INVITED PAPER

An Introduction to Adaptive QAM Modulation Schemes for Known and Predicted Channels

Relatively simple Quadrature Amplitude Modulation (QAM) systems illustrate the advantages, in error performance and spectral efficiency, of adaptive modulation over fixed modulation.

By Arne Svensson, *Fellow IEEE*

ABSTRACT | A major disadvantage with fixed modulation (nonadaptive) on channels with varying signal-to-noise ratio (SNR) is that the bit-error-rate (BER) probability performance is changing with the channel quality. Most applications require a certain maximum BER and there is normally no reason for providing a smaller BER than required. An adaptive modulation scheme, on the contrary, can be designed to have a BER which is constant for all channel SNRs. The spectral efficiency of the fixed modulation is constant, while it, in general, will increase with increasing channel SNRs for the adaptive scheme. This in effect means that the average spectral efficiency of the adaptive scheme is improved, while at the same time the BER is better suited to the requirement of the application. Thus, the adaptive link becomes much more efficient for data transmission. The major disadvantage is that the transmitter needs to know the channel SNR such that the best suitable modulation is chosen and the receiver must be informed on the used modulation in order to decode the information. This leads to an increased overhead in the system as compared with a fixed modulation system. In this paper, we introduce adaptive modulation systems by presenting some of the simpler adaptive quadrature amplitude modulation schemes and their performance for both perfectly known and predicted channels.

KEYWORDS | Adaptive modulation; channel prediction; flat fading channel; quadrature amplitude modulation (QAM)

Manuscript received December 11, 2006; revised April 21, 2007.
The author is with the Department of Signals and Systems, Chalmers University of Technology, SE-412 96 Göteborg, Sweden (e-mail: arnes@chalmers.se).
Digital Object Identifier: 10.1109/JPROC.2007.904442

I. INTRODUCTION

Adaptive modulation is a method to improve the spectral efficiency of a radio link for a given maximum required quality (error probability). The idea of adapting the modulation and coding to the channel conditions is not at all new; it has been mentioned in numerous papers at least since the 1970s.[1] It is, however, not until much later that optimum schemes for this purpose became available. Many papers on good schemes started to appear in the middle of the 1990s.

The purpose of this paper is to introduce the reader to the topic of adaptive modulation to get an understanding of the differences between fixed and adaptive modulations schemes. We are especially focusing on illustrating the big advantage in both error performance and spectral efficiency of adaptive schemes compared with fixed schemes on varying channels. This is done by describing some of the simpler adaptive quadrature amplitude modulation (QAM) schemes, when the channel is perfectly known in the transmitter and when the predicted channel is available in the transmitter. For simplicity, only the most simple spectrally flat channels are considered, where one parameter alone describes the channel. The approach taken is to describe, in some detail, some of the schemes for perfectly known channels published in [1] and some of the schemes for predicted channels in [2]. The reason for choosing these schemes is that they are reasonably simple schemes which are optimized for different criteria. Moreover, they illustrate the design rules and performance of adaptive schemes in a simple

[1] It is not known to this author when this idea was published for the first time and, therefore, we will not give any citation here.

and illustrative way. Some examples of other contributions to adaptive modulation are presented in [3]–[16]. There are many other studies on various topics related to adaptive modulation schemes published in the literature. To list all of these contributions is outside the scope of this introductory paper on adaptive modulation. The interested reader is referred to other papers in this Special Issue, which all taken together should give a rather complete picture on adaptive modulation and transmission schemes, and to the open literature.

The rest of this paper is organized in the following way. In Section II, some basic background on fixed modulations and their performance on simple channels are given. Here, also the topic of adaptive modulation (and coding) is introduced. Then, in Section III, some of the optimum adaptive schemes designed for a known channel from [1] are introduced and their performance is given and compared with the performance of fixed modulations. A similar description of some optimum adaptive schemes for predicted channels is given in Section IV. Channel coding is used in fixed modulations to improve the power efficiency, and this can de done also in adaptive modulations as briefly discussed in Section V. Finally, some concluding remarks are given in Section VI.

II. BACKGROUND ON FIXED MODULATIONS AND INTRODUCTION TO ADAPTIVE MODULATIONS

Many different fixed modulation methods have been designed for various channels and applications [17]–[24]. A modulation method is used to carry digital information over a channel.[2] Since different channels have different properties, the modulation method must convey the information in a form suitable for the particular channel in mind. This is typically done by assigning a waveform to each possible transmitted symbol. Then, the waveform is transmitted over the channel and in the receiver a detector is used to find which of the possible waveforms was transmitted. The frequency spectrum required to transmit the signal depends on the correlation properties of the sequence of symbols to be transmitted and the set of waveforms used to convey the information. More details on this can be found in the textbooks cited above.

Since all practical wireless channels change the transmitted waveforms by at least adding noise and interference, the detection process is never free from making errors. Many channels change the transmitted waveform in more complicated ways. The simplest form of alteration due to a channel is commonly modeled as a

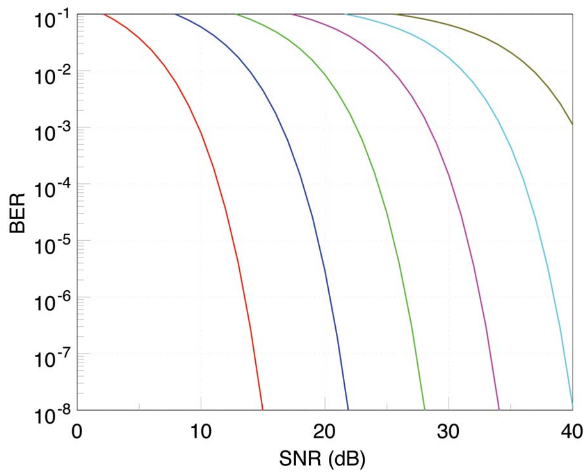

Fig. 1. *BER versus SNR per symbol for Gray coded QAMs. The curves from left to right correspond to 2, 4, 6, 8, 10, and 12 bits per symbol.*

multiplication by a fixed attenuation and a phase shift of the received carrier,[3] while more complex forms of fading typically are modeled by a filter. Given a modulation method and a channel, the detector can be designed in many different ways. The optimum detector is supposed to find the most likely transmitted symbol or message given a received signal. This detector might be very complex to implement and for that reason many less complex suboptimum detection schemes have been devised. Different detection schemes are, however, outside the scope of this paper and the interested reader is referred to [17]–[24] and references therein.

A. Fixed Modulation in Noise

In summary, a fixed modulation method carries a given number of bits per symbol over a channel and the detector detects the bits (or symbols) with a given bit (or symbol) error probability. The actual bandwidth required to transmit the modulation without distortion also depends on the set of waveforms used, but, in this paper, we will not go into details on this. Bandwidth efficiency will be measured by the average number of bits per transmitted symbol which equals the maximum spectral efficiency measured in bits per second per Hertz (b/s/Hz) for the modulations considered in this paper. The average error probability, defined as the average number of errors per transmission divided by the average number of transmitted bits, depends not only on the detector but also on the channel. For a channel which only adds white Gaussian noise,[4] the error probability is completely specified by the signal-to-noise ratio (SNR) in the detector. In Fig. 1, we

[2] Since this Special Issue is on adaptive modulation and transmission on wireless channels, we will here limit ourselves to wireless channels. The same principles also apply to other channels but the channel properties may be different.

[3] In wireless channels, the transmitted signal must be located in a given frequency band and for this reason a carrier frequency is used to obtain a bandpass signal in the required frequency band.

[4] Such a channel is normally referred to as an additive white Gaussian noise (AWGN) channel.

show an example of bit error probability[5] versus SNR per symbol (received SNR) for Gray coded quadrature amplitude modulation (QAM) with optimum (symbol) detection for 2 (red), 4 (blue), 6 (green), 8 (magenta), 10 (cyan), and 12 bits per symbol (brown), respectively.[6]

When this modulation is used for an application requiring at most say 10^{-5} in BER, it is clear that 2 bits per symbol can be used and fulfill this requirement when SNR is at least 12.6 dB. However, at the expense of increasing the transmit power by 6.9 dB, such that the SNR reaches 19.5 dB, 4 bits per symbol can be used which means that the bandwidth efficiency is doubled. With another increase of 6.1 dB in transmit power, 6 bits per symbol can be used, etc. Instead of increasing the transmit power to obtain a gain in SNR, a similar gain can be obtained by moving the transmitter and receiver closer together (if possible), since the transmitted power decays with increasing distance [25]–[28]. The problem in most applications is that there is a maximum allowed transmit power and the distance between the transmitter and receiver may vary and is not always known at the transmitter. Thus, with a fixed modulation, one has to choose a modulation method that gives a high enough SNR to obtain the required error probability at the maximum separation between transmitter and receiver. This, in fact, means that the error probability will be much lower than required at all smaller distances between the transmitter and receiver.

B. Fixed Modulation in Fading

Most wireless channels are affected by fading in addition to added noise and interference [25]–[28]. Fading is due to multipath propagation between the transmit and receive antennas. In its simplest form, the time delays between these multipath components are small compared with the symbol time of the modulation, resulting in so-called flat fading. The effect is that the signals arriving at the receive antenna experience different carrier phases causing the power of the received signal (the sum of all the multipath components) to depend on the carrier phases of the multipath components. A flat fading channel is often modeled as an AWGN channel with an exponentially distributed instantaneous SNR and a uniformly distributed carrier phase of the received signal. This particular fading channel is referred to as a Rayleigh-fading channel since the received amplitude is Rayleigh distributed.[7] An example

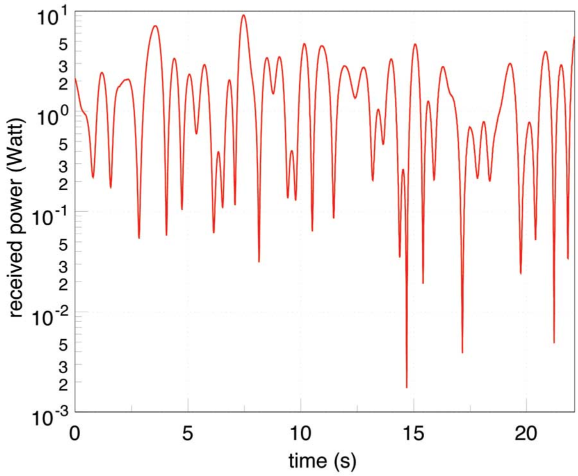

Fig. 2. *Example of received power versus time when a sinusoidal tone with constant power is transmitted on a flat Rayleigh-fading channel. The Doppler frequency is 1 Hz. The time axis will scale with the inverse of the Doppler frequency for other Doppler frequencies.*

of the received power versus time, when a sinusoidal tone with constant power is transmitted on a Rayleigh-fading channel with 1-Hz Doppler frequency, is shown in Fig. 2. The time axis will scale with the inverse of the Doppler frequency for other Doppler frequencies [25]–[28]. In this example, the average received power is set to 1 W. It is easy to see from this example, that the multipath propagation can result in a power (and thus SNR) variation of up to 40 dB from time to time. By combining the results in Figs. 1 and 2, it is clear that the instantaneous BER varies significantly from very low values when the instantaneous received signal power is large to 0.5 when the instantaneous received signal power approaches zero.[8]

In the ideal case, we assume that the receiver is able to perfectly track the carrier phase of the received signal, but still the average BER will be much worse as compared with the results in Fig. 1. The average BER is, in fact, the average of the instantaneous BER, as given in Fig. 1 for QAM, over the probability density function (pdf) of the instantaneous SNR. That leads to a BER that is proportional to the inverse of the average SNR (for Rayleigh-fading) rather than exponentially decaying with increasing SNR. As an example, in Fig. 3, we show the average BER versus average SNR for 4QAM on a flat Rayleigh-fading channel. To obtain an average BER of 0.001, we now require an SNR of 27 dB instead of 9.7 dB on the nonfading channel. In fixed modulations, a link with fading is normally designed based on a given operation point of the average SNR, resulting in large

[5]In this paper we use BER as abbreviation for bit error probability.

[6]The expressions for the BER of QAM are rather complicated for higher constellations. In this graph, we have used the approximation given in [1, eq. (9)], see also [17]. This approximation is accurate for all considered QAMs except at high BER. The same approximation has been used in the design and performance evaluation of the adaptive modulations discussed later in this paper. The expression itself is not given in this paper since it is not important for the understanding of the adaptive modulations and their properties.

[7]When a line-of-sight component exists in addition to the multipath components, the channel is referred to as a Ricean-fading channel.

[8]A Ricean-fading channel leads to somewhat better BER performance for a given average SNR, but it will still experience significant variation in instantaneous BER.

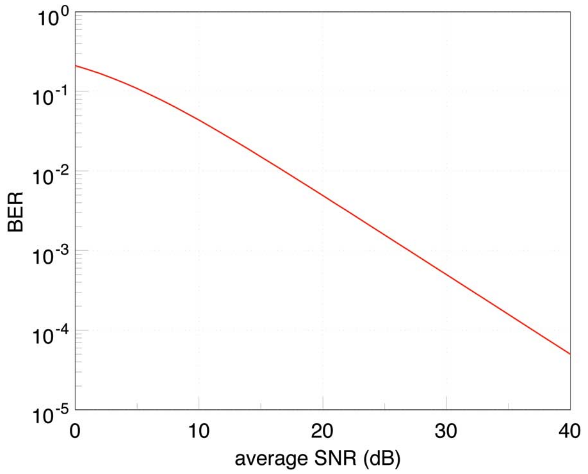

Fig. 3. *BER versus SNR per symbol for QAMs with 2 bits per symbol on a Rayleigh-fading channel.*

variations in instantaneous BER due to fading, also when the transmitter–receiver distance is unchanged. As we will see later, this variation can be removed or reduced by adaptive modulation and in addition a significant gain in spectral efficiency can be obtained.

More advanced fading channels, when the delay difference between the multipath components are larger than about 1/10th of the symbol period are outside the scope of this paper. The interested reader is referred to [25]–[28] for more details on such channels.

C. Adaptive Modulation

With fixed modulation, the modulator (transmitter) does not have (use) any information on the received SNR or other channel parameters available. It is usually designed for a certain minimum (average) SNR, which is related to the maximum coverage distance of the link, in such a way that the maximum allowed error probability is guaranteed within the coverage area. In an adaptive modulation method, on the other hand, channel information is made available to the transmitter. In its simplest form, the instantaneous SNR is made available but for more complex channels, more channel information can be made available. A simple block diagram, showing only the important part, of an adaptive modulation scheme, is shown in Fig. 4.

The green box represents the channel. It can be anything from pure addition of noise to complex time-varying filtering of the transmitted signal plus addition of noise and interference. In this paper, we will limit the channel to be a simple flat fading channel which only attenuates the signal amplitude, changes the carrier phase of the transmitted signal, and adds Gaussian noise. The channel is totally out of the control of the link designer.

The two blue boxes represent the modulation in the transmitter and the detection of the received signal in the receiver. These schemes have to be designed properly by the link designer, given the knowledge available about the channel. Almost all transmitters in wireless systems use some form of channel coding to improve the quality of the wireless link and for this reason the blue transmitter block also includes the word *coding*. The coding can either be in the traditional form of coding followed by modulation (each done independent of the other) or joint coding and modulation [17]–[24], [29]. The detection block, of course, has to be designed for the selected coding and modulation.

The pink block represents channel estimation. Most detection schemes assume that some channel parameters are already estimated and made available to the detector. One typical such parameter is the carrier phase offset, which is assumed known by so-called coherent detectors [30]–[32]. For some signaling constellations, both the amplitude and the carrier phase of the received signal need to be known. For even more advanced channels, an impulse response model of the channel might need to be estimated and made available to the detector. This block also has to be designed by the link designer.

The blocks described so far appear also in fixed modulation. The two remaining blocks are, however, specific to adaptive modulations. In its simplest form, when the channel changes very slowly, the estimated channel parameters are made available to the transmitter. From these parameters, the transmitter decides the modulation and coding parameters to be used, and this is referred to rate/power adaptation (the left red block) in Fig. 4. However, the transmitter is by no means limited to changing the rate and/or power only, but could also change other parameters in the modulation and coding scheme which influence the performance of the scheme. We will describe how this is done for adaptive QAM modulations in more detail later, but on a flat fading channel, typically the rate and power is adapted such that the required BER (or lower) is obtained with the highest possible spectral efficiency in the channel bandwidth available for the link.

Since processing of information is involved in channel estimation and rate adaptation, both these operations

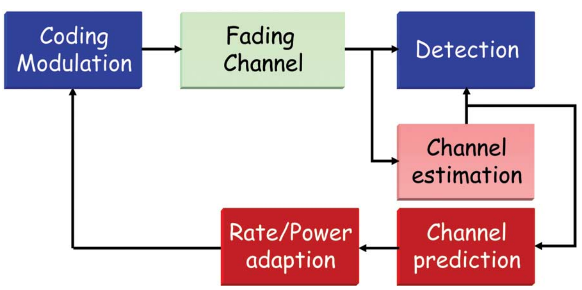

Fig. 4. *Major functions in an adaptive modulation system.*

will result in some latency. Moreover, on many links, the channel parameters or the modulation parameters have to be transmitted on a return channel from the receiver to the transmitter, which adds additional latency. If the channel changes significantly during this time period, the modulation parameters are outdated, resulting in poor adaptation. To overcome this latency, a second red block denoted channel prediction is included in the block diagram. The purpose of this block is to use the current and previous channel estimates to form a model of the channel and use this model to predict future channel parameters.[9] In this case, it is the predicted parameters that are used for the rate adaptation.

In the rest of this paper, we will describe the simplest forms of adaptive QAM modulations and their performance. This will be done for two cases; when the channel power gain is perfectly known in the transmitter and a prediction of the channel power gain with a certain prediction accuracy is available in the transmitter. Other contributions in this Special Issue will deal with more advanced adaptive transmission schemes, including channel prediction and coding of channel information to be efficiently transmitted on the return channel. We will also briefly discuss some other contributions to adaptive modulation and coding and the sensitivity of such schemes to errors in parameters.

III. ADAPTIVE MODULATION ON A KNOWN CHANNEL

In this section, we will give an overview of adaptive QAM modulation when used on a flat fading channel. The channel gain $\sqrt{g[l]}$ which affects the lth transmitted symbol is assumed perfectly known in both the receiver and the transmitter.[10] Moreover, the receiver is assumed to be an ideal coherent receiver which knows the channel phase without error. Referring to Fig. 4, this means that the channel estimation block delivers error-free estimates of the channel gain, and the predictor is not needed. We also assume that the feedback link has no latency. This case is studied in detail in [1] and we will use the same notation here as in [1]. The interested reader is referred to [1] for more details.

The average transmit power, the variance of the noise in the receiver, the bandwidth, and the average channel power gain, respectively, will be denoted $\overline{S}$, σ^2, B, and $\overline{g}$. We can assume that $\overline{g}=1$ with appropriate scaling of $\overline{S}$. The instantaneous received SNR is $\gamma[l]=\overline{S}g[l]/\sigma^2$, when the transmit power is constant and equal to $\overline{S}$. Since $g[l]$ is stationary, the pdf of $\gamma[l]$ is independent of l and will be denoted $p(\gamma)$. For the Rayleigh-fading channel

$$p(\gamma)=\begin{cases}\frac{1}{\Gamma}\exp\left(-\frac{\gamma}{\Gamma}\right), & \gamma\geq 0\\ 0, & \gamma<0\end{cases}\tag{1}$$

where $\Gamma=\overline{S}/\sigma^2$ is equal to the average SNR in the receiver. This instantaneous SNR and its pdf only reflect the influence of the channel on the SNR and not the influence of a varying transmit power. In general, in an adaptive modulation scheme, the transmit power will vary depending on $\gamma[l]$ and will, thus, be denoted $S(\gamma[l])$. The instantaneously received SNR is then $\gamma[l]S(\gamma[l])/\overline{S}$. It should be noted that the pdf of this received SNR is different from the pdf in (1) when the modulation uses a varying transmit power. To make the notation simpler, we will, when the context is clear, omit the time index l and simply write γ and $S(\gamma)$.

A. Adaptive Rate, Maximum BER, and Constant Power

One simple form of adaptive modulation is when only the transmission rate $R=R(\gamma[l])$ is changed when the channel power gain changes. From Fig. 1, it is clear that BER can be kept below a certain maximum value, although the number of bits per symbol $k=k(\gamma[l])$ is increased with increasing channel power gain.[11] In this case, however, no transmission should be done below a certain value of the channel power gain or the BER will be higher than the maximum allowed value. Assuming that we use N different constellations (modulations), each with k_i bits per symbol, this means that we use the ith constellation when $\gamma_i\leq\gamma<\gamma_{i+1}$, where $0\leq i\leq N-1$ and $\gamma_N=\infty$. These intervals are referred to as rate regions.[12] The transmit power becomes

$$S(\gamma)=\begin{cases}S, & \gamma\geq\gamma_0\\ 0, & \gamma<\gamma_0\end{cases}\tag{2}$$

where S is given from

$$E[S(\gamma)]=S\int_{\gamma_0}^{\infty}p(\gamma)d\gamma=\overline{S}\tag{3}$$

[9]The channel model might in some cases be available beforehand, and then the predictor does not need to update the model but only find the predicted channel values.

[10]Fig. 2 shows an example realization of the channel power gain versus time; thus $g[l]$ are samples drawn at symbol rate from the curve shown.

[11]Please note that SNR on the horizontal axis in Fig. 1 is proportional to the channel power gain. Thus, this axis could just as well be labeled with channel power gain.

[12]Please note that γ is the received SNR when the transmit power is constant and that this is proportional to the channel power gain. Thus, the rate regions could just as well be defined as regions of the channel power gain. The end points γ_i are in the same way defined in the scale of the received SNR when the transmitted power is constant as is done in [1]. The actual received SNR will be larger since we will be able to increase the transmit power as seen next.

such that the average transmitted power is the same as when $S(\gamma) = \overline{S}$ for all γ. Here, $E(\cdot)$ denotes the expected value, which in this case is evaluated over the pdf of the SNR. This, in fact means that the transmitted power can be increased when transmission occurs, resulting in a somewhat higher received SNR for a given channel power gain as compared with the case when the transmit power is constant. Thus, the channel can be used at somewhat lower channel power gains without violating the BER requirement, which will increase the spectral efficiency of the link.

A communication link should normally operate at or below a certain maximum BER. This design goal parameter will be denoted $\mathrm{BER}_{\mathrm{dg}}$. Thus, $\mathrm{BER}(\gamma S/\overline{S}) \leq \mathrm{BER}_{\mathrm{dg}}$ for all $\gamma \geq \gamma_0$, where $\mathrm{BER}(\cdot)$ is a function relating the BER to the instantaneous SNR for the modulation scheme considered.[6] For this particular scheme, this will be fulfilled if

$$\mathrm{BER}_i(\gamma_i S/\overline{S}) \leq \mathrm{BER}_{\mathrm{dg}}; \quad 0 \leq i \leq N-1 \tag{4}$$

where $\mathrm{BER}_i(\gamma_i S/\overline{S})$ refers to the BER for modulation i (used when $\gamma_i \leq \gamma < \gamma_{i+1}$) at received SNR equal to $\gamma_i S/\overline{S}$ (which corresponds to the lowest SNR for this modulation). Now, it remains to find the N rate region boundaries γ_i fulfilling (4) with equality. These rate regions will depend on the average SNR (Γ) of the channel.

The two most important performance measures of this adaptive modulation scheme is the spectral efficiency and the average BER for a given value of N, a given value of $\overline{S}$, and a given channel (a given Γ). Assuming Nyquist data pulses at the lowest possible bandwidth $1/T_s$, where T_s is the symbol period of the modulation, to avoid intersymbol interference [17], the spectral efficiency becomes

$$\frac{R}{B} = \sum_{i=0}^{N-1} k_i \int_{\gamma_i}^{\gamma_{i+1}} p(\gamma) d\gamma. \tag{5}$$

The average BER can, for adaptive modulation schemes, be defined in at least two different ways, as discussed in [1]. Here, we choose to use the same definition as in [1] which means

$$\overline{\mathrm{BER}} = \frac{E[\text{number of error bits per transmission}]}{E[\text{number of bits per transmission}]} \tag{6}$$

and is given by

$$\overline{\mathrm{BER}} = \frac{\sum_{i=0}^{N-1} k_i \int_{\gamma_i}^{\gamma_{i+1}} \mathrm{BER}_i(\gamma S/\overline{S}) p(\gamma) d\gamma}{\sum_{i=0}^{N-1} k_i \int_{\gamma_i}^{\gamma_{i+1}} p(\gamma) d\gamma}. \tag{7}$$

The performance of the adaptive modulation schemes described above will be illustrated by four different examples of design parameters. The modulations used when $N = 3$ are QAM with 2, 4, and 6 bits per symbol. When $N = 6$, we use QAM with 2, 4, 6, 8, 10, and 12 bits per symbol. The schemes have been designed for a maximum BER of 0.001 and 10^{-7}, respectively. The channel is a flat Rayleigh-fading channel defined according to (1). From Fig. 3, we know that an average SNR of 27 dB is required to obtain an average BER of 0.001 with 4QAM (2 b/s/Hz) when this channel is used with a fixed modulation. The spectral efficiency and average BER of the adaptive schemes are shown in Figs. 5 and 6, respectively.

From Fig. 5, it is clearly seen that a spectral efficiency of 2 b/s/Hz can be obtained already at an average SNR of 15 dB for both the considered adaptive modulations when designed for a maximum BER of 0.001 (green and red curves). The actual average BER, as seen in Fig. 6, is actually almost a factor of 10 lower. To continue the comparison with a fixed 4QAM, we find that, at 27 dB, the spectral efficiency is about 5.5 b/s/Hz with six modulations used and designed for a maximum BER of 0.001 (the actual BER is again about the same as at 15 dB). With three modulations, the spectral efficiency starts to approach the floor region at 27 dB, but the spectral efficiency is still around 5 b/s/Hz. The reason for the floor region is that there is no modulation with more than 6 bits per symbol and this limits the spectral efficiency to 6 b/s/Hz for this scheme. However, since QAM with 6 bits per symbol is

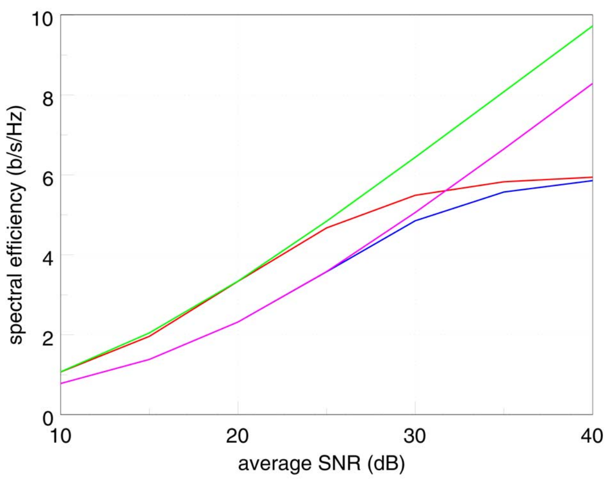

Fig. 5. ***Maximum spectral efficiency versus SNR per symbol for adaptive QAM modulations with adaptive rate only when $N = 3$ and $\mathrm{BER}_{\mathrm{dg}} = 0.001$ (red), $N = 3$ and $\mathrm{BER}_{\mathrm{dg}} = 10^{-7}$ (blue), $N = 6$ and $\mathrm{BER}_{\mathrm{dg}} = 0.001$ (green), and $N = 6$ and $\mathrm{BER}_{\mathrm{dg}} = 10^{-7}$ (magenta). These schemes are designed to have BER $\leq \mathrm{BER}_{\mathrm{dg}}$. Please note that the used fixed modulations have a constant spectral efficiency equal to the major grid lines on the vertical axis (12 bits per symbols is outside the figure).***

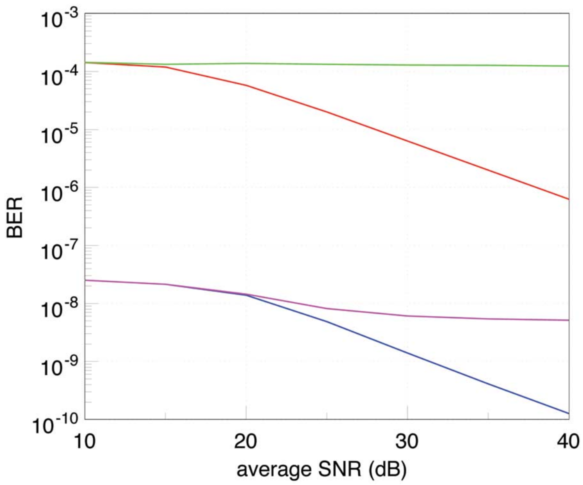

Fig. 6. ***Average BER versus SNR per symbol for adaptive QAM modulations with adaptive rate only when $N = 3$ and $\mathrm{BER}_{\mathrm{dg}} = 0.001$ (red), $N = 3$ and $\mathrm{BER}_{\mathrm{dg}} = 10^{-7}$ (blue), $N = 6$ and $\mathrm{BER}_{\mathrm{dg}} = 0.001$ (green), and $N = 6$ and $\mathrm{BER}_{\mathrm{dg}} = 10^{-7}$ (magenta). These schemes are designed to have $\mathrm{BER} \leq \mathrm{BER}_{\mathrm{dg}}$.***

now used also when more bits per symbol could have been used if higher constellations were available, the BER will drop even further below 0.001. From Fig. 6, we see that the average BER with three modulations at 27 dB is around 10^{-5} (red curve). With more modulations, this drop in BER can be avoided and the spectral efficiency increased as clearly demonstrated from the figures. However, it may be more complex to use more constellations and the higher constellations will also put higher requirements on synchronization accuracy [31].

When the adaptive modulations are designed for other maximum BER values, a similar performance behavior will result. With a maximum BER of 10^{-7} as illustrated, the spectral efficiency will be somewhat lower than at BER equal to 0.001, but there is still a significiant gain compared with a fixed modulation scheme. From Fig. 5, we can see that an average SNR increase of about 3–4 dB is needed when maximum BER is 10^{-7} to obtain the same spectral efficiency as with BER 0.001. For this maximum BER, the actual average BER is again a factor of almost 10 smaller than the maximum BER at the lower range of average SNRs, while it is a factor of more than 10 smaller than the maximum BER at the higher range of SNRs.

B. Adaptive Rate, Average BER, and Constant Power

One drawback with the design procedure described above is that the instantaneous BER is lower than the design goal $\mathrm{BER}_{\mathrm{dg}}$ at all instantaneous SNRs except the rate region boundary points γ_i for $0 \leq i \leq N-1$. Therefore, the average BER will also be lower than $\mathrm{BER}_{\mathrm{dg}}$ for all channels. Another design rule avoiding this drawback is to require the average BER, as calculated from (7), to become equal to the design goal instead, i.e., require that $\overline{\mathrm{BER}} = \mathrm{BER}_{\mathrm{dg}}$. According to [1], it is hard to find the rate region boundaries for this case. A suboptimal solution to this optimization problem, which is proposed in [1], is to assume that all the rate region boundaries for the average BER constraint optimization problem are equal to a constant (smaller than one) times the corresponding rate boundaries for the maximum BER design above. Then, the constant needs to be found such that the average BER constraint is fulfilled. This constant will depend on the average SNR. This optimization problem can be solved numerically and results for the $N = 6$ examples above are available in [1, Fig. 9]. The gain in average SNR for a given spectral efficiency is about 1.5 dB when BER is equal to 0.001 but less than 1 dB when BER is 10^{-7}.

C. Constant Rate and Adaptive Power

In the sections above, the channel power gain defined the rate used in the transmitter, while the power was kept constant. It is of course also possible to do the opposite; to keep the (bit) rate constant but adapt the power such that a BER constraint is fulfilled. This means that a fixed modulation with k bits per symbol is used when $\gamma \geq \gamma_0$. If the transmit power $S(\gamma)$ is chosen such that BER becomes equal to $\mathrm{BER}_{\mathrm{dg}}$ for all $\gamma \geq \gamma_0$, an instantaneous BER constraint is fulfilled. This is obtained when the transmit power is chosen as

$$S(\gamma) = \begin{cases} \frac{\overline{S}}{\gamma}\mathrm{BER}^{-1}(\mathrm{BER}_{\mathrm{dg}}), & \gamma \geq \gamma_0 \\ 0, & \gamma < \gamma_0 \end{cases} \tag{8}$$

where $\mathrm{BER}^{-1}(\cdot)$ is defined such that $\mathrm{BER}^{-1}(\mathrm{BER}(\gamma)) = \gamma$ and will be referred to as the inverse BER function.[13] Here, γ_0 must be chosen such that the average power becomes $\overline{S}$, i.e.,

$$\int_{\gamma_0}^{\infty} S(\gamma)p(\gamma)d\gamma = \overline{S}. \tag{9}$$

After evaluating the transmit power function $S_i(\gamma)$ and the corresponding lower SNR threshold $\gamma_{0,i}$ for all considered modulations $0 \leq i \leq N-1$ using the formulas above, the spectral efficiency becomes

$$\frac{R}{B} = \max_{k_i; 0\leq i\leq N-1}\left\{k_i \int_{\gamma_{0,i}}^{\infty} p(\gamma)d\gamma\right\}. \tag{10}$$

[13]For many modulations it is difficult, if not impossible, to find an analytical expression for the inverse BER function. This is the main reason that approximative BER expressions are used when designing the adaptive modulations.

From (7), which now only contains one term in the sums, it is straight forward to verify that the average BER becomes equal to the value $\mathrm{BER}_{\mathrm{dg}}$ which the scheme was designed for.

The spectral efficiency for this scheme with the same six modulations as used above is shown in Fig. 7. To simplify comparison, we have also included the corresponding spectral efficiency curves for the adaptive rate schemes with maximum BER constraint also shown in Fig. 5 above. In Fig. 7, we clearly see that the adaptive power schemes slightly outperform the adaptive rate schemes. However, when compared with the adaptive rate schemes with average BER constraint in [1], we conclude that the adaptive rate and adaptive power schemes have very similar spectral efficiency when both are designed to have the same average BER. To illustrate the adaptive transmit power for this scheme, we show $S(\gamma)/\overline{S}$ in Fig. 8 for the scheme designed for $\mathrm{BER}_{\mathrm{dg}} = 0.001$ when the average SNR (Γ) is 30 dB. Here, we see that no transmission takes place below around 17.5 dB. Above 17.5 dB, the power is adjusted in such a way that the BER is kept at 0.001. This is often referred to as water filling to the inverse BER function according to (8) and (9).

In the design above, the instantaneous BER becomes identical for all SNRs, due to the water filling transmit power allocation. A less restrictive BER requirement is that the average BER is equal to the design goal, i.e., requiring $\overline{\mathrm{BER}} = \mathrm{BER}_{\mathrm{dg}}$. In this case, the instantaneous BER can vary with SNR as long as the average over the SNR distribution is not affected. Thus, the transmit power does not have to be water filling to the inverse BER function but can be more general. This problem can be solved with a Lagrangian optimization procedure and we refer the interested reader to [1, p. 1570]. From the solution, one finds that the power allocation becomes different [1, Fig. 11] and the BER varies with instantaneous SNR, but the spectral efficiency is almost not changed.

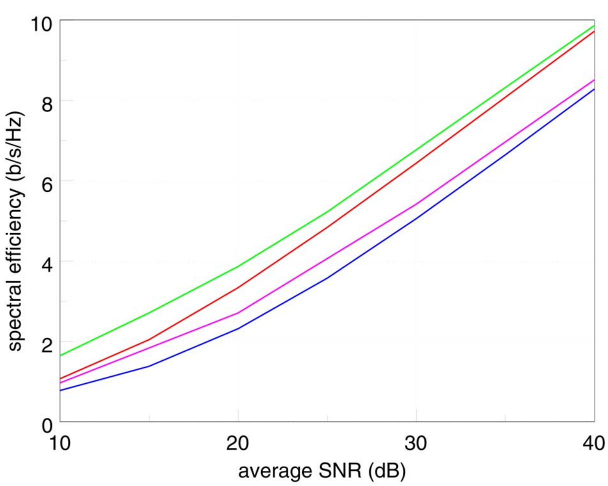

Fig. 7. ***Maximum spectral efficiency versus SNR per symbol for adaptive QAM modulations with adaptive power only when N = 6 and BER$_{\mathrm{dg}}$ = 0.001 (green) and N = 6 and BER$_{\mathrm{dg}}$ = 10^{-7} (magenta). These schemes are designed to have BER = BER$_{\mathrm{dg}}$ for all SNRs. The corresponding curves for an adaptive rate only scheme, designed to have BER ≤ BER$_{\mathrm{dg}}$, (from Fig. 5) are also shown (red and blue).***

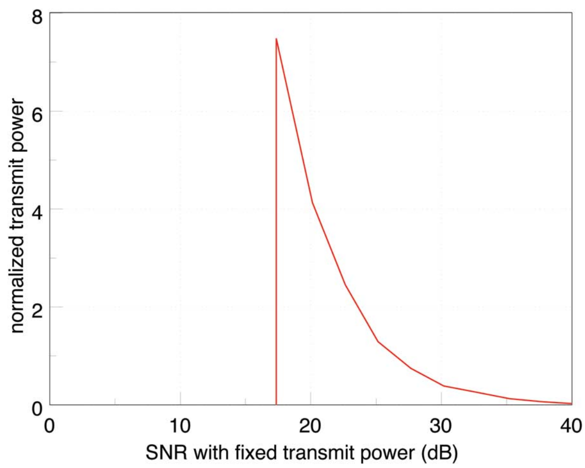

Fig. 8. ***$S(\gamma)/\overline{S}$ for the adaptive power and constant rate scheme, designed for BER = BER$_{\mathrm{dg}}$ = 0.001, when the average SNR (Γ) is 30 dB. The SNR on the abscissa is the SNR in the receiver when a fixed power equal to $\overline{S}$ is transmitted. This SNR is directly related to the channel power gain. The actual SNR in the receiver is equal to the product of the corresponding values on the horizontal axis and vertical axis.***

D. Adaptive Rate and Power

Let us now turn to the most general case of adaptive QAM modulation, where both the rate and the power are chosen based on channel power gain information. Here, we will cover the somewhat simpler problem of an instantaneous BER constraint in some details and only briefly discuss the more general case of an average BER constraint. Just as in Section III-A above, we assume N modulations, with the ith used when $\gamma_i \leq \gamma < \gamma_{i+1}$ and carrying k_i bits per symbol. When the instantaneous BER is required to be equal to $\mathrm{BER}_{\mathrm{dg}}$ for all SNRs, the transmit power must be chosen according to (8). A Lagrangian method can then be used to find the rate region boundaries γ_i, using the Lagrangian equation

$$J(\gamma_0, \ldots, \gamma_{N-1}, \lambda) = \sum_{i=0}^{N} k_i \int_{\gamma_i}^{\gamma_{i+1}} p(\gamma) d\gamma + \lambda \left[\sum_{i=0}^{N-1} \int_{\gamma_i}^{\gamma_{i+1}} S(\gamma) p(\gamma) d\gamma - \overline{S} \right] \quad (11)$$

where $\lambda \neq 0$ is a Lagrangian multiplier which needs to be optimized. The optimum rate regions and the optimum value of λ are found by solving the equation system obtained from $\partial J(\gamma_0, \ldots, \gamma_{N-1}, \lambda)/\partial \gamma_i = 0$ for $0 \le i \le N-1$ and $\partial J(\gamma_0, \ldots, \gamma_{N-1}, \lambda)/\partial \lambda = 0$. The solution becomes

$$\gamma_0 = \frac{\mathrm{BER}_0^{-1}(\mathrm{BER}_{\mathrm{dg}})}{k_0}\rho \tag{12}$$

and

$$\gamma_i = \frac{\mathrm{BER}_i^{-1}(\mathrm{BER}_{\mathrm{dg}}) - \mathrm{BER}_{i-1}^{-1}(\mathrm{BER}_{\mathrm{dg}})}{k_i - k_{i-1}}\rho; \quad 1 \le i \le N-1 \tag{13}$$

where ρ is found from the average power constraint in (9). Since BER is constant for all SNRs, the average BER becomes $\overline{\mathrm{BER}} = \mathrm{BER}_{\mathrm{dg}}$.

The spectral efficiency of this scheme is illustrated in Fig. 9. The modulations used are the same as before. In this case, we do not show the average BER since it becomes equal to the value used for the design, i.e., 0.001 and 10^{-7}, respectively, in this case. The behavior of the spectral efficiency is very similar to the schemes discussed before but there is a small gain in spectral efficiency for a given average SNR as compared with the scheme that adapts only power or rate. The gain is small when the average SNR is small but increases with increasing average SNR. In Fig. 10, $S(\gamma)/\overline{S}$ is shown for the scheme with $N = 6$ and BER 0.001 when the average SNR is 30 dB. The SNR corresponding to the discontinuities in the curve are the rate region boundaries γ_i. No transmission takes place below around 8 dB. Above 8 dB, the power is adapted according to water-filling of the inverse BER function of each modulation according to (8) such that the average power is constant according to (9).

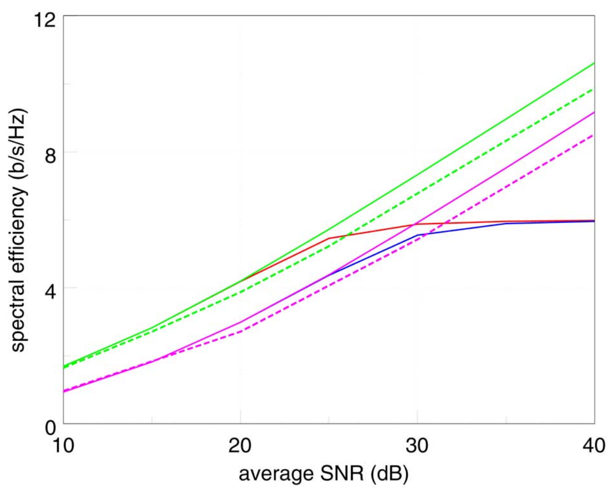

Fig. 9. ***Maximum spectral efficiency versus SNR per symbol for adaptive QAM modulations with adaptive rate and power, when N = 3 and BER$_{\mathrm{dg}}$ = 0.001 (red), N = 3 and BER$_{\mathrm{dg}}$ = 10^{-7} (blue), N = 6 and BER$_{\mathrm{dg}}$ = 0.001 (green), and N = 6 and BER$_{\mathrm{dg}}$ = 10^{-7} (magenta). As a comparison, the dashed lines show the spectral efficiency when N = 6 for the adaptive power and fixed rate scheme which are also shown in Fig. 7. These schemes are designed to have BER = BER$_{\mathrm{dg}}$ for all SNRs.***

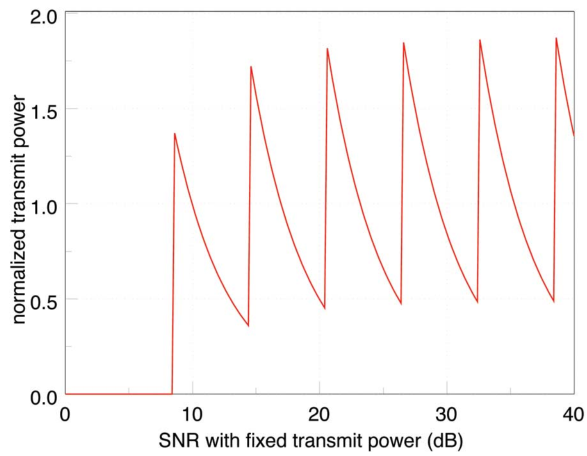

Fig. 10. ***$S(\gamma)/\overline{S}$ for the adaptive power and rate scheme designed for an instantaneous BER of 0.001 when the average SNR (Γ) is 30 dB. Please refer to the legend of Fig. 8 for further explanation of the curve.***

Also in this general case, an average BER constraint can be used instead of an instantaneous BER constraint. Then, the power does not have to be chosen such that BER is equal to the design goal for all SNRs, but it is enough that BER on average is equal to the constraint. This problem can also be solved using a Lagrangian method but now the Lagrangian equation has to include a third term that corresponds to the average BER constraint in addition to the two terms used already above in (11). A suboptimum solution to this problem is described in detail in [1]. The gain in spectral efficiency using this constraint is very small compared with the results shown in Fig. 9.

IV. ADAPTIVE MODULATION ON A PREDICTED CHANNEL

In Section III, the channel power gain was assumed perfectly known by the transmitter when the modulation schemes were to be chosen. In a practical implementation, the channel power gain is never known but must be estimated, which means that the estimated value will have to be used when selecting modulation. Moreover, in most systems, there is a delay, which sometimes is significant, between the time for which a channel power gain is estimated until this estimated gain is available in the transmitter. Several factors contribute to this delay but the most dominant term is the time it takes to transmit the estimate from the receiver to the transmitter on a return channel. In systems where the uplink and downlink use different carrier frequencies such that the channels are not reciprocal, this delay may be significant. Other factors influencing the delay are the processing time in transmitter and receiver.

In all systems where the time delay for making the channel power gain estimate available in the transmitter is on the order of the channel coherence time[14] or larger, the estimate will be outdated before it is available in the transmitter. This is common to many wireless systems where either the transmitter or the receiver (or both) are moving. In these systems, the error of the estimated value will simply be too large to make it usable for selecting the modulation scheme. Therefore, a predictor that predicts the channel power gain into future is necessary. In [2], a channel power predictor and the corresponding optimum adaptive modulations are presented for flat fading channels. In this section, we will in some detail present the adaptive modulation schemes based on predicted channels. We will focus on illustrating the results rather than giving all the necessary equations to find the solutions. The mathematics is a bit more complicated than for the schemes assuming a known channel and, therefore, we will not go into too much detail. The interested reader is referred to [2] for details.

The BER will now depend not only on the instantaneously received SNR γ (or channel power gain g), which is defined as the SNR for a constant transmit power $\overline{S}$, but also on the predicted SNR $\widehat{\gamma}$. Here $\widehat{\gamma}$ denotes the predicted value of γ and is a function of the predicted value of the channel power gain. The instantaneous received SNR becomes $\gamma S(\widehat{\gamma})/\overline{S}$, i.e., it is a function of both the actual channel SNR and the predicted SNR. The BER can now be evaluated using the standard formulas for AWGN by replacing SNR in these formulas with $\gamma S(\widehat{\gamma})/\overline{S}$. We will denote this BER by $\mathrm{BER}(\gamma, \widehat{\gamma})$. Then, average BER for a given predicted SNR can be obtained from

$$\mathrm{BER}(\widehat{\gamma}) = \int_0^{\infty} \mathrm{BER}(\gamma, \widehat{\gamma}) f_{\gamma}(\gamma|\widehat{\gamma}) d\gamma \tag{14}$$

where $f_{\gamma}(\gamma|\widehat{\gamma})$ denotes the conditional pdf of SNR γ given a predicted SNR $\widehat{\gamma}$.[15]

The rate regions are now specified as regions of the predicted SNR values and the boundaries are denoted $\widehat{\gamma}_i$ where $0 \le i \le N$ and $\widehat{\gamma}_N = \infty$ as before. The spectral efficiency becomes

$$\frac{R}{B} = \sum_{i=0}^{N-1} k_i \int_{\widehat{\gamma}_i}^{\widehat{\gamma}_{i+1}} f_{\widehat{\gamma}}(\widehat{\gamma}) d\widehat{\gamma} \tag{15}$$

where $f_{\widehat{\gamma}}(\widehat{\gamma})$ is the pdf of the predicted SNR. Similarly, the average BER is given by

$$\overline{\mathrm{BER}} = \frac{\sum_{i=0}^{N-1} k_i \int_{\widehat{\gamma}_i}^{\widehat{\gamma}_{i+1}} \mathrm{BER}_i(\widehat{\gamma}) f_{\widehat{\gamma}}(\widehat{\gamma}) d\widehat{\gamma}}{\sum_{i=0}^{N-1} k_i \int_{\widehat{\gamma}_i}^{\widehat{\gamma}_{i+1}} f_{\widehat{\gamma}}(\widehat{\gamma}) d\widehat{\gamma}}. \tag{16}$$

It is outside the scope of this paper to go into detail on predictors since this is the topic of another paper in this special issue [33]. In this paper, we will use the results for the predictor presented in [2] to illustrate the performance of adaptive modulation schemes on predicted channels. The interested reader is referred to [2] for the details on the predictor and the expressions for the pdfs needed above.[16]

A. Adaptive Rate, Maximum BER, and Constant Power

The optimum adaptive modulation scheme with adaptive rate and constant power, designed for a maximum BER, when the SNR is predicted, is found in a very similar way as the corresponding scheme discussed in Section III-A when the channel is known. The transmit power is

$$S(\widehat{\gamma}) = \begin{cases} S, & \widehat{\gamma} \ge \widehat{\gamma}_0 \\ 0, & \widehat{\gamma} < \widehat{\gamma}_0 \end{cases} \tag{17}$$

where S must fulfill

$$S \int_{\widehat{\gamma}_0}^{\infty} f_{\widehat{\gamma}}(\widehat{\gamma}) d\widehat{\gamma} = \overline{S}. \tag{18}$$

Then, according to the maximum BER constraint

$$\mathrm{BER}_i(\widehat{\gamma}_i) \le \mathrm{BER}_{\mathrm{dg}}; \quad 0 \le i \le N-1 \tag{19}$$

[14]The channel coherence time is a measure of the time variability of the channel. A quite common definition is the time over which the correlation between two channel power gains are equal to 0.5; see, e.g., [27] for more details.

[15]In this section, we use $f(\cdot)$ to denote a pdf rather than $p(\cdot)$ as we did in the previous section. This is to follow the notation used in [2].

[16]The expressions for $f_{\gamma}(\gamma|\widehat{\gamma})$ and $f_{\widehat{\gamma}}(\widehat{\gamma})$ are given in [2, eq. (16)] and [2, eq. (17)], respectively.

where, as before, $\text{BER}_i(\cdot)$ refers to the BER for modulation i. From these equations, we can solve for S and $\widehat{\gamma}_i$ for $0 \leq i \leq N-1$.

B. Adaptive Rate, Average BER, and Constant Power

As with the schemes for known channel power gain, the drawback with the scheme described in Section IV-A is that the average BER will be lower than BER_{dg}. With a Lagrangian method, we can instead design for an average BER constraint to assure that $\overline{\text{BER}} = \text{BER}_{\text{dg}}$. The Lagrangian equation to be used is given as

$$J(\widehat{\gamma}_0,\ldots,\widehat{\gamma}_{N-1},\lambda) = \sum_{i=0}^{N} k_i \int_{\widehat{\gamma}_i}^{\widehat{\gamma}_{i+1}} f_{\widehat{\gamma}}(\widehat{\gamma})d\widehat{\gamma} + \lambda\left[\sum_{i=0}^{N-1}\int_{\widehat{\gamma}_i}^{\widehat{\gamma}_{i+1}} \left(\text{BER}(\widehat{\gamma}) - \text{BER}_{\text{dg}}\right) f_{\widehat{\gamma}}(\widehat{\gamma})d\widehat{\gamma}\right] \quad (20)$$

where $\lambda \neq 0$ is a Lagrangian multiplier which needs to be optimized. The optimum solution is now found by solving $\partial J/\partial\widehat{\gamma}_i = 0$ for $0 \leq i \leq N-1$ and $\partial J/\partial\lambda = 0$. In addition to this, the transmit power must fulfill (18).

C. Adaptive Rate and Power

In this section, we choose both rate and power based on the predicted SNR. The BER constraint can either be $\text{BER}(\widehat{\gamma}) = \text{BER}_{\text{dg}}$ for all $\widehat{\gamma}$ or $\overline{\text{BER}} = \text{BER}_{\text{dg}}$. The former is an instantaneous BER constraint while the latter is an average BER constraint which is less restrictive. In the former, $S(\widehat{\gamma})$ has to be chosen such that $\text{BER}(\widehat{\gamma}) = \text{BER}_{\text{dg}}$ for all $\widehat{\gamma}$ but the average of $S(\widehat{\gamma})$ must be equal to $\overline{S}$. The Lagrangian equation to use for this case is

$$J(\widehat{\gamma}_0,\ldots,\widehat{\gamma}_{N-1},\lambda) = \sum_{i=0}^{N} k_i \int_{\widehat{\gamma}_i}^{\widehat{\gamma}_{i+1}} f_{\widehat{\gamma}}(\widehat{\gamma})d\widehat{\gamma} + \lambda\left[\sum_{i=0}^{N-1}\int_{\widehat{\gamma}_i}^{\widehat{\gamma}_{i+1}} S(\widehat{\gamma}) f_{\widehat{\gamma}}(\widehat{\gamma})d\widehat{\gamma} - \overline{S}\right]. \quad (21)$$

The interested reader is referred to [2] for details in this solution.

The average BER constraint can be solved using a Lagrangian equation with a third term in addition to the two other terms. The third term is identical to the second term in (20) and will assure that the average BER becomes equal to the design BER. The two Lagrangian multipliers will in this case be λ_1 and λ_2, respectively.

D. Numerical Results

In this subsection, we will illustrate the performance of the adaptive modulations designed for predicted SNRs. The same six modulations as before, have been used in the numerical examples. The three schemes above are in the figures below denoted:

1) I-BER, V-Pow for the scheme designed for an instantaneous BER constraint with adaptive power according to (21);
2) I-BER, C-Pow for the scheme designed for a maximum (instantaneous) BER constraint with constant power according to Section IV-A; and
3) A-BER, C-Pow for the scheme designed for an average BER constraint with constant power according to (20).

The accuracy of the predictors are given by the normalized mean-square error $\sigma^2_{\epsilon_p}$ of the channel gain prediction error; see [2] for detailes on how this measure is calculated for the predictor considered. As a reference, $\sigma^2_{\epsilon_p}$ would be equal to 0.5 when the predicted SNR is equal to the average power which is a really bad predictor. Here we show results for $\sigma^2_{\epsilon_p} = 0.001$ and $\sigma^2_{\epsilon_p} = 0.1$. The former case corresponds to an almost perfect predictor while the latter is a quite poor predictor.

In Fig. 11, we show the normalized transmit power versus predicted SNR for the two considered predictors when the average SNR on the channel is 20 dB. The solid curve corresponds to the almost perfect predictor and this curve is very similar to the curve in Fig. 10. The reason for the small difference is that the average SNR is 30 dB

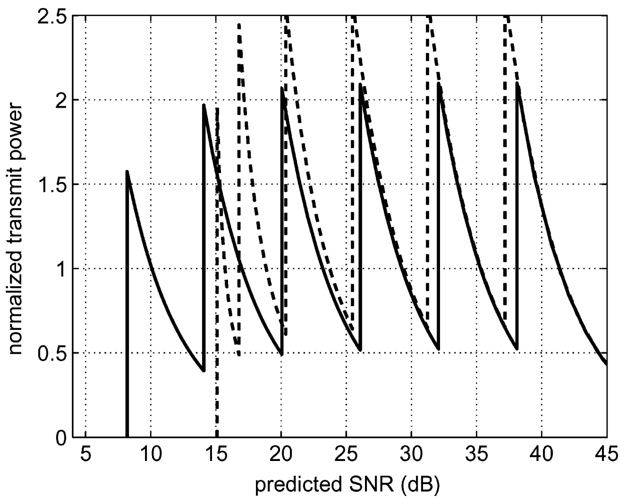

Fig. 11. *$S(\widehat{\gamma})/\overline{S}$ for the adaptive power and rate scheme designed for an instantaneous BER of 0.001 when the average SNR (Γ) is 20 dB. Solid and dashed lines correspond to $\sigma^2_{\epsilon_p} = 0.001$ and $\sigma^2_{\epsilon_p} = 0.1$, respectively. (From [2], copyright IEEE 2004.)*

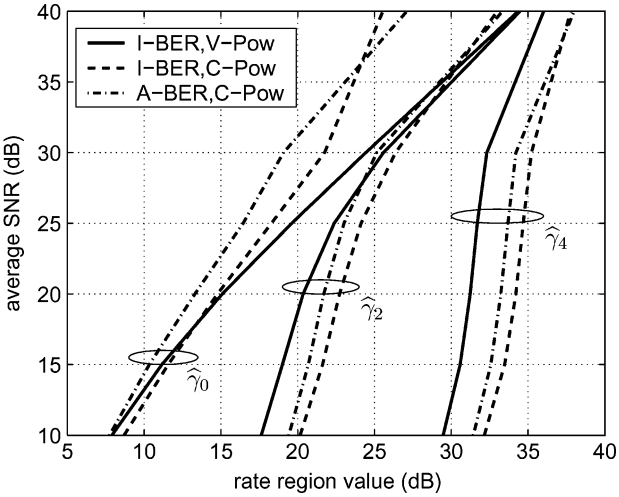

Fig. 12. *Selected optimum rate region boundaries $\hat{\gamma}_0$, $\hat{\gamma}_2$, and $\hat{\gamma}_4$ (horizontal axis) as a function of average SNR (Γ) (vertical axis) for the adaptive schemes using predicted SNR. For a given average SNR on the vertical axis, each graph shows, on the horizontal axis, the corresponding lower rate region value above which the given modulation should be used. Thus, the set of curves correspond to the lower limit of using 2 bits per symbol, 6 bits per symbol, and 10 bits per symbol, respectively. The design BER is 0.001 and $\sigma^2_{\epsilon_p} = 0.1$. (From [2], copyright IEEE 2004.)*

in Fig. 10 while it is 20 dB in Fig. 11. For the predictor with $\sigma^2_{\epsilon_p} = 0.1$, the normalized transmit power is quite different. The discontinuities in the curve corresponds to the boundaries of the rate regions. We can clearly see that the boundary values of the lower regions increase quite significantly when the performance of the predictor becomes worse, while the boundaries of the upper regions on the contrary decrease. In each rate region, the power follows a water-filling to the inverse BER function with respect to predicted power in order to keep the BER constant.

It is interesting to note, from this figure and some other figures in [2], that the rate boundaries are raised for SNR lower than the average SNR, while they are usually reduced for higher average SNR values when the prediction error variance increases. This is explained in some detail in [2]. The effect is that the scheme becomes cautious when entering into fading dips (low SNRs) and save the corresponding transmit power to be able to transmit at higher power when the channel becomes good. This strategy leads to optimized spectral efficiency.

In Fig. 12, the optimum rate region boundaries for three of the modulations are shown for different average SNRs. This is shown here for the predictor with $\sigma^2_{\epsilon_p} = 0.1$. Here it is clearly seen that the lower region boundary $\hat{\gamma}_0$ is increased a lot with increasing average SNR. The increase is faster for the scheme with variable power since here the power can be increased to compensate for the increase in the region boundary. The increase is smaller for the higher boundaries.

The spectral efficiency for the schemes optimize for predicted channel power gains are shown in Fig. 13. The three curves without marks are for the almost perfect predictor and these should be compared with the results for the corresponding schemes in Section III which are designed for perfect channel information. Such a comparison reveals that the schemes using a predictor with $\sigma^2_{\epsilon_p} = 0.001$ have basically the same spectral efficiency as the corresponding schemes designed for perfect channel knowledge. This is of course not surprising and should be the case. The more interesting comparison is to compare the spectral efficiency for the good and bad predictors. From Fig. 13, it is clear that the degradation due to large prediction errors is relatively small. At the lower average SNRs, an increase of about 1 dB in average power will compensate for the loss in spectral efficiency of the predictor with the large error variance as compared with the almost perfect predictor. For the large SNRs, the corresponding average power increase is 5–7 dBs.

E. Discussion

The results shown above clearly reveal that even with a quite bad predictor, the total spectral efficiency is increasing quickly with increasing average SNR, which makes the adaptive schemes much more attractive than the fixed modulation schemes. With these schemes, it is also possible to design a scheme that gives essentially the same BER for all channel SNRs which is not the case with fixed modulations. A return channel must, however, be available for transmission of the channel gain information to the transmitter. The rate needed on this channel and

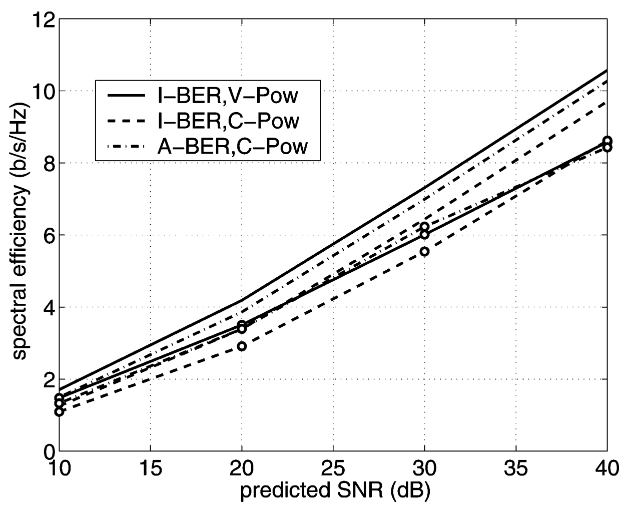

Fig. 13. *Maximum spectral efficiency versus SNR per symbol for adaptive QAM modulations which are optimized for predicted channel power gains. The design BER is 0.001. The lines without a marker are for $\sigma^2_{\epsilon_p} = 0.001$ and the lines with the rings are for $\sigma^2_{\epsilon_p} = 0.1$. (From [2], copyright IEEE 2004.)*

compression methods for this feedback information are discussed in another paper in this special issue [34]. A system using the principles discussed here are described in [35] and is also part of this special issue. The adaptive schemes discussed above, are all based on the assumption that the predictor error variance is know by the rate adaptation. In practice, this parameter must also be estimated, but this problem has not yet been dealt with in the literature.

An alternative to designing the adaptive modulation scheme for the predicted SNRs is to use the rate regions which are optimum for perfectly known SNRs with predicted SNRs. This means that the regions become far from optimum when the prediction error variance becomes large. This is not a solution to recommend since the BER would significantly increase due to the nonoptimized rate regions. Some examples of BER for such cases are given in [2, Fig. 8]. For the larger predictor error variance, the BER become much higher than the design goal and in some cases the link will become useless.

V. ADAPTIVE CODED MODULATION SCHEMES

So far in this paper, we have only dealt with adaptive QAM schemes. The design methods presented can, however, be used for any set of modulations, including coded modulation schemes, as long as an invertible BER versus SNR expression exists. If no such exact expression is available, reasonably good approximations can be used in the design phase. Such an approach has been taken in [36]. In [36], the class of trellis coded modulation used in the ITU-T V.34 modem standard was used in an adaptive modulation scheme with predicted channel power gains. The conclusion is that there is a performance gain by using trellis coding instead of QAM, but the gain is relatively small. For a design BER of 0.001, the trellis coded scheme gives the same spectral efficiency as the QAM scheme at about 1 dB lower average power when the predictor error variance is small. For larger predictor error variances, the difference is even smaller. For smaller design BERs, the advantage of trellis coded modulations is somewhat larger. For a good predictor, an increase in SNR of 3–4 dB is required for QAM to meet the performance of trellis coded modulation. For a large predictor error variance, the difference is still quite small. Thus, the power gains of trellis coded modulation that we see in fixed modulations are not yet reached when these schemes are used in adaptive modulations.

We have no intention here to summarize results from other studies of adaptive coded modulation schemes. The reader is referred to [37]–[45] and other papers in this special issue for more details on other adaptive coding and modulation schemes.

VI. DISCUSSION AND CONCLUSIONS

In this paper, the basic principles of adaptive modulation has been introduced. Several optimum schemes, both when the channel is known by the transmitter and when then the predicted channel is available to the transmitter, were described. More details on these schemes can be found in [1] and [2]. From the results presented here and elsewehere, it is clear that adaptive schemes have a large performance advantage compared with fixed schemes in systems when the channel is not constant during the whole transmission. This is actually the case for most schemes. The channel variation can be due to many things, the most common being that the distance between the transmitter and receiver is varying and/or the channel experience multipath propagation. On all these channels, the adaptive scheme can ideally be designed to obtain exactly the required BER value with a generally much higher spectral efficiency than a fixed modulation link. In practice, there will be some BER variation due to errors in predicted parameters which influence the choice of modulation, but this variation is insignificant compared with the variation in fixed modulations.

But, as usual, there is no scheme with only advantages. The main disadvantages with adaptive schemes are as follows.

1) that a return channel from the receiver to the transmitter must be available for providing the channel information to the transmitter; and
2) that information on the choice of modulation and coding parameters must be transmitted on the forward link from the transmitter to the receiver in order to be able to decode the information.

In some duplex systems, where the forward and reverse links experience reciprocal channels, the feedback information does not need to be transmitted using any additional channel resources. This is because the channel can be predicted in the receiver of the uplink and this information can be used to predict the channel on the downlink. The downlink transmitter is located in the same unit as the uplink receiver and thus the information is readily available without additional channel use. This is also true in the other end of the link. There are, however, many systems where this is not the case and the feedback information has to consume channel resources.

The information that the receiver needs to know from the transmitter in order to decode the received information must almost always be transmitted on the channel. One can think of so-called blind schemes where detection takes place without this information, but then it is likely that the performance will be greatly reduced. This forward information will consume channel resources that otherwise would be used to transmit useful data. However, in most cases, this can be afforded, since the overall spectral efficiency of the link is significantly increased. At the end, the amount of bandwidth needed for both the feedback and forward information depends on the rate at which the transmitter needs to adapt to the channel and thus on the channel variability over time (or frequency for some schemes). If the channel varies too quickly, it might not be possible to use adaptive schemes. One reason is the excessive additional information which needs to be transmitted, while the other is that the predictions of channel information will be poor. However, for the majority of systems, the additional information will not pose any problem, and adaptive schemes are the natural choice. A system using adaptive modulation is presented in [35] (in this special issue). ■

REFERENCES

[1] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: A unified view," *IEEE Trans. Commun.*, vol. 49, no. 9, pp. 1561–1571, Sep. 2001.

[2] S. Falahari, A. Svensson, T. Ekman, and M. Sternad, "Adaptive modulation systems for predicted wireless channels," *IEEE Trans. Commun.*, vol. 52, no. 2, pp. 307–316, Feb. 2004.

[3] C. Köse and D. L. Goeckel, "On power adaptation in adaptive signaling systems," *IEEE Trans. Commun.*, vol. 48, no. 11, pp. 1769–1773, Nov. 2000.

[4] B. Choi and L. Hanzo, "Optimum mode-switching-assisted constant-power single- and multicarrier adaptive modulation," *IEEE Trans. Veh. Technol.*, vol. 52, no. 3, pp. 536–560, May 2003.

[5] K. M. Kamath and D. L. Goeckel, "Adaptive-modulation schemes for minimum outage probability in wireless systems," *IEEE Trans. Commun.*, vol. 52, no. 10, pp. 1632–1635, Oct. 2004.

[6] K. L. Baum, T. A. Kostas, P. J. Sartori, and B. K. Classon, "Performance characteristics of cellular systems with different link adaptation strategies," *IEEE Trans. Veh. Technol.*, vol. 52, no. 6, pp. 1497–1507, Nov. 2003.

[7] S. Chatterjee, W. Fernando, and M. Wasantha, "Adaptive modulation based MC-CDMA systems for 4G wireless consumer applications," *IEEE Trans. Consumer Electron.*, vol. 49, no. 4, pp. 995–1003, Nov. 2003.

[8] T. Ikeda, S. Sampei, and N. Morinaga, "TDMA-based adaptive modulation with dynamic channel assignment for high-capacity communication systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 2, pp. 404–412, Mar. 2000.

[9] J. S. Blogh, P. J. Cherriman, and L. Hanzo, "Dynamic channel allocation techniques using adaptive modulation and adaptive antennas," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 2, pp. 312–321, Feb. 2001.

[10] J. S. Blogh, P. J. Cherriman, and L. Hanzo, "Comparative study of adaptive beam-steering and adaptive modulation-assisted dynamic channel allocation algorithms," *IEEE Trans. Veh. Technol.*, vol. 50, no. 2, pp. 398–415, Mar. 2001.

[11] E. Armanious, D. D. Falconer, and H. Yanikomeroglu, "Adaptive modulation, adaptive coding, and power control for fixed cellular broadband wireless systems: Some new insights," in *Proc. IEEE Wireless Communications Networking Conf.*, Mar. 2003, vol. 1, pp. 238–242.

[12] T.-S. Yang, A. Duel-Hallen, and H. Hallen, "Reliable adaptive modulation aided by observations of another fading channel," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 605–611, Apr. 2004.

[13] T. Keller and L. Hanzo, "Adaptive multicarrier modulation: A convenient framework for time-frequency processing in wireless communications," *Proc. IEEE*, vol. 88, no. 5, pp. 611–640, May 2000.

[14] T. Keller and L. Hanzo, "Adaptive modulation techniques for duplex OFDM transmission," *IEEE Trans. Veh. Technol.*, vol. 49, no. 5, pp. 1893–1906, Sep. 2000.

[15] P. Xia, S. Zhou, and G. B. Giannakis, "Adaptive MIMO-OFDM based on partial channel state information," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 202–213, Jan. 2004.

[16] C.-J. Ahn and I. Sasase, "The effects of modulation combination, target BER, doppler frequency, and adaptation interval on the performance of adaptive OFDM in broadband mobile channel," *IEEE Trans. Consumer Electron.*, vol. 48, no. 1, pp. 167–174, Feb. 2002.

[17] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.

[18] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital Communications*, 3rd ed. Norwell, MA: Kluwer, 2004.

[19] I. A. Glover and P. M. Grant, *Digital Communications*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2004.

[20] J. B. Anderson, *Digital Transmission Engineering*, 2nd ed. Piscataway, NJ: IEEE Press, 1999.

[21] J. Kurzweil, *An Introduction to Digital Communications*. Hoboken, NJ: Wiley, 2000.

[22] R. E. Ziemer and R. L. Peterson, *Introduction to Digital Communication*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2000.

[23] B. Sklar, *Digital Communications: Fundamentals and Applications*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2001.

[24] S. Benedetto and E. Biglieri, *Principles of Digital Transmission With Wireless Application*. Norwell, MA: Kluwer/Plenum, 1999.

[25] A. Molisch, *Wireless Communications*. Hoboken, NJ: Wiley, 2005.

[26] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[27] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Englewood, NJ: Prentice-Hall, 2002.

[28] G. L. Stüber, *Principles of Mobile Communication*, 2nd ed. Norwell, MA: Kluwer, 2001.

[29] J. B. Anderson and A. Svensson, *Coded Modulation Systems*. Norwell, MA: Kluwer/Plenum, 2003.

[30] H. Meyr and G. Ascheid, *Synchronization in Digital Communications: Phase-Frequency-Locked Loops and Amplitude Control*. Hoboken, NJ: Wiley, 1990.

[31] H. Meyr, M. Moeneclay, and S. A. Fechtel, *Digital Communication Receivers*. Hoboken, NJ: Wiley, 1998.

[32] U. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers*. New York: Plenum, 1998.

[33] A. Duel-Hallen, "Fading channel prediction for mobile radio adaptive transmission systems," *Proc. IEEE*, vol. 95, no. 12, Dec. 2007.

[34] T. Eriksson and T. Ottosson, "Compression of feedback for adaptive modulation and scheduling," *Proc. IEEE*, vol. 95, no. 12, Dec. 2007.

[35] M. Sternad, T. Svensson, T. Ottosson, A. Ahlén, A. Svensson, and A. Brunstrom, "Towards systems beyond 3G based on adaptive OFDMA transmission," *Proc. IEEE*, vol. 95, no. 12, Dec. 2007.

[36] S. Falahati, A. Svensson, M. Sternad, and H. Mei, "Adaptive trellis-coded modulation over predicted flat-fading channels," in *Proc. IEEE Vehicular Technology Conf.*, Oct. 2003.

[37] B. Vucetic, "An adaptive coding scheme for time-varying channels," *IEEE Trans. Commun.*, vol. 39, no. 5, pp. 653–663, May 1991.

[38] S. M. Alamouti and S. Kallel, "Adaptive trellis-coded multiple phase shift keying for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 42, no. 6, pp. 2305–2314, Jun. 1994.

[39] V. K. N. Lau and M. D. MacLeod, "Variable-rate adaptive trellis coded QAM for high bandwidth efficiency applications in Rayleigh fading channels," in *Proc. IEEE Vehicular Technology Conf.*, May 1998.

[40] A. J. Goldsmith and S. Chua, "Adaptive coded modulation for fading channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 595–602, May 1998.

[41] K. J. Hole, H. Holm, and G. E. Øien, "Adaptive multidimensional coded modulation over flat-fading channels," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 7, pp. 1153–1158, Jul. 2000.

[42] V. K. N. Lau and M. D. MacLeod, "Variable rate trellis-coded QAM for flat-fading channels," *IEEE Trans. Commun.*, vol. 49, no. 9, pp. 1550–1560, Sep. 2001.

[43] D. L. Goeckel, "Adaptive coding for time-varying channels using outdated fading estimates," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 844–855, Jun. 1999.

[44] S. Vishwanath and A. J. Goldsmith, "Adaptive turbo-coded modulation for flat-fading channels," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 964–972, Jun. 2003.

[45] P. Örmeci, X. Liu, D. L. Goeckel, and R. D. Wesel, "Adaptive bit-interleaved coded modulation," *IEEE Trans. Commun.*, vol. 49, no. 9, pp. 1572–1581, Sep. 2001.

ABOUT THE AUTHOR

Arne Svensson (Fellow, IEEE) was born in Vedåkra, Sweden, on October 22, 1955. He received the M.Sc. (Civilingenjör) degree in electrical engineering from the University of Lund, Sweden, in 1979, and the Dr.Ing. (Teknisk Licentiat) and Dr. Techn. (Teknisk Doktor) degrees at the Department of Telecommunication Theory, University of Lund, in 1982 and 1984, respectively.

Currently he is with the Department of Signal and Systems at Chalmers University of Technology, Gothenburg, Sweden, where he was appointed Professor and Chair in Communication Systems in April 1993 and head of department in January 2005. Before July 1987, he held various teaching and research positions [including Research Professor (Docent)] at the University of Lund. Between August 1987 and December 1994, he was with several Ericsson companies in Mölndal, Sweden, where he worked with both military and cellular communication systems. He has a broad interest in wireless communication systems and his main expertise is on physical layer algorithms for wireless communication systems. He also has a consulting company BOCOM which offers expertise in wireless communications. He is coauthor of *Coded Modulation Systems* (Norwell, MA: Kluwer/Plenum, 2003). He has also published 4 book chapters, 34 journal papers/letters, and more than 150 conference papers.

Dr. Svensson received the IEEE Vehicular Technology Society Paper of the Year Award in 1986, and in 1984, the Young Scientists Award from the International Union of Radio Science, URSI. He was an editor of the Wireless Communication Series of IEEE Journal of Selected Areas in Communications until 2001, and is now an editor for IEEE Transactions on Wireless Communications and chief editor for a special issue on *Adaptive Modulation and Transmission* to be published in Proceedings of the IEEE late 2007. He has been a guest editor of a special issue on *Ultra-Wideband Communication Systems: Technology and Applications*, which was published 2006, and is now guest editor for a special issue on *Multicarrier Systems*, which will be published late 2007, both for EURASIP Journal on Wireless Communications and Networking. He is also guest editor of a special issue on COST289 research for Springer *Journal on Wireless Personal Communications*, which will appear in 2008. He is a member of four IEEE Societies. He was a member of the council of SER (Svenska Elektro-och Dataingenjörers Riksförening) between 1998 and 2002, and is currently a member of the council of NRS (Nordic Radio Society). Between 2002 and 2004, he was the chair of the SIBED program committee at Vinnova.