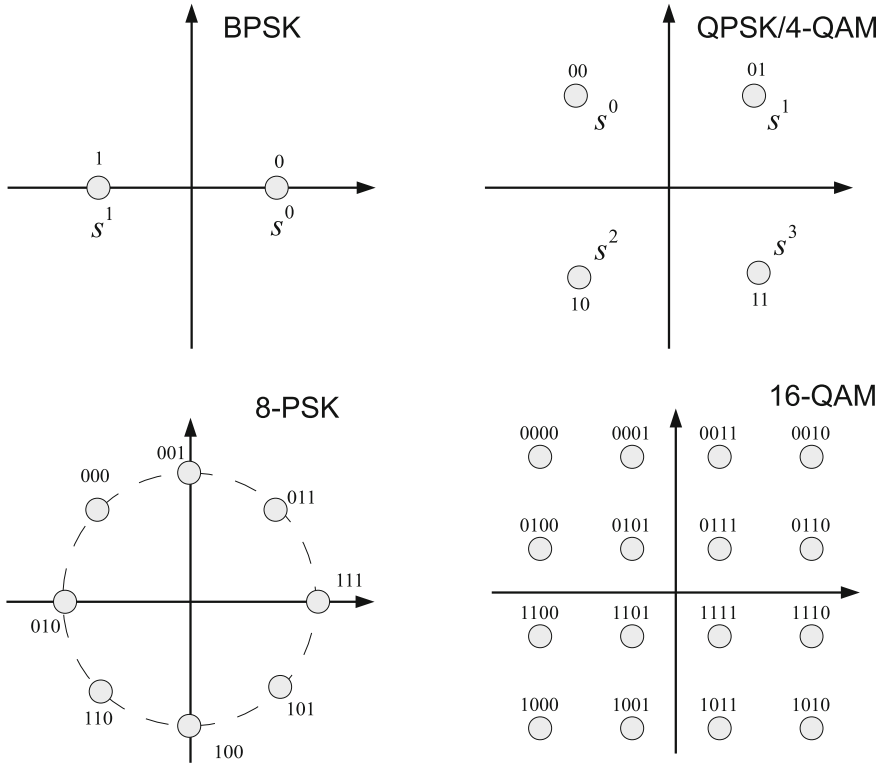


Fig. 2.6 Different modulation types: binary phase shift keying, 4-QAM, 8

Figure 2.6 shows different signal constellations which can be used for the modulation. 2^Q different signal points exist, depending on the number of mapped bits. For the phase-shift keying with 2, 4, or 8 signal points (BPSK, QPSK, 8-PSK) each signal point has the same energy E_S . The figure shows as well a quadrature amplitude modulation (QAM) with 16 signal points. The energy for different signal points may vary for QAM modulation types, while the average signal energy \bar{E}_S will be normalized, with

$$\bar{E}_S = \frac{1}{2^Q} \sum_{k=0}^{2^Q-1} (Re(s_k)^2 + Im(s_k)^2). \tag{2.16}$$

The amplitudes ($Re(s)$, $Im(s)$) of the complex constellation points are chosen accordingly to ensure $\bar{E}_S = 1$. The bit to symbol mapping shown in Fig. 2.6 is a so called Gray mapping. A Gray mapping or Gray coding ensures that two neighboring symbols differ in only one bit. This type of mapping is most often used in communication systems which use the introduced bit-interleaved coded modulation transmission.

The demodulator is the first stage after the channel model, see Fig. 2.1. The input to the demodulator is the sequence of received values y , where each value y_i comprises the information of Q bits corrupted by noise.

The task of the demodulator is to calculate for each of these Q bits, an estimate of the originally sent bit. These bit estimates can be given as so-called hard-values or soft-values.

- **Hard-values:** These values are represented by just one bit per transmitted bit. For the BPSK example in Fig. 2.2 this would be the sign bit of the received information, which can be extracted in this case by a simple threshold decision.
- **Soft-values:** These values include a probability information for each bit. Thus we pass a sign value and an associated confidence information, which is derived from the received sample.

The advantage of the soft-values is an increased possible coding gain of the channel decoder. This gain is up to 3 dB when comparing channel decoding with hard-input or soft-input information, respectively [1]. Nearly all modern channel decoders require soft-values as input information.

In order to obtain soft-values we have to calculate the a posteriori probability (APP) that the bit x_i (or symbol s_j) was transmitted under the condition that y_j was received. Also possible to calculate a symbol probability $P(s_j|y_j)$ it is often required to calculate directly the corresponding bit probability $P(x_i|y_j)$. This is due to the BICM system at which the bit information is interleaved to ensure independent probabilities at the channel decoders input. For a simple modulation scheme with one bit mapped to one symbol we will calculate one a posteriori probability. In the case of a higher mapping we have to extract Q APP values, one for each mapped bit. Thus, the demodulator has to calculate for one received symbol y_j :

$$P(x_i|y_j), \dots, P(x_{i+Q-1}|y_j) \quad (2.17)$$

We can not directly calculate this probability, but instead we have to use Baye's theorem, which then results in:

$$P(x_i|y_j) = P(y_j|x_i) \cdot P(x_i)/P(y_j) \quad (2.18)$$

$P(x_i|y_j)$ is denoted as a posteriori probability, $P(y_j|x_i)$ as conditional probability, and $P(x_i)$ as a priori information. The calculation of $P(y_j|x_i)$ and $P(y_j)$ depends on the channel characteristics, while $P(x_i)$ reflects an a priori information about the source characteristics. For the demodulation process we do not calculate the probability of $P(x_i|y_j)$ directly; rather we operate in the so called log-likelihood domain, which calculates

$$\lambda(x_i|y_j) = \ln \frac{P(y_j|x_i = 0) \cdot P(x_i = 0)/P(y_j)}{P(y_j|x_i = 1) \cdot P(x_i = 1)/P(y_j)} \quad (2.19)$$

$P(y_j)$ depends only on the channel characteristics and is a constant for a certain received y_j . The term cancels out while the calculation can be split into two parts.

$$\lambda(x_i|y_j) = \ln \frac{P(y_j|x_i = 0)}{P(y_j|x_i = 1)} + \ln \frac{P(x_i = 0)}{P(x_i = 1)} = \lambda(y_j|x_i) + \lambda(x_i) \quad (2.20)$$

For binary sources without a priori information, i.e., with an assumed equal distribution of 1s and 0s, the probabilities $P(x_i = 1) = P(x_i = 0)$ are equivalent and thus cancel out (i.e., $\lambda(x_i) = \ln \frac{P(x_i=0)}{P(x_i=1)} = 0$).

Log-likelihood ratios (LLR) have several advantages over operating on the probabilities directly. These advantages become evident for evaluating $\lambda(y_j|x_i)$ when taking the characteristics of a channel into account, e.g. an AWGN channel. The mathematical operations for BPSK demodulator is presented in the following. The demodulation for a 4 amplitude modulation is presented in Appendix D, respectively.

BPSK Demodulator

In the case of BPSK modulation only one bit is mapped to every symbol s . Typically the mapping is $x_i = 0 \rightarrow s_i = +1$ and $x_i = 1 \rightarrow s_i = -1$. For each received value one LLR value is calculated, which is here denoted with $\lambda(y|x)$. For the sake of simplicity, we omit the indices $i = j$ for the BPSK demodulation. In order to calculate $\lambda(y|x)$ under the assumption of an AWGN channel we have to evaluate the Gaussian density function with a shift of its mean value according to the amplitude of the transmitted symbol s , thus:

$$p(y|s) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(y-s)^2}{2\sigma^2} \right\}. \quad (2.21)$$

For the BPSK mapping given above the computation of the demodulator results in:

$$\begin{aligned} \lambda(y|x) &= \ln \frac{P(y|s = +1)}{P(y|s = -1)} \\ &= \ln \frac{\exp \left\{ -\frac{1}{2\sigma^2} (y-1)^2 \right\}}{\exp \left\{ -\frac{1}{2\sigma^2} (y+1)^2 \right\}} \\ &= \frac{1}{2\sigma^2} (y+1)^2 - \frac{1}{2\sigma^2} (y-1)^2 \\ &= \frac{2}{\sigma^2} y \end{aligned} \quad (2.22)$$

Thus the channel LLR value $\lambda = \frac{2}{\sigma^2}y$ of the received symbol is just the received sample corrected by the *channel reliability factor* $L_{ch} = \frac{2}{\sigma^2}$.¹ We can summarize the advantage of the log-likelihood logarithm of Eq. 2.20 for this demodulator example: Only the exponents are of interest, while all normalizers of the density function ($1/\sqrt{2\pi\sigma^2}$) cancel out. Furthermore, in the logarithmic domain multiplications and divisions turn into additions, and subtractions, respectively, which is of great advantage for a hardware realization.

2.3 Channel Capacity

As mentioned before, the goal of the channel code is to transmit the information bits reliably through a discrete unreliable channel at the highest practicable rate. The channel capacity gives us the bound for a reliable (error free) transmission with respect to a given signal-to-noise ratio. For evaluating this question we have to introduce the definition of the information content I . The information content of a discrete event x is defined as:

$$I(x) = \log_2 \frac{1}{P(x)}, \quad (2.23)$$

with $P(x)$ the probability of the single event x , see Appendix C. The logarithm of base two constrains this information measure to the dimension ‘bit’. This is the smallest and most appropriate unit, especially when dealing with digital data. Often we are interested in the average information content of a discrete random variable X with probability density function $p(x)$. This is denoted as entropy H and can be calculated via the expectation with respect to I :

$$H(X) = E \left\{ \log_2 \frac{1}{p(x)} \right\} = \sum_{x \in \Omega_x} p(x) \log_2 \frac{1}{p(x)}. \quad (2.24)$$

$p(x)$ defines the probability of the random variable X , with Ω_x defining the respective discrete sample space. The entropy of the random variable X is maximized when each element occurs equally likely. Assuming Ω_x has 2^K elements, the entropy results in:

$$H(X) = \sum_{i=0}^{2^K-1} \frac{1}{2^K} \log_2 \frac{1}{\frac{1}{2^K}} = 2^K 2^{-K} K = K \text{ [bit]}. \quad (2.25)$$

Assuming two discrete random variable X and Y we can define a joint entropy:

¹ A channel model related to the AWGN channel is the so called fading channel. In this channel model each received value has an additional attenuation term a , and the channel reliability factor turns into $L_{ch} = a \frac{2}{\sigma^2}$, with a being the fading factor.

$$H(X; Y) = E \left\{ \log_2 \frac{1}{p(x, y)} \right\} = \sum_{x \in \Omega_x} \sum_{y \in \Omega_y} p(x, y) \log_2 \frac{1}{p(x, y)}. \quad (2.26)$$

The joint entropy is very important since we can describe the dependency of the information between two variables. Its relation is closely linked to Bayes theory, see Eq.C.9. However, since the entropy is defined in the logarithm domain, additions instead of multiplications result within the chain rule:

$$H(X, Y) = H(X|Y) + H(Y) = H(Y|X) + H(X) \quad (2.27)$$

The mutual information $I(X; Y)$ gives us the information that two random variables X and Y share. If this shared information is maximized we have as much knowing of one variable to give information (reduce uncertainty) above the other. Figure 2.7 shows the dependencies between the mutual information and the entropies. The illustration shows a possible transmission of variable X with entropy $H(X)$ via a channel. Information will get lost during transmission which is expressed as conditional entropy (equivocation) $H(X|Y)$. In turn irrelevant information will be added $H(Y|X)$. The final received and observable entropy is $H(Y)$. The mutual information of two random variables can be calculated by:

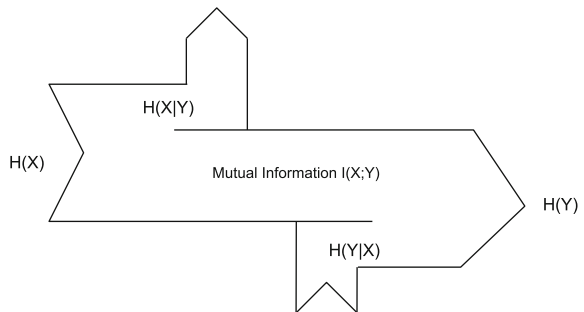
$$I(X; Y) = H(Y) - H(Y|X) = H(X) - H(X|Y). \quad (2.28)$$

We can define the mutual information as well directly via the joint and marginal probabilities

$$I(X; Y) = \sum_{x \in \Omega_x} \sum_{y \in \Omega_y} p(x, y) \log_2 \frac{p(x, y)}{p(x) \cdot p(y)}. \quad (2.29)$$

Note that the mutual information of the input X and the output Y of a system can be estimated by Monte Carlo simulations. By tracking the density functions by histogram measurements we can evaluate Eq.2.29. This is a pragmatic approach especially when the density functions are not entirely known. As mentioned before, the channel capacity gives an upper bound for a reliable (error free) transmission

Fig. 2.7 Illustration of the mutual information and its dependencies



with respect to a given signal-to-noise ratio. Assuming a random input variable X and the observation Y , the capacity of the utilized channel is its maximum mutual information

$$C = \sup_{p(x)} I(X; Y). \quad (2.30)$$

Equation 2.30 requires the knowledge of the joint distribution of X and Y and the distribution and the channel. Three assumptions have to be met to achieve the maximum capacity which is denoted as Shannon limit:

- the input variable X requires to be Gaussian distributed,
- the noise has to be additive with Gaussian distribution,
- the length of the input sequence has to be infinite.

In the case of an AWGN channel with a constant noise power of N_0 the channel capacity can be evaluated to [2]:

$$C = \log_2 \left(1 + \frac{E_S}{N_0} \right) = \log_2 \left(1 + \frac{R E_b}{N_0} \right). \quad (2.31)$$

When using a transmission rate $R < C$ an error probability as low as desired can be achieved. With Eq. 2.31 we can evaluate the bound for the required bit energy to noise power by assuming the code rate to be equal to the capacity

$$\frac{E_b}{N_0} = \frac{2^R - 1}{R}. \quad (2.32)$$

The Shannon limit gives a bound for the minimum bit energy required for reliable transmission assuming an infinite bandwidth, i.e. the code rate approaches zero ($R \rightarrow 0$). The absolute minimum bit energy to noise power required for reliable transmission is $E_b/N_0 = -1.59$ dB.

The bandwidth B is limited for many communications systems and quite expensive from an economic point of view. The total noise power is given by $N = B \cdot N_0$, the total signal energy $S = R_b E_b$ is defined via the data rate R_b . R_b is the number of bits we can transmit within a certain amount of time. According to the Nyquist criterion this will be $2BT$ symbols within a time unit of T . The channel capacity for a channel with limited bandwidth B can be expressed as

$$C_B = B \cdot \log_2 \left(1 + \frac{S}{N} \right) \quad [bits/s]. \quad (2.33)$$

We can see that it is possible to trade off bandwidth for signal-to-noise ratio. The introduced capacity before, without bandwidth, defines the spectral efficiency. This spectral efficiency gives us the number of bits per second per Hertz [$bits/s/Hz$] which is an important metric since it defines the amount of information that can be transmitted per Hertz of bandwidth.

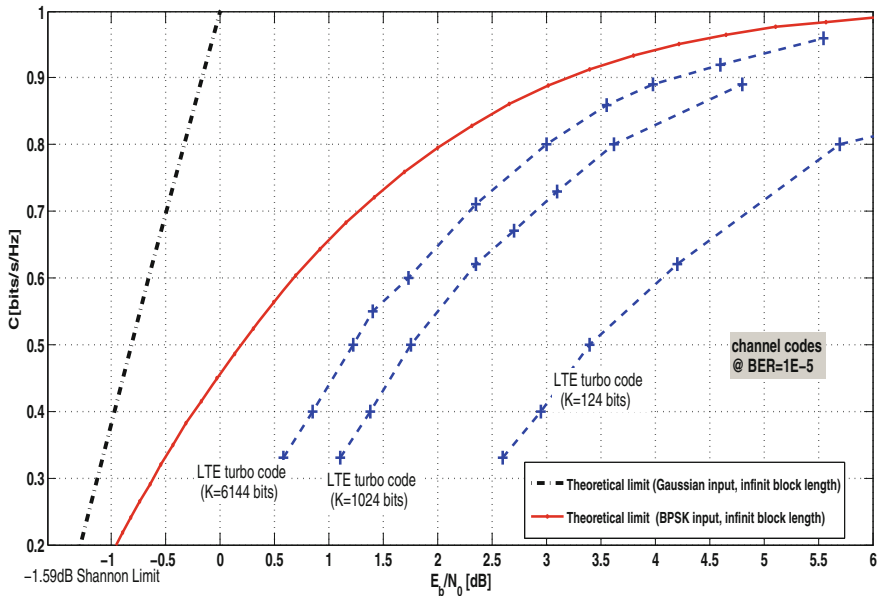


Fig. 2.8 Spectral and power efficiency for Shannon limit, theoretical limit applying BPSK modulation, and turbo codes with different block sizes

Figure 2.8 shows the spectral efficiency versus the signal-to-noise ratio E_b/N_0 . Plotted is the optimum theoretical limit which can only be reached for Gaussian distributed values given as input to a channel. By constraining the channel input to be binary (e.g. Binary Phase Shift Keying (BPSK) [3]) the resulting capacity is lower than the optimum capacity. The random input variable S has a sample space of $s \in \{-1, +1\}$ while the output is modeled as a continuous random variable $y \in R$. Then the mutual information evaluates to:

$$I(S; Y) = \frac{1}{2} \sum_{s_i \in \{+1, -1\}} \int_{-\infty}^{+\infty} p(y|s_i) \cdot \log_2 \left(\frac{2p(y|s_i)}{p(y|s_i = -1) + p(y|s_i = +1)} \right) dy \quad (2.34)$$

The resulting theoretical limit for BPSK input is shown as well in Fig. 2.8.

The channel capacity is the tight upper bound for a reliable transmission, however, it tells us not how to construct good codes which can approach this limit. Furthermore, the limit is only valid for infinite block sizes. Therefore, Fig. 2.8 shows coding schemes with different block lengths and coding rates. The utilized coding schemes here are turbo codes utilized in the LTE standard.

All codes are plotted at a reference bit error rate of $BER = 10^{-5}$. The results are collected from different simulations over a simple AWGN channel and BPSK modulation, with K the number of information bits ($K = 124$, $K = 1024$, $K = 6144$).

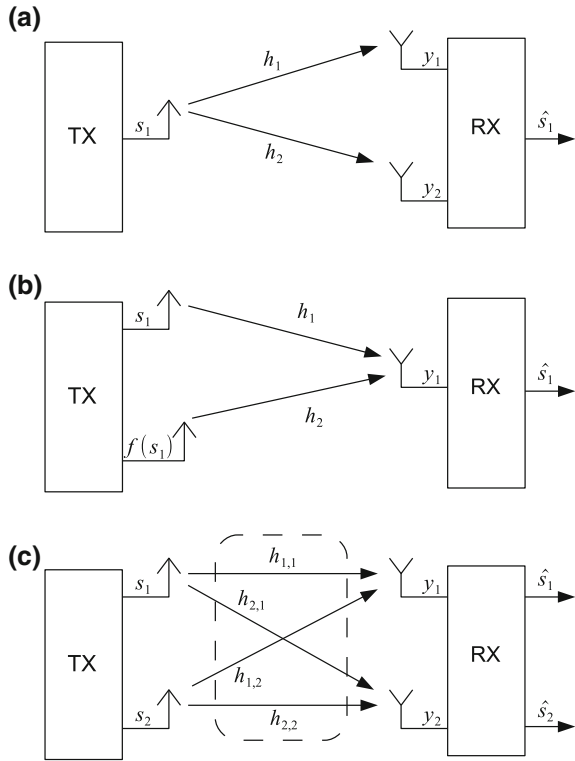
The different code rates are obtained by a so called puncturing scheme defined in the communication standard. The simulated LTE turbo codes are introduced in Chap. 6. One important aspect can be seen in this figure. The larger the block length the smaller the gap to the theoretical limit. For block sizes of $K > 100k$ bits the gap to the theoretical limit is less than 0.2 dB.

2.4 Multiple Antenna Systems

Multiple antenna system is a general expression for a transceiver system which relies on multiple antennas at the receiver and/or at the transmitter side. The major goal of using a multiple antenna system is to improve the system performance. This could either mean the transmission of more information for a given amount of energy, or to achieve a better quality of service for a fixed amount of energy [4]. By using multiple antennas different types of gains can be obtained.

- **Array gain:** The array gain can increase the signal-to-noise ratio by a coherent combination of the received signal. Achieving the coherent detection we need the knowledge of the channel to align the phases of signals. The average SNR grows proportionally to the number of received signals, i.e., multiple receive antennas collect more signal energy compared to a single receive antenna.
- **Diversity gain:** Diversity gain is obtained when transmitting information multiple times via different transmission channels. Different channels may undergo a different fading, which means, the attenuation of the channel path will change. The assumption is that different paths do not fade concurrently. The receiver will gather the diversified information and has then the possibility to mitigate the fading effects. The higher the number of paths, the higher the probability to mitigate occurring fading effects. There exist different sources of diversity: diversity in time (also called temporal diversity), diversity in frequency, and diversity in space. Temporal diversity can be introduced in an artificial way by the transmitter, by transmitting the same information at different points in time. Frequency diversity can be utilized when transmitting at different frequencies. Spatial diversity uses the distance between receive or transmit antennas. The goal of diversity is always to improve the reliability of communication. The maximum diversity gain we can obtain is of $d = M_T M_R$ th order, i.e. each individual path can be exploited. M_T defines the number of transmit antennas, M_R defines the number of receive antennas, respectively. The diversity d is defined as the slope of the average error probability at high signal-to-noise ratio, i.e., the average error probability can be made to decay like $1/SNR^d$ [5].
- **Multiplexing gain:** Maximum gain in spatial multiplexing is obtained when at each time step and each antenna a new symbol is transmitted. The MIMO detector on the receiver side has the task to separate the symbols which are interfered and corrupted by noise. MIMO channels offer a linear increase $r = \min(M_T, M_R)$ in capacity [5]. It was shown in the high SNR regime that the rate of the system (capacity)

Fig. 2.9 Three different multiple antenna systems



can grow with $r \log(SNR)$. The factor r is denoted as spatial multiplexing gain and defines the number of degrees of freedom of a system (bit/s/Hz). There exists a trade off between gain in diversity and gain in spatial multiplexing [5].

Three major scenarios exist to set up a multiple antenna system which are shown in Fig. 2.9. All three possibilities are used in the LTE communication standard, a good overview can be found in [6]. Here, only the obtained diversity or multiplexing gain is highlighted with respect to the different antenna constellations.

Single-Input Multiple-Output Transmission

The most straight forward system using multiple antennas is the single-input multiple-output (SIMO) system which is shown in Fig. 2.9a. The receiver collects two versions of the sent symbol s_1 which is denoted as receive diversity. The channel coefficients h_1 and h_2 define the different attenuations of both paths. The highest degree of diversity is obtained when the channel coefficients are uncorrelated. Assuming M_R receive antennas we can obtain diversity of order M_R . The SIMO system reflects a typical point-to-point transmission from a mobile device to a base station (uplink).

Multiple-Input Single-Output Transmission

Multiple antennas can be applied as well at the transmitter which the goal to obtain transmit diversity, see Fig. 2.9b. Again, to maximize the diversity the channel coefficients have to be uncorrelated. This can be achieved by an appropriate distance between transmit antennas. Diversity can be in addition introduced by applying a function to the sent information, denoted here as $f(s_1)$. The function indicates the introduction of a time or frequency offset. One prominent example is the so called space-time coding [7] which is as well introduced in the LTE communication standard. The point-to-point multiple-input single-output (MISO) system can be efficiently exploited in the downlink of wireless transmission systems.

Multiple-Input Multiple-Output Transmission

MIMO systems can achieve either a diversity gain and/or a multiplexing gain. Especially the multiplexing gain is of large interest. For achieving the highest multiplexing gain the symbol stream is demultiplexed to multiple transmit antennas. At each transmit antenna and in each time slot we transmit independent symbols. The receiver side collects the superposed and noise disturbed samples from multiple receive antennas. This spatial multiplexing yields in a linear increase of system capacity, i.e. the spectral efficiency is increased by a factor of $\min(M_T, M_R)$ for a given transmit power. LTE defines a spatial multiplexing scheme using 4 transmit and 4 receive antennas (4×4) while applying a 64-QAM modulation. The number of bits per second per Hertz would be 24 [bits/s/Hz] without channel coding. However, spatial multiplexing should be combined with a channel coding scheme which is often done according to a BICM system as shown in Fig. 2.1.

Ergodic Channel and Quasi-Static Channel

Figure 2.9c shows a two-by-two antenna system. The transmission of a data and the received vector can be modeled in general as:

$$\mathbf{y}_t = \mathbf{H}_t \cdot \mathbf{s}_t + \mathbf{n}_t, \quad (2.35)$$

with \mathbf{H}_t the channel matrix of dimension $M_R \times M_T$ and \mathbf{n}_t the noise vector of dimension M_R . The entries in \mathbf{H}_t can be modeled as independent, complex, zero-mean, unit variance, Gaussian random variables [8]. The presented channel model is again a time discrete model. In this model fading coefficients are represented by the matrix \mathbf{H}_t which is in contrast to the already introduced AWGN channel in which no fading coefficients are present. The average transmit power E_S of each antenna is normalized to one, i.e. $E\{\mathbf{s}_t \mathbf{s}_t^H\} = \mathbf{I}^2$. We assume an additive Gaussian noise at each

² \mathbf{I} is the identity matrix and $()^H$ denotes the Hermitian transpose.

receive antenna, with $E\{\mathbf{n}_t \mathbf{n}_t^H\} = N_0 \mathbf{I}$. The signal-to-noise power at each receive antenna can be expressed as $SNR = \frac{M_T E_s}{N_0}$. We have to distinguish two major cases for simulating the MIMO channel:

- Quasi-static channel: \mathbf{H}_t remains constant for a longer time period and is assumed to be static for one entire codeword of the channel code.
- Ergodic channel: \mathbf{H}_t changes in each time slot and successive time slots are statistically independent.

Assuming a quasi-static channel we cannot guarantee an error free transmission since we cannot average across multiple channel realizations. The fading coefficients of \mathbf{H}_t could by chance prohibit a possible correction of the sent signal. Another definition of the channel capacity has to be used which is called outage capacity [9]. The outage capacity is associated with an outage probability P_{out} which defines the probability that the current channel capacity is smaller than the transmission rate R .

$$P_{out} = P(C(\rho) < R) \quad (2.36)$$

$C(\rho)$ defines the capacity of a certain channel realization with the signal-to-noise ρ of the current fading coefficients. The outage probability can be tracked via Monte-Carlo simulations. The capacity C of a given channel model differs for a quasi-static or ergodic channel.

For ergodic channels the capacity is higher as shown in [9]. The capacity C for an ergodic channel can be calculated as expectation via the single channel realizations:

$$C = E \left\{ \log_2 \det \left(\mathbf{I} + \frac{SNR}{M_T} \cdot \mathbf{H}_t \mathbf{H}_t^H \right) \right\} \quad [bits/s/Hz]. \quad (2.37)$$

SNR is the physically measured signal-to-noise ratio at each receive antenna. Assuming an ergodic channel the channel capacity increases linearly with $\min(M_T, M_R)$. Transmission rates near MIMO channel capacity can be achieved by a bit-interleaved coded MIMO system. The resulting BICM-MIMO systems and the challenges for designing an appropriate MIMO detector are shown in Chap. 8. A comprehensive overview of MIMO communications can be found in [8].

References

1. Bossert, M.: Kanalcodierung, 2nd edn. B. G. Teubner, Stuttgart (1998)
2. Moon, T.K.: Error Correction Coding: Mathematical Methods and Algorithms. Wiley-Interscience, Hoboken (2005)
3. Proakis, J.G.: Digital Communications, 3rd edn. McGraw-Hill, Inc., Boston (1995)
4. Boelskei, H.: Fundamental tradeoffs in MIMO wireless systems. In: Proceedings of the IEEE 6th Circuits and Systems Symposium on Emerging Technologies: Frontiers of Mobile and Wireless, Communication, vol. 1, pp. 1–10 (2004). doi:[10.1109/CASSET.2004.1322893](https://doi.org/10.1109/CASSET.2004.1322893)
5. Zheng, L., Tse, D.N.C.: Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels. IEEE Trans. Inf. Theory **49**(5), 1073–1096 (2003). doi:[10.1109/TIT.2003.810646](https://doi.org/10.1109/TIT.2003.810646)

6. Dahlman, E., Parkvall, S., Skoeld, J., Beming, P.: 3G Evolution. Elsevier, HSPA and LTE for Mobile Broadband, Oxford (2008)
7. Alamouti, S.M.: A simple transmit diversity technique for wireless communications. *IEEE J. Sel. Areas Commun.* **16**(8), 1451–1458 (1998)
8. Paulraj, A.J., Gore, D.A., Nabar, R.U., Bölcskei, H.: An overview of MIMO communications—a key to gigabit wireless. *Proc. IEEE* **92**(2), 198–218 (2004). doi:[10.1109/JPROC.2003.821915](https://doi.org/10.1109/JPROC.2003.821915)
9. Telatar, I.: Capacity of multi-antenna Gaussian channels. *Eur. Trans. Telecommun.* **10**, 585–595 (1999)