DIGITAL AMPLITUDE MODULATION

Due to the technical development, especially in the area of semiconductors, powerful digital devices such as digital signal processors (DSPs) are available today for application in data transmission systems. In comparison with analog transmission techniques, systems with digital modulation allow an identical reproduction of the transmitted data by means of error control coding. Thus, data transmission with a very high bandwidth efficiency can be achieved.

The task of digital amplitude modulation is to map digital data to a continuous-time signal that can be transmitted over a channel. It is essential that the generated signals be chosen such that the receiver can distinguish between the received signals after transmission over the channel.

A digital M-ary amplitude modulation is defined as follows: A number $m = \log_2 M$ bits of a binary data sequence is assigned unambiguously to one out of M complex values called modulation symbols. The set of all M modulation symbols is called a constellation diagram of the modulation scheme. In the transmitter the sequence of mN data bits is mapped to a sequence of N modulation symbols S_n, $n = 1, \ldots, N$. From this sequence of modulation symbols the complex baseband signal is composed according to the following equation:

$$s(t) = \sum_n S_n p(t - nT)$$

where $p(t)$ is an arbitrary modulation pulse, for example, with rectangular shape. The modulation schemes characterized by the preceding equation are also termed linear modulation schemes (as opposed to continuous phase-modulation methods). Alternatively to the direct mapping of the data bits (absolute encoding) to the modulation symbols, the information can instead be contained in the quotient of two successive modulation symbols (differential encoding).

The resulting baseband signal is mixed up to the carrier frequency and transmitted over the channel. After transmission, the receiver determines the modulation symbols, which are received as

$$R_n = H_n S_n + N_n \tag{1}$$

where the complex factor H_n describes the channel influence that can cause an attenuation and a phase shift of the modulation symbols. The term N_n describes additive noise. A widely used special case is additive white Gaussian noise (AWGN).

Based on R_n the receiver has to decide on one of the possibly transmitted modulation symbols S_n that have been distorted by the channel. After this decision the data bits are determined by demapping the detected modulation symbol. If a wrong modulation symbol is detected, bit errors occur.

An advantage of digital modulation schemes is the fact that such bit errors can possibly be corrected by means of error control coding.

In some transmission environments (e.g., wireless), the signal may reach the receiver via several paths with different delays due to fraction and reflection. This behavior is called multipath propagation and leads to different attenuations within the frequency band (the so-called frequency selectivity of the channel). The received signal consists of several replicas of the transmitted signal, which causes the problem of interference of an adjacent modulation symbol. To counteract this intersymbol interference (ISI), a channel equalization has to be performed in the receiver, which might be a complex

Figure 1. Block diagram of linear amplitude modulation.

task, depending on the maximum delay $\tau_{\max}$ of the channel and the symbol duration T.

An alternative to modulating the information on a single carrier consists of dividing the total bandwidth into many (say K) narrow-band subcarriers. On each of the subcarriers, a portion of the total information is transmitted. This technique is called orthogonal frequency division multiplexing (OFDM). Consequently, the transmission channel appears to be frequency-nonselective within the subcarrier bandwidth, which is an approximation with reasonable error. Thus, the influence of the channel can be described by a single channel transfer factor $H_{n,k}$ for each subcarrier, where $k = 0, \ldots, K - 1$ refers to the subcarrier number. Thus, Eq. (1) can be modified:

$$R_{n,k} = H_{n,k} S_{n,k} + N_{n,k} \tag{2}$$

It follows that the task of channel equalization is reduced to estimating this complex factor for each subcarrier. Furthermore, ISI can be avoided totally if a periodic extension is added to the OFDM block that is at least of length $\tau_{\max}$.

MODULATION AND DEMODULATION

Modulation

As described earlier, a carrier is modulated by a complex modulation symbol S_n. Generally, this complex value is generated by mapping a sequence of m bits to a specific point in the constellation diagram followed by an optional differential encoding; see Fig. 1. Figures 2 to 5 show constellation diagrams of different modulation schemes, namely amplitude shift keying (M-ASK), quadrature amplitude modulation (M-QAM), phase shift keying (M-PSK), and amplitude and phase shift keying (M-APSK).

Absolute Encoding. If absolute encoding is applied, the data bits to be transmitted are directly mapped to the modulation symbols S_n. Examples of this modulation technique are PSK and QAM (see earlier).

Differential Encoding. Differential encoding for narrowband channels implies that the information is contained in the quotient of two successive modulation symbols. The coherence time (1) of the channel has to be large compared to the symbol duration to assure that the corresponding channel transfer factors H_n and H_{n-1} are approximately equal.

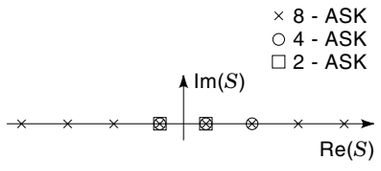

Figure 2. Constellation diagram for M-ASK.

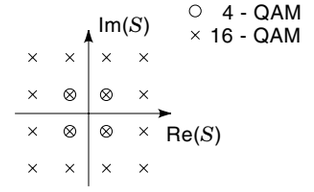

Figure 3. Constellation diagram for M-QAM.

The operation of differential encoding can be described analytically by a simple multiplication:

$$S_n = S_{n-1} B_n \tag{3}$$

In the case of differential phase shift keying modulation (M-DPSK) for example, $B_n \in \{e^{j2\pi m/M} | m = 0, \ldots, M - 1\}$. Unfortunately, M-DPSK has a poor performance if M is large. In this case a combined differential amplitude and phase modulation should be applied, which is discussed in detail later.

Demodulation

In this section it is described how the transmitted data bits can be recovered from the received symbol sequence. This could more exactly be termed hard output demodulation. In the case of channel coding soft decision decoding may be preferable, which means that soft output demodulation is required. This topic is discussed in more detail in the channel coding section.

Depending on the type of modulation (absolute or differential) there are different ways to demodulate the received symbol sequence, which are discussed in detail in the following subsections.

Coherent Demodulation. Coherent demodulation has to be applied if an absolute encoding is used in the transmitter. Originally, the term *coherent* means that the mixer in the receiver is synchronized in frequency and phase with the carrier frequency of the transmitted signal. If any kind of amplitude modulation such as QAM is applied, then the attenuations H_n caused by the channel have to be known, too. To generate this information in the receiver, a channel estimation has to be performed that provides estimates $\hat{H}_n$ for the channel transfer factors. The decision is based on the quotient

$$D_n^{\mathrm{c}} = \frac{R_n}{\hat{H}_n} = S_n + \frac{N_n}{\hat{H}_n} \Rightarrow \hat{S}_n = \mathrm{dec}\{D_n^{\mathrm{c}}\} \tag{4}$$

For D_n^{c} the receiver makes a decision according to given thresholds.

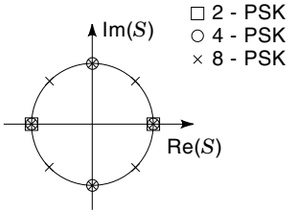

Figure 4. Constellation diagram for M-PSK.

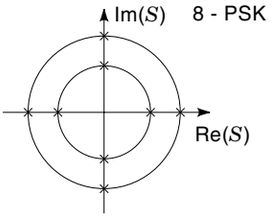

Figure 5. Constellation diagram for M-APSK.

The channel estimation can be performed on the basis of known symbols (pilot symbols) that are included in the signal before transmission. In the first step, the receiver extracts the transfer factors at those times when pilot symbols have been transmitted. Then the transfer factors between the positions of pilot symbols can be interpolated by means of filtering. If the transfer function is filtered in the time direction with a given density of pilot symbols, this requires the maximal Doppler frequency of the channel to be sufficiently small according to the sampling theorem.

Noncoherent Demodulation. If differential encoding is used in the transmitter, the demodulation can be performed either noncoherently (nc) or quasicoherently (qc). With noncoherent demodulation the decision is based on the quotient of two successive symbols

$$D_n^{\mathrm{nc}} = \frac{R_n}{R_{n-1}} = \frac{S_{n-1}B_nH_n + N_n}{S_{n-1}H_{n-1} + N_{n-1}} \tag{5}$$

$$\hat{B}_n = \mathrm{dec}\{D_n^{\mathrm{nc}}\} \tag{6}$$

In general, successive channel-transfer factors are strongly correlated so that $H_n \approx H_{n-1}$ and therefore cancel out in Eq. (5) (if the noise influence is neglected). Unfortunately, the D^{nc} is affected by twice the noise power of D^{c}, leading to a higher bit error rate than coherent demodulation (with perfect channel state information) does. Note that with noncoherent demodulation, no channel estimation has to be performed. Thus, the computation complexity in the receiver is relatively low.

Quasicoherent Demodulation. Quasicoherent demodulation is another way to demodulate differential modulation. Similar to coherent demodulation, the channel influence is first removed before differential decoding takes place. Due to differential encoding and decoding, there is no need to determine the channel phase exactly but only up to an ambiguity of $2\pi/N_{\mathrm{p}}$ rad, where N_{p} denotes the number of phase states. In this case, no pilot symbols are required to estimate the unknown channel-transfer factors (2).

Apart from differential decoding the processing in a quasicoherent receiver is similar to that in a coherent one, that is,

$$\hat{B}_n = \frac{\hat{S}_n}{\hat{S}_{n-1}} = \frac{\mathrm{dec}(R_n/\tilde{H}_n)}{\mathrm{dec}(R_{n-1}/\tilde{H}_{n-1})} = \frac{\mathrm{dec}(D_n^{\mathrm{c}})}{\mathrm{dec}(D_{n-1}^{\mathrm{c}})}$$

In the case of an incorrect decision this error influences two successive symbols due to differential encoding. Therefore, the error probability with quasicoherent demodulation is approximately twice as high as with coherent detection, provided that the error rate is relatively small (<10%). But doubling the error rate with respect to coherent demodulation results in a smaller signal-to-noise ratio loss than noncoherent demodulation does (at least in case of perfect channel estimation).

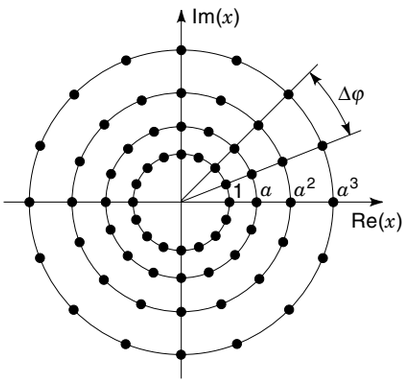

Figure 6. 64-(D)APSK signal space diagram.

Multilevel Differential Modulation

In order to increase the bandwidth efficiency, the well-known M-DPSK can be extended to a differential amplitude and phase shift keying modulation (M-DAPSK), which shows a substantial performance improvement over M-DPSK for $M \geq 16$.

Modulation. DAPSK can be described as a differently encoded APSK, the signal-space diagram of which is defined by the signal set

$$\Psi = \{a^A e^{j\Delta\varphi P} | A \in \{0, \ldots, N_{\mathrm{a}} - 1\}, P \in \{0, \ldots, N_{\mathrm{p}} - 1\}\}$$

where N_{a} denotes the number of amplitudes and N_{p} the number of phases. As an example, a 64-APSK signal space diagram with $N_{\mathrm{a}} = 4$ and $N_{\mathrm{p}} = 16$ is depicted in Fig. 6. Note that the amplitudes are spaced by a factor a. The mapping of the m input bits is done separately for amplitude and phase using m_{a} and m_{p} bits, respectively. To minimize the bit error rate, Gray mapping should be used.

The number of amplitude circles $N_{\mathrm{a}} = 2^{m_{\mathrm{a}}}$, the number of phases per amplitude circle $N_{\mathrm{p}} = 2^{m_{\mathrm{p}}}$, and the amplitude ratio a are free parameters that have to be optimized in the dependency of the number of signal states $M = N_{\mathrm{a}}N_{\mathrm{p}} = 2^m$ and the demodulation method (quasicoherent or noncoherent). The optimized values for the different number of bits per symbol are given in Table 1.

Table 1. Optimal Modulation Parameters for Quasi-coherent and Noncoherent Demodulation

			a	
	N_a	N_p	Quasicoherent Demodulation	Incoherent Demodulation
$M < 16$	1	M	—	—
$M = 16$	2	8	2.0	1.8
$M = 32$	2	16	1.6	1.45
$M = 64$	4	16	1.4	1.38
$M = 128$	4	32	1.3	1.21

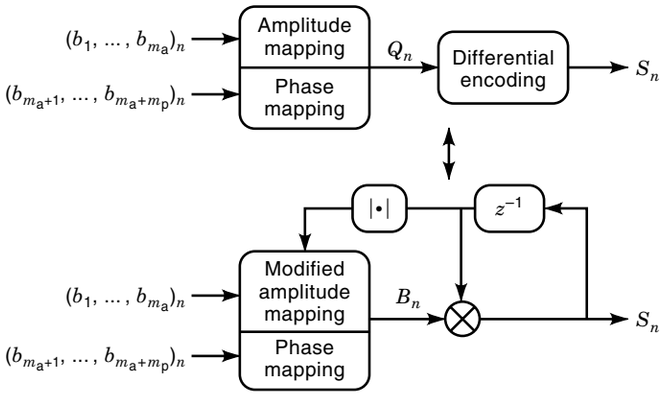

Figure 7. DAPSK modulation.

Differential encoding of the APSK symbols Q_n can be performed as follows:

$$S_n = Q_n \odot S_{n-1} = a^{[A(Q_n)+A(S_{n-1})] \bmod N_a} e^{j(2\pi/N_P)[P/Q_n)+P(S_{n-1})]}$$

To ensure that $S_n \in \Psi$, the amplitude exponents $A(Q_n)$ and $A(S_{n-1})$ are added modulo the number of amplitude circles. This is not explicitly necessary for the phase due to its inherent periodicity. Alternatively, the modulo operation can be integrated into the amplitude mapping (see Fig. 7), so that differential encoding can again be described by a multiplication according to Eq. (3):

$$S_n = B_n S_{n-1}$$

where

$$B_n \in \Psi' = \left\{ a^{A'} e^{j\Delta\varphi P} \left| \begin{matrix} A' \in \{-N_{a+1}, \ldots, N_a - 1\} \\ P \in \{0, \ldots, N_p - 1\} \end{matrix} \right. \right\}$$

As an example, this modified mapping is depicted in Table 2 for $m_a = 2$.

Demodulation. Hard output demodulation of DAPSK signals requires thresholds to be defined for the decision variable D^{nc} (noncoherent demodulation) and D^{c} (quasicoherent demodulation). Since the bits are independently mapping to amplitude and phase, it is sufficient to define corresponding amplitude and phase thresholds. In the case of quasicoherent demodulation with i.i.d. symbols it is well known that the optimum threshold is given by half the distance between two adjacent valid phases or amplitudes:

$$T_p^c = \Delta\varphi/2 + i\Delta\varphi, \qquad i = 0, \ldots, N_p - 1$$
$$T_a^c = a^j \frac{1+a}{2}, \qquad j = 0, \ldots, N_a - 2$$

In the case of noncoherent demodulation the same thresholds are valid for the phase, that is, $T_p^{nc} = T_p^c$ but not for the amplitude. Due to the division of two successively received noise symbols [see Eq. (5)], the exact thresholds are hard to calculate. But there exists a good approximation given by the geometric average of two adjacent valid amplitudes (3):

$$T_a^{nc} \approx a^{j/2}\sqrt{a}, \qquad j = 0, \ldots, N_a - 2$$

Table 2. Modified Amplitude Mapping in the Case of Four Amplitude Circles

		$(b_{a1}b_{a2})_n$			
$\lvert B_n \rvert$		00	01	11	10
$\lvert S_{n-1} \rvert$	1	1	a	a^2	a^3
	a^1	1	a	a^2	$1/a$
	a^2	1	a	$1/a^2$	$1/a$
	a^3	1	$1/a^3$	$1/a^2$	$1/a$

THE OFDM TRANSMISSION TECHNIQUE

In this section we give a brief overview of the history of OFDM and an analytical description of this transmission technique. Finally, we show the technical structure of an OFDM transmission system.

History of OFDM

The basic principles of OFDM have already been proposed in several publications in the 1960s. However, these ideas could not be realized efficiently, since powerful semiconductor devices were not available at that time. Today, even a relatively complex OFDM transmission system with high data rates is technically feasible and such systems can be taken advantage of in frequency-selective channels.

In a classical FDM system narrow-band signals are generated independently, assigned to various frequency bands, transmitted, and separated by filters at the receiver. The new aspect of OFDM is that the various signals are generated jointly by an inverse fast Fourier transform (IFFT) and that their spectra overlap. As a result, generating the signal is simplified and the bandwidth efficiency of the system is improved.

It is interesting to note that as early as 1961 a code division multiplexing (CDM) scheme has been proposed in which sine and cosine functions were used as orthogonal signals (4). The resulting signal can already be compared with an OFDM signal. However, the fact that this system was identical to a frequency division multiplex was not important for the proposal, and the benefits in frequency-selective channel were not recognized.

Since 1966 FDM systems with overlapping spectra were proposed in several publications (5–7). The next step was the proposal to realize an FDM system with the discrete Fourier transform (DFT) (8). Finally, in 1971 Weinstein and Ebert proposed a complete OFDM system (9) that included generating the signal with an FFT and adding a guard interval in the case of multipath channels. This system will be referred to in the following.

In the further development, OFDM has been discussed for channels with frequency-nonselective fading in Ref. 10. In Ref. 11 it has been proposed for broadcast applications for mobile reception. Meanwhile, OFDM systems have been standardized in the DAB (Digital Audio Broadcasting) and DTVB (Digital Terrestrial Video Broadcasting) projects.

The most important advantage of the OFDM transmission technique as compared to single-carrier systems is obtained in frequency-selective channels. The signal processing in the receiver is rather simple in this case, because after transmission over the radio channel the orthogonality of the OFDM subcarriers is maintained and the effect of the channel is reduced to a multiplication of each subcarrier by a complex transfer factor. Therefore, equalizing the signal is very simple, whereas equalization may not be feasible in the case of single-carrier transmission with the same bandwidth.

Mathematical Description

Transmitter. An OFDM signal consists of K subcarriers spaced by the frequency distance Δf. Thus, the total system bandwidth B is divided into N equidistant subchannels. All subcarriers are mutually orthogonal within a time interval of length $T_S = 1/\Delta f$. The kth subcarrier signal is described analytically by the function $g_k(t)$, $k = 0, \ldots, K-1$ (a rectangular pulse shaping is assumed:

$$g_k(t) = \begin{cases} e^{j2\pi k \Delta f t} & \forall t \in [0, T_S] \\ 0 & \forall t \notin [0, T_S] \end{cases}$$

Since the system bandwidth B is subdivided into K narrowband subchannels, the OFDM block duration T_S is K times as large as in the case of a single-carrier transmission system covering the same bandwidth. This subcarrier signal $g_k(t)$ is extended by a cyclic prefix (called guard interval) with the length T_G, yielding the following signal:

$$g_k(t) = \begin{cases} e^{j2\pi k \Delta f t} & \forall t \in [-T_G, T_S] \\ 0 & \forall t \notin [-T_G, T_S] \end{cases}$$

The guard interval is added to the subcarrier signal in order to avoid intersymbol interferences (ISI) that occur in multipath channels. At each receiver the guard interval is removed and only the time interval $[0, T_S]$ is evaluated. From this point of view the guard interval is a pure system overhead. The total OFDM block duration is $T = T_S] + T_G]$.

It is an important advantage of the OFDM transmission technique that ISI, which occur in all multipath channels, can be reduced considerably. If the guard interval length T_G is larger than the maximal delay in the radio channel, no ISI occur at all and the orthogonality of the subcarriers is not affected.

Each subcarrier can be modulated independently with the complex modulation symbol $S_{n,k}$, $k = 0, 1, \ldots, K-1$ so that within the symbol duration T the following signal of the nth OFDM block is formed:

$$s_n(t) = \frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} S_{n,k} g_k(t - nT)$$

The total continuous-time signal consisting of all OFDM blocks is

$$s(t) = \frac{1}{\sqrt{K}} \sum_{n=0}^{\infty} \sum_{k=0}^{K-1} S_{n,k} g_k(t - nT)$$

Due to the rectangular pulse shaping of the signal, the spectra of the subcarriers are sinc functions, for example, for the kth subcarrier:

$$G_k(f) = T \mathrm{sinc}(\pi T(f - k\Delta f))$$

where $\mathrm{sinc}(x) = \sin(x)/x$. The spectra of the subcarriers overlap, but the subcarrier signals are mutually orthogonal and the modulation symbols $S_{n,k}$ can be recovered by a correlation:

$$\begin{aligned} \langle g_k, g_t \rangle &= \int_0^{T_S} g_k(t) \overline{g_t(t)}\, dt \\ &= T_S \delta_{k,l} \\ S_{n,k} &= \frac{\sqrt{K}}{T_S} \langle s_n(t), \overline{g_k(t - nT)} \rangle \end{aligned}$$

In a practical application the OFDM signal $s_n(t)$ is generated in a first step as a discrete-time signal in the digital signal processing part of the transmitter. As the bandwidth of an OFDM system is $B = K\,\Delta f$, the signal must be sampled with the sampling time $\Delta t = 1/B = 1/K\,\Delta f$. The samples of the signal are written as $s_{n,i}$, $i = 0, 1, \ldots, K-1$ and can be calculated as

$$s_{n,i} = \frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} S_{n,k} e^{j2\pi ik/N}$$

This equation describes exactly the inverse discrete Fourier transform (IDFT) which is typically realized by an inverse FFT (IFFT).

Receiver. The subcarrier orthogonality is not affected at the output of the radio channel if the guard interval length T_G is larger than the maximal multipath delay. Therefore, the received signal $r_n(t)$ can be separated into the orthogonal subcarrier signals by a correlation according to

$$R_{n,k} = \frac{\sqrt{K}}{T_S} \langle r_n(t), \overline{g_k(t - nT)} \rangle \tag{7}$$

The correlation at the receiver can be realized as a discrete Fourier transform (DFT) or an FFT, respectively:

$$R_{n,k} = \frac{1}{\sqrt{K}} \sum_{i=0}^{K-1} r_{n,i} e^{-j2\pi ik/K}$$

Here $r_{n,i}$ is the ith sample of the received signal $r_n(t)$ and $R_{n,k}$ is the received complex symbol of the kth subcarrier. The FFT and IFFT algorithms can be implemented very efficiently (12).

If the subcarrier spacing Δf is chosen much smaller than the coherence bandwidth and the symbol duration T much smaller than the coherence time of the channel, then the transfer function of the radio channel can be considered constant within the bandwidth Δf of each subcarrier and the duration of each modulation symbol. In this case, the effect of the radio channel is only a multiplication of each subcarrier

signal $g_k(t)$ by a complex transfer factor $H_{n,k}$. As a result, the received complex symbol after the FFT is

$$R_{n,k} = H_{n,k} S_{n,k} + N_{n,k} \tag{8}$$

where $N_{n,k}$ is additive noise of the channel.

Comparing OFDM systems with single-carrier transmission it is obvious that the simplicity of channel equalization is an important advantage of OFDM. However, there are also some difficulties connected with OFDM. First, the signal can have very-high-amplitude peaks that are limited by a nonlinear power amplifier in the transmitter. In this case the orthogonality of subcarriers is disturbed and interferences are produced both in the OFDM band and in adjacent channels. Therefore, the power amplifier is operated at an input backoff factor. These topics have been analyzed in Refs. 13 and 14.

Second, synchronization must be much more exact in an OFDM system because a frequency offset between transmitter and receiver affects the subcarrier orthogonality. It has been shown for an uncoded system that the maximum acceptable frequency offset is lower than 3% of the subcarrier spacing (15).

CHANNEL CODING ASPECTS

As it has already been explained, in a narrow-band channel the modulation symbols are attenuated by a complex transfer factor H_n. If the channel is a multipath channel with many propagation paths and without a line-of-sight (LOS) path, then the amplitude of the transfer factors is Rayleigh distributed according to the central limit theorem (16).

This means that even at a very large average signal-to-noise ratio (SNR) the situation of a deep fade (i.e., a small value $|H_n|$) may occur, which leads to bit errors. For this reason, the use of channel coding is a very important topic. Without channel coding the typical flat fading curve is obtained for the bit error rate (BER). With channel coding, large SNR gains are achieved. The code provides diversity in this situation. The additional gain that can be realized with soft decision instead of hard decision decoding is very large.

Therefore, convolutional codes are a reasonable choice, because soft-decision decoding can easily be performed by the Viterbi algorithm, which is usually applied in order to decode convolutional codes. Furthermore, the code rate of convolutional codes can be adjusted in a very flexible way by puncturing the code. This can be applied in order to adapt the error correcting capability of the code to the channel state.

Soft-Output Demodulation

In order to apply soft-decision decoding, the decoder needs metric increments for the received modulation symbols instead of ready decisions. This means that the demodulator has to provide soft outputs λ_n^ν for each symbol R_n, where the superscript ν enumerates the (possibly transmitted) modulation symbols in the set Ψ.

Coherent Detection. According to Eq. (2), the received modulation symbol in the nth time interval is

$$R_n = H_n S_n + N_n$$

where S_n is the transmitted symbol, H_n is the complex channel transfer factor, and N_n is additive white Gaussian noise with the noise power $2\sigma^2$. The probability density function (PDF) of R_n is a complex Gaussian distribution:

$$p(R_n|S_n) = \frac{1}{2\pi\sigma^2} e^{-|R_n - H_n S_n|^2/2\sigma^2}$$

A maximum likelihood sequence estimator would have to choose one out of all possibly transmitted symbol sequences μ:

$$\langle S_n \rangle^{(\mu)} = \langle S_n(\mu) \rangle$$

This leads to the estimate $\widehat{\langle S_n \rangle}$:

$$\widehat{\langle S_n \rangle} = \arg\max P(\langle R_n \rangle | \langle S_n \rangle^{(\mu)}) \tag{9}$$

$$= \arg\max_\mu \prod_n p(R_n|S_n(\mu)) \tag{10}$$

$$= \arg\min_\mu \sum_n |R_n - H_n S_n(\mu)|^2 \tag{11}$$

The summands in Eq. (11) are calculated by the demodulator for each possibly transmitted modulation symbol S^ν and passed to the decoder as metric increments

$$\lambda_n^\nu = |R_n - H_n S^\nu|^2$$

for further evaluation. If only metric increments for each bit are required, these could also be computed by the demodulator. The decoder estimates the symbol sequence or the corresponding bit sequence on the basis of the given metric increments according to Eq. (11).

This metric, consisting of squared Euclidean distances as metric increments, can be used with all coherent modulation techniques because it is only based on the assumptions that the transmitted modulation states are statistically independent and that the received modulation states are Gaussian distributed due to the AWGN.

Noncoherent Detection. In the case of differential modulation techniques we can calculate soft outputs in a very similar way. We choose decision variables having a PDF that can be approximated by a Gaussian distribution and determine the variance of this distribution. Then we use a metric consisting of squared distances with these appropriate decision variables weighted according to the determined variance.

For higher level M-DPSK ($M \geq 8$) the PDF of the phase difference φ_n of successive symbols as the usual decision variable can already be approximated by a Gaussian distribution:

$$p(\varphi_n|\psi_n) \approx \frac{1}{\sqrt{2\pi\sigma_\varphi^2}} e^{-|\varphi_n - \psi_n|^2/2\sigma_\varphi^2}$$

where ψ_n is the transmitted and φ_n the received phase difference. This results in the following decision criterion of the Viterbi algorithm for M-DPSK signals:

$$\widehat{\langle \psi_n \rangle} = \arg\min_\mu \sum_n \left|H_n\right|^2 \left|\varphi_n - \psi_n(\mu)\right|^2$$

Differential amplitude and phase shift keying (DAPSK) consists of a DPSK and a differential amplitude modulation. The amplitude information is contained in the quotient B_n of the amplitudes of two successive modulation symbols, which is the usual decision variable. For the purpose of generating a metric for the Viterbi algorithm, we use the new decision variable W_n instead of B_n. W_n is approximately Gaussian distributed again:

$$W_n = \ln\left|\frac{R_n}{R_{n-1}}\right|, \qquad V_n = \ln\left|\frac{S_n}{S_{n-1}}\right|$$

$$p(W_n|V_n) \approx \frac{1}{\sqrt{2\pi\sigma_v^2}} e^{-|W_n - V_n|^2/2\sigma_w^2}$$

$$\sigma_w^2 = \sigma_\varphi^2 = \frac{\sigma^2}{|H_n S_n|^2} + \frac{\sigma^2}{|H_{n-1} S_{n-1}|^2}$$

By using the decision variables $W_{n,k}$ and $\varphi_{n,k}$ a metric with squared distances can also be given for DAPSK modulation. The transmitted symbols B_n are described by V_n and ψ_n:

$$\langle\widehat{B_n}\rangle = \arg\min_\mu \sum_n d_n^2(\mu) RI_n^2 \tag{12}$$

$$d_n(\mu) = [W_n - V_n(\mu)]^2 + [\varphi_n - \psi_n(\mu)]^2 \tag{13}$$

$$RI_n^2 = \frac{1}{\left|1/R_n\right|^2 + \left|1/R_{n-1}\right|^2} \tag{14}$$

Again each summand in Eq. (12) is the metric increment for the corresponding symbol B_n. For all possible symbols $B_{n,k}^\nu$, metric increments can be calculated as

$$\lambda_n^\upsilon = (d_n^\upsilon)^2 RI_n^2$$

by the demodulator. Through the term d_n^ν the position of the received symbols in the constellation diagram is considered. RI_n is information about the reliability of the subcarrier (*reliability information*).

Quasicoherent Detection. In the case of quasicoherent detection, the two successively received modulation symbols are evaluated separately in order to calculate the metric increments. The Viterbi decoder uses the following metric:

$$\langle\widehat{B_n}\rangle = \arg\max_\mu \prod_n p(R_n, R_{n-1}|B_n(\mu))$$
$$\approx \arg\min_\mu \lambda_n(\mu)$$

with the metric increments

$$\lambda_n(\mu) = \min_{S^\kappa/S^\rho = B_n(\mu)} (|R_n - H_n S^\kappa|^2 + |R_{n-1} H_{n-1} S^\rho|^2)$$

Again, metric increments for all possible B_n^ν can be calculated by the demodulator and passed to the decoder.

Soft-Decision Decoding

As it has just been described, metric increments λ_n^ν are calculated for each modulation symbol S_n^ν or B_n^ν by the demodulator. The task of the decoder is to decide for a bit sequence $\langle b_{n,j}\rangle$ which leads to the minimal metric. The subscript j enumerates the bits that are transmitted by a modulation symbol. If the Viterbi algorithm is to be used for decoding, metric increments $\lambda_{n,j}$ for each bit are required instead of metric increments for each symbol in order to make the following decision:

$$\langle b_{n,j}\rangle = \arg\max_\mu P(\langle R_n\rangle|\langle b_{n,j}\rangle^{(\mu)}) \tag{15}$$

$$\approx \arg\min_\mu \sum_j \lambda_{n,j}(\mu) \tag{16}$$

A metric increment $\lambda_{n,j}^\kappa$ for the assumption that $b_{n,j} = \kappa \in \{-1, 1\}$ can be obtained by

$$\lambda_{n,j}^\kappa = \arg\min_{S_n^\upsilon | b_{n,j} = \kappa} \lambda_n^\upsilon \tag{17}$$

In the case of differential modulation B_n^ν replaces S_n^ν in Eq. (17). The Viterbi decoder chooses the sequence of bits $\langle b_{n,j}\rangle$ that minimizes the sum in Eq. (16). For all bits $b_{n,j}$ of a subcarrier k, the corresponding metric increments are determined on the basis of the modulation symbol with the highest conditional probability $P(R_n|S_n)$ and by which the assumed bit (-1 or 1) is transferred.

This way of calculating and using a bit metric for the Viterbi algorithm is an approximation. For an optimal decision the metric increments for all bits in the bit sequence that are associated with the same subcarrier would have to refer to the same modulation symbol.

Multilevel Coding

In the discussion above, modulation and channel coding have been considered separately. A joint consideration of coding and modulation can achieve a better performance by taking advantage of the fact that different error events occur with very different probabilities. A very general approach of combining coding and modulation is multilevel coding (17).

If a modulation scheme with m bits per symbol is used, the data bits are encoded into m sequences of l bits by m different encoders. These bit sequences are mapped to l modulation symbols.

The set of all modulation symbols is partitioned into two subsets such that the error probability within each subset is minimized. In the case of nondifferential modulation this means that the Euclidean distance of the modulation symbols in each subset must be maximized. Subsequently, each of the subsets is recursively partitioned again until the subsets contain only a single modulation symbol. A well-known example for this set partitioning is given in Fig. 8 for 8-PSK.

Based on this set partitioning, the data bits are mapped to modulation symbols in the following way. One data bit is taken from each of the m coded sequences. The first data bit selects one of the two subsets of the first partitioning level. Each of the following data bits selects one of the two subsets of the next partitioning level until the modulation symbol to be transmitted is defined after m selections.

The receiver evaluates the transmitted modulation symbols and subsequently determines the coded bit sequences. Provided that the receiver is able to decode the first bit sequence that corresponds to the first partitioning level without errors, this information can be used in order to evaluate the

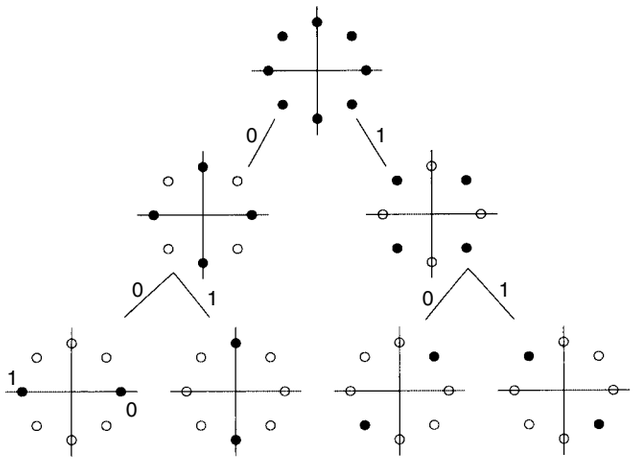

Figure 8. Set partitioning for 8-PSK.

second partitioning level and so on. As the error probability within the subsets decreases from level to level, the codes need less redundancy in order to correct errors.

For differential modulation, the set of all possible quotients $B\pi = S_n/S_{n-1}$ has to be partitioned so that the error probability within each subset is maximized. In the case of M-DPSK this leads to the same set partitioning that is known for M-PSK. For M-DAPSK the set partitioning is discussed in the following.

We assume that the M-DAPSK symbol B_1 with logarithm of amplitude V_1 and phase difference ψ_1 is transmitted, that the symbol B_2 with V_2 and ψ_2 belongs to the same subset as B_1, and that the decision variables that the receiver calculates from the received symbol B' are $W = V_1 + \Delta V$ and $\varphi_1 + \Delta\psi$.

The probability density functions of W and φ are approximately Gaussian as it has been shown earlier. The approximated joint PDF of W and φ is

$$p(W, \varphi|V_1, \psi_1) \approx \frac{1}{2\pi\sigma_w^2} e^{-(1/2\sigma_w^2)(|W-V_1|^2+|\varphi-\psi_1|^2)}$$
$$= \frac{1}{2\pi\sigma_w^2} e^{-(1/2\pi\sigma_w^2)(|\Delta V|^2+|\Delta\psi|^2)}$$

The receiver decides incorrectly if

$$p(W, \varphi|V_2, \psi_2) > p(W, \varphi|V_1, \psi_1)$$
$$\Rightarrow (W - V_2)^2 + (\varphi - \psi_2)^2 < (W - V_1)^2 + (\varphi - \psi_1)^2$$

The decision variables W and φ make the situation very similar to the partitioning of a nondifferential modulation scheme like M-QAM where the Euclidean distance within the subsets is maximized. Here, the error probability within a subset is

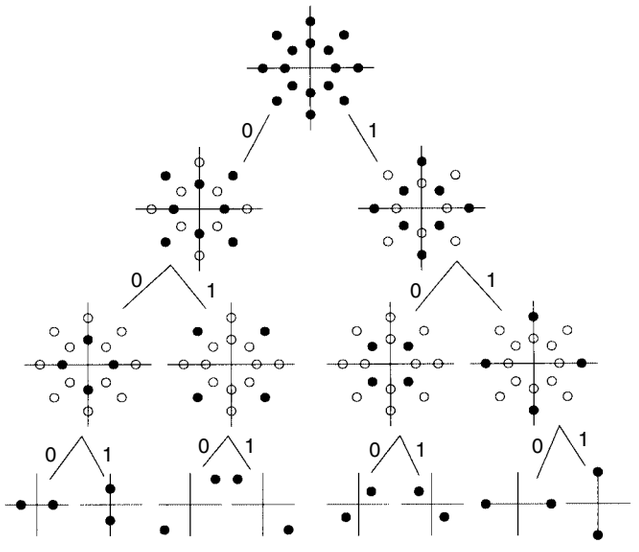

Figure 9. Set partitioning for 16-DAPSK.

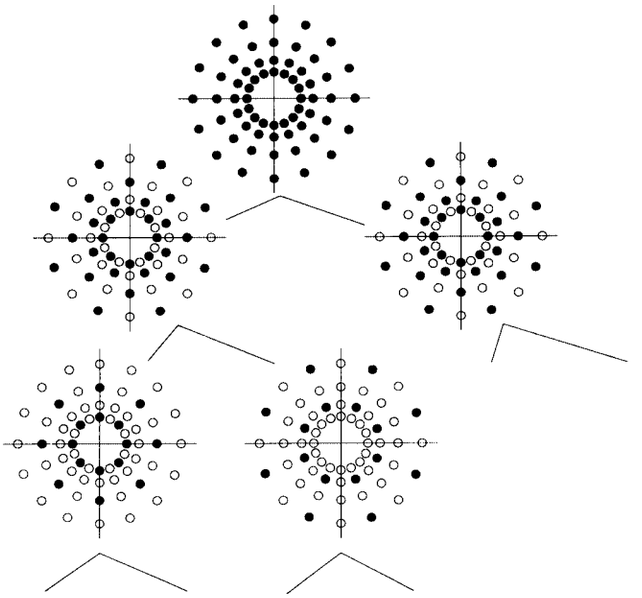

Figure 10. Set partitioning for 32-DAPSK.

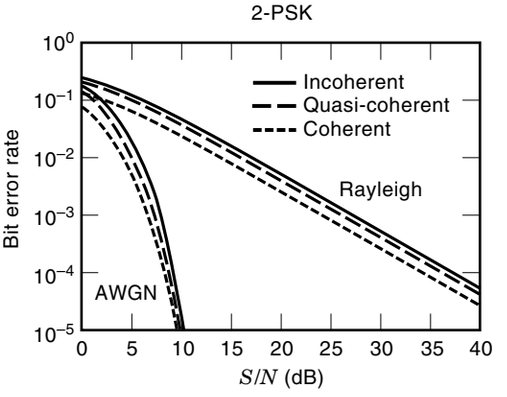

Figure 11. Set partitioning for 64-DAPSK.

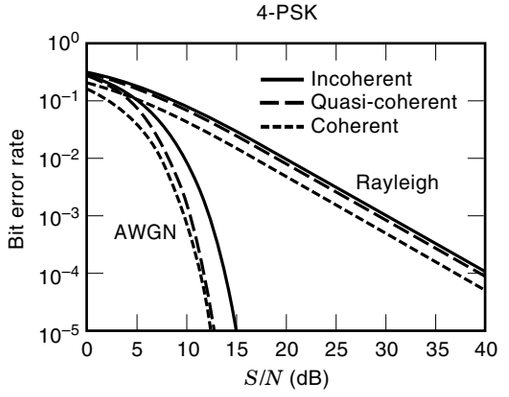

Figure 12. BER performance of 2-(D)PSK. (AWGN and Rayleigh fading channel)

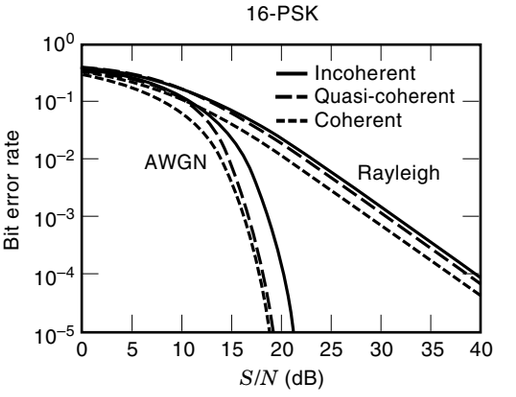

Figure 14. BER performance of 8-(D)PSK. (AWGN and Rayleigh fading channel)

minimized if the smallest distance $d_{i,j}$ between two symbols B_i and B_j in the subset is maximized:

$$d_{ij}^2 = (V_i - V_j)^2 + (\psi_i - \psi_j)^2$$

The distance measure $d_{i,j}$ is nothing else but a Euclidean distance referring to the coordinates V and ψ. The optimal set partitioning is given for 16-, 32-, and 64-DAPSK in Figs. 9 to 11. The partitioning diagrams refer to the Q symbols that represent the information to be transmitted.

PERFORMANCE

In Figs. 12 to 18, the performance of various amplitude modulation schemes is presented in terms of BER versus the signal-to-noise-ratio. The performance is analyzed in two transmission channels, namely the AWGN channel and the Rayleigh fading channel.

The results have been obtained on the basis of simulations, assuming a Gray mapping of the bits to the modulation symbols, that is, the assigned bit patterns of adjacent symbols differ by only one bit. Closed analytical forms of the bit error rate p_b for some of the depicted curves can be found in the literature (e.g., Ref. 1).

For example, the BER for coherently detected 2-PSK and 4-PSK in an AWGN channel is given by

$$p_b = \tfrac{1}{2}\mathrm{erfc}(\sqrt{\gamma_b})$$

where $\gamma_b = E_b/N_0 = (1/m)(S/N)$ (matched filter reception: $m = \log_2(M)$, which is the number of bits per symbol). For coherently detected 8-PSK, the corresponding expression is derived in Ref. 18.

For large values of M and $\gamma_b \gg 1$, an approximation for the BER of coherently detected M-PSK can be derived (1):

$$p_b(M) \approx \frac{1}{\log_2(M)}\,\mathrm{erfc}\left(\sqrt{ld(M)\gamma_b}\sin\frac{\pi}{M}\right)$$

As to noncoherently detected 2-DPSK and 4-DPSK, the BER is expressed by

$$p_b = \tfrac{1}{2}e^{-\gamma_b}$$

and

$$p_b = Q(a,b) - \tfrac{1}{2}I_0(ab)e^{-(a^2+b^2)/2}$$

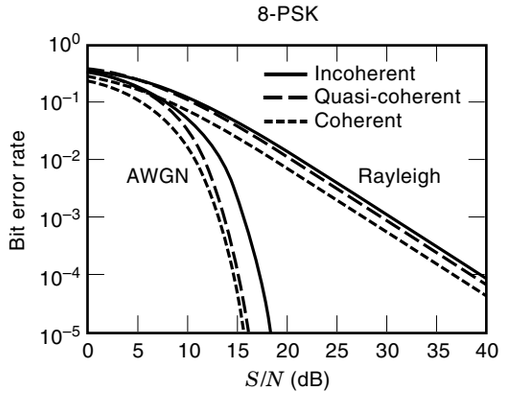

Figure 13. BER performance of 4-(D)PSK. (AWGN and Rayleigh fading channel)

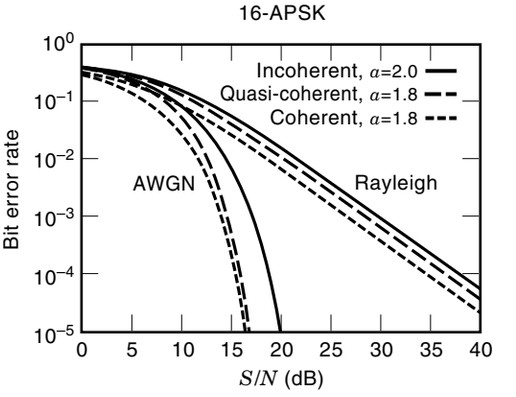

Figure 15. BER performance of 16-(D)PSK. (AWGN and Rayleigh fading channel)

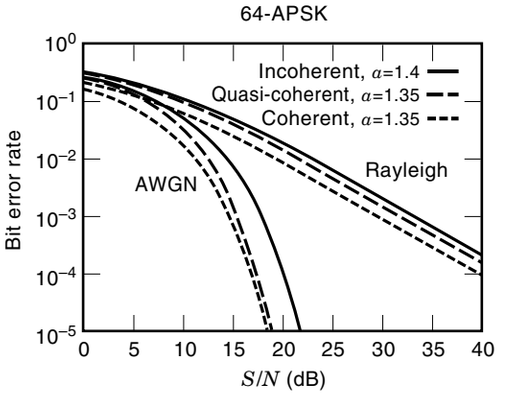

Figure 16. BER performance of 16-(D)APSK. (AWGN and Rayleigh fading channel)

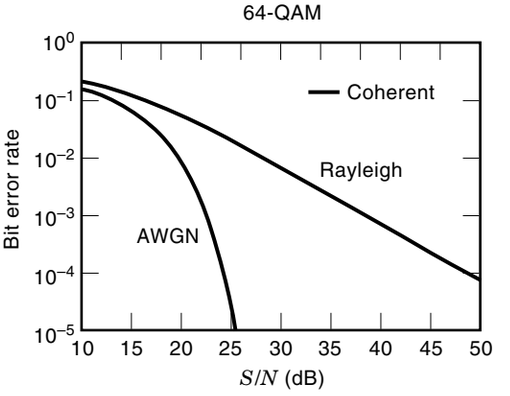

Figure 18. BER performance of 16-QAM. (AWGN and Rayleigh fading channel)

respectively, where

$$a = \sqrt{2\gamma_b\left(1 - \frac{1}{\sqrt{2}}\right)}$$

$$b = \sqrt{2\gamma_b\left(1 + \frac{1}{\sqrt{2}}\right)}$$

$$Q(a,b)e^{-(a^2+b^2)/2}\sum_{k=0}^{\infty}\left(\frac{a}{b}\right)^k I_k(ab)$$

and $I_k(x)$ is the kth-order modified Bessel function of the first kind.

In general, for noncoherent detection results from the fact that the decision variable is formed by the quotient of two (noisy) modulation symbols, hence the noise power is doubled. On the other hand, the computation complexity for coherent reception is higher due to the need of a channel estimation and equalization. Furthermore, a realistic (imperfect) channel estimation would introduce additional errors.

As far as quasicoherent detection is concerned, corresponding error rates can be derived from the equations for coherent reception, based on the approximation that the BER of quasi-coherent detection is twice as high as with coherent detection (due to differential encoding).

In the case of coherent M-QAM modulation, an approximation of the BER is presented in Ref. 19:

$$p_b = 2\frac{\left(1 - \frac{1}{\sqrt{M}}\right)}{\log_2 M}\,\mathrm{erfc}\left(\sqrt{\frac{3\log_2\sqrt{M}}{M-1}\gamma_b}\right)$$

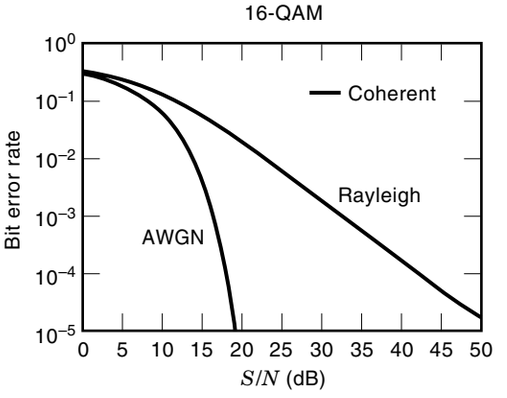

Figure 17. BER performance of 64-(D)APSK. (AWGN and Rayleigh fading channel)

BIBLIOGRAPHY

1. J. G. Proakis, *Digital Communications,* 3rd ed., New York: McGraw-Hill, 1995.
2. A. J. Viterbi and A. M. Viterbi, Nonlinear estimation of PSK-modulated carrier phase with application to burst digital transmission, *IEEE Trans. Commun.,* **IT-29**: 543–551, 1983.
3. V. Engels and H. Rohling, Multilevel differential modulation techniques (64-DAPSK) for multicarrier transmission systems, *Eur. Trans. Telecommun.,* **6** (6): 633–640, 1995.
4. G. A. Franco and G. Lachs, An orthogonal coding technique for communications, *IRE Int. Conv. Rec.,* **9**: Pt. 8: 126–133, 1961.
5. R. W. Chang, Synthesis of band-limited orthogonal signals for multichannel data transmission, *Bell Sys.Tech. J.,* **45**: 1775–1796, 1966.
6. B. Saltzberg, Performance of an efficient parallel data transmission system, *IEEE Trans. Commun.,* **15**: 805–811, 1967.
7. M. S. Zimmerman and A. L. Kirsch, The AN/GSC-10 (KATHRYN) variable rate data modem for HF radio, *IEEE Trans. Commun.,* **15**: 197–203, 1967.
8. S. Darlington, On digital single-sideband modulators, *IEEE Trans. Circuit Theory,* **17**: 409–414, 1970.
9. S. B. Weinstein and P. M. Ebert, Data transmission by frequency-division multiplexing using the discrete Fourier transform, *IEEE Trans. Commun.,* **19**: 628–634, 1971.
10. L. J. Cimini, Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing, *IEEE Trans. Commun.,* **33**: 665–675, 1985.
11. M. Alard and R. Lassalle, Principles of Modulation and Channel Coding for Digital Broadcasting for Mobile Receivers, *EBU Techn. Rev.* **224**: August 1987, pp. 168–190.
12. A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing,* Englewood Cliffs, NJ: Prentice-Hall, 1989.

13. A. E. Jones and T. A. Wilkinson, Combined coding for error control and increased robustness to system nonlinearities in OFDM, *IEEE 46th Veh. Technol. Conf.,* 904–908, 1996.
14. S. H. Mller et al., OFDM with reduced peak-to-average power ratio by multiple signal representation, *Ann. Telecommun.,* **52**: 58–67, 1997.
15. M. Aldinger, Multicarrier COFDM scheme in high bitrate radio local area networks, *Proc. Wireless Comput. Networks '94,* Sept. 1994, pp. 969–973.
16. A. Papoulis, *Probability, Random Variables, and Stochastic Processes,* 3rd ed., New York: McGraw-Hill, 1991.
17. H. Imai and S. Hirakawa, A new multilevel coding method using error-correcting codes, *IEEE Trans. Inf. Theory,* **23**: 371–377, 1977.
18. P. J. Lee, Computation of the Bit Error Rate of Coherent M-ary PSK with Gray Code Bit Mapping, *IEEE Trans. Commun.,* **34**: 488–491, 1986.
19. B. Sklar, *Digital Communications,* Englewood Cliffs, NJ: Prentice-Hall, 1988.

Reading List

A. C. Bingham, Multicarrier modulation for data transmission: An idea whose time has come, *IEEE Commun. Mag.,* **28** (5): 5–14, May 1990.

M. Bossert and A. Donder, Channel estimation and equalization in orthogonal frequency division multiplexing systems, *Proc. VDE Conf., ITG Fachberichte 135,* Neu-Ulm, 485–492, 1995.

Y. C. Chow, A. R. Nix, and J. P. McGeehan, Analysis of 16-APSK Modulation in AWGN and Rayleigh Fading Channel, *Electron. Lett.,* **28** (17): 1608–1610, 1992.

Digital Land Mobile Radio Communications, Final report 14 March 1984–13 September 1988, COST 207, Luxembourg: Office for Publications of the European Communities, 1989, ISBN 92-825-9946-9.

P. Höher, TCM on frequency-selective land-mobile fading channels. In M. Luise and E. Biglieri (eds.), *Proc. 5th Int. Workshop Digital Commun.,* Pisa, Italy, 1991.

P. Höher, A statistical discrete-time model for the WSSUS multipath channel, *IEEE Trans. Veh. Technol.,* **41**: 461–468, 1992.

T. Lauterbach and M. Unbehaun, Multimedia environments for mobiles (MEMO)—Interactive multimedia services to portable and mobile terminals, *Proc. ACTS Mobile Commun. Summit,* Aalborg, Denmark, October 1997.

H. Rohling and V. Engels, Differential amplitude phase shift keying (DAPSK)—A new modulation method for DTVB, *Int. Broadcasting Conv.,* Amsterdam, 1995, pp. 102–108.

H. Rohling and R. Grünheid, OFDM transmission technique with flexible subcarrier allocation, *Proc. ICT 1996,* pp. 567–571.

M. Sandell, *Design and analysis of estimators for multicarrier modulation and ultrasonic imaging,* Ph.D. thesis, Lulea Univ. of Technology, Lulea, Sweden, 1996.

S. G. Wilson, J. Freebersyser, and C. Marshall, Multi-symbol detection of MPSK, *Proc. GLOBECOM '89,* Dallas, Texas, 1989, pp. 1692–1697.

Y. Yasuda, K. Kashiki, and Y. Hirata, High-rate punctured convolutional codes for soft decision Viterbi decoding, *IEEE Trans. Commun.,* **32**: 315–319, 1984.

HERMANN ROHLING
RAINER GRÜNHEID
THOMAS MAY
KARSTEN BRÜNINGHAUS
Technical University Braunschweig