

Mixed Pseudo Analogue-Digital Speech and Audio Coding

Gemischt Pseudoanalog-Digitale
Sprach- und Audiocodierung

Von der Fakultät für Elektrotechnik und Informationstechnik
der Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades eines Doktors der
Ingenieurwissenschaften genehmigte Dissertation

vorgelegt von

Diplom-Ingenieur
Carsten Eric Johannes Hoelper
aus Korschenbroich

Berichter: Universitätsprofessor Dr.-Ing. Peter Vary
Universitätsprofessor Dr.-Ing. Stefan Heinen

Tag der mündlichen Prüfung: 3. Dezember 2010

Diese Dissertation ist auf den Internetseiten der Hochschulbibliothek online verfügbar.

AACHENER BEITRÄGE ZU DIGITALEN NACHRICHTENSYSTEMEN

Herausgeber:

Prof. Dr.-Ing. Peter Vary
Institut für Nachrichtengeräte und Datenverarbeitung
Rheinisch-Westfälische Technische Hochschule Aachen
Muffeter Weg 3a
52074 Aachen
Tel.: 0241-80 26 956
Fax.: 0241-80 22 186

Bibliografische Information der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar

1. Auflage Aachen:

Wissenschaftsverlag Mainz in Aachen
(Aachener Beiträge zu digitalen Nachrichtensystemen, Band 27)
ISSN 1437-6768
ISBN 3-86130-653-0

© 2010 Carsten Hoelper

Wissenschaftsverlag Mainz
Süsterfeldstr. 83, 52072 Aachen
Tel.: 02 41 / 2 39 48 oder 02 41 / 87 34 34
Fax: 02 41 / 87 55 77
www.Verlag-Mainz.de

Herstellung: Druckerei Mainz GmbH,
Süsterfeldstr. 83, 52072 Aachen
Tel.: 02 41 / 87 34 34; Fax: 02 41 / 87 55 77
www.Druckservice-Aachen.de

Gedruckt auf chlorfrei gebleichtem Papier

"D 82 (Diss. RWTH Aachen University, 2010)"

Acknowledgements

This thesis was written between 2005 and 2010, mostly during my time as a research and teaching assistant at the *Institut für Nachrichtengeräte und Datenverarbeitung (IND)* at the *Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen*.

First of all, I would like to express my sincere gratitude to my supervisor, Prof. Dr.-Ing. Peter Vary, whose continuous support made this work possible.

Also I would like to thank my colleagues at IND for providing a very pleasant and enjoyable working environment. For many inspiring discussions and proof-reading of the manuscript, I particularly thank Dipl.-Ing. Tobias Breddermann, Dipl.-Ing. Hauke Krüger, and Dipl.-Ing. Laurent Schmalen and I want to thank my former colleague Dr.-Ing. Marc Adrat for his valuable input especially during the initial phase of this research.

Special thanks go to the students who contributed to this work.

Last but not least, I want to thank my wife and sons, Aneliya, Maksim, and Darian, my family and my friends for supporting and encouraging me all the time.

Aachen, July 2010

Carsten Hoelper

Abstract

Current speech, audio, and video coding and transmission systems are either analogue or digital, with a strong shift from analogue systems to digital systems during the last decades for the benefit of exploiting digital channel coding for error correction. Combining both, digital and analogue schemes results in the benefit of saving transmission bandwidth, complexity, and of improving the achievable quality at any given signal-to-noise ratio on the channel within the range of interest.

The combination was achieved by transmitting pseudo analogue samples of the unquantized residual signal of a linear predictive digital filter. This principle, called Mixed Pseudo Analogue-Digital (MAD) transmission, is applied to both, narrow-band, and wideband speech, as well as to audio signals.

After introduction of the MAD transmission principle, this contribution examines the performance of the novel scheme for speech and audio transmission over a channel modelled as fading Additive White Gaussian Noise (AWGN with flat fading) with Rayleigh fading. An implementation of MAD transmission is compared to the GSM Adaptive Multi-rate speech codec mode 12.2 kbit/s (Enhanced Fullrate Codec, EFR), which uses a comparable transmission bandwidth if channel coding is included.

The simulative results are backed by a thorough information theoretical analysis of the principles used in MAD transmission, pointing out that the increased performance mainly stems from the combination of digitally transmitting the spectral envelope of the signal while at the same time the Gaussian residual signal is the optimum input for the AWGN channel.

Modulation schemes using the Archimedes Spiral for mapping the pseudo analogue residual to a 2-dimensional signal space are theoretically motivated and developed to enhance the quality of the basic system.

Finally, possible applications like MAD microphones and headsets are suggested and further prospects like channel adaptive MAD are briefly given.

Kurzfassung

Heutige Systeme zur Codierung und Übertragung von Sprach-, Audio- und Videodaten sind entweder digital oder analog, wobei es in den letzten Jahrzehnten einen starken Trend von analogen hin zu digitalen Systemen gegeben hat, die mittels Kanalcodierung Übertragungsfehler ausgleichen können. Eine Kombination von digitaler und analoger Welt spart sowohl Übertragungsbandbreite als auch Komplexität, während die erzielbare Qualität im interessanten Bereich der Signal-Rausch-Abstände auf dem Kanal gesteigert wird.

Die Kombination wird hier erreicht, indem die pseudo-analogen Abtastwerte des Restsignals nach linearer Prädiktion unquantisiert übertragen werden. Dieses Prinzip, gemischt pseudoanalog-digitale Übertragung (MAD: Mixed Pseudo Analogue-Digital transmission), wird auf Telefonsprache, breitbandige Sprache und Audioübertragung angewendet.

Die vorliegende Arbeit untersucht die Leistungsfähigkeit des MAD Übertragungssystems anhand von Sprach- und Audioübertragung über einen AWGN Kanal mit flachem Rayleigh Fading. MAD Übertragung wird mit dem Adaptiven Multiraten-codec im 12.2 kbit/s Modus (Enhanced Fullrate Codec) verglichen, der mit Kanalcodierung eine vergleichbare Bruttobandweite auf dem Kanal benötigt.

Die simulativen Ergebnisse werden mit einer informationstheoretischen Betrachtung hinterlegt. Es wird gezeigt, dass nach digitaler Codierung der spektralen Einhüllenden ein Gaussförmiges Restsignal zu übertragen bleibt, welches ideal an den AWGN Kanal angepasst ist.

Modulationsschemata, die mit der Archimedesspirale das pseudoanaloge Restsignal in einen zweidimensionalen Signalraum überführen, werden vorgestellt und theoretisch beleuchtet um die Qualität des MAD Systems weiter zu steigern.

Schließlich werden Anwendungsszenarien wie ein MAD Mikrophon, Hörgerät oder Headset umrissen und es wird ein kurzer Ausblick auf kanaladaptive MAD Übertragung gegeben.

Contents

1	Introduction	1
1.1	Analogue Transmission	2
1.2	Digital Transmission	2
1.3	Mixed Transmission	3
2	Fundamentals	5
2.1	Source Coding	5
2.1.1	Audio Bandwidth	6
2.1.2	Linear Prediction	6
2.1.3	Residual Coding	8
2.2	Channel Coding	11
2.2.1	Block Codes	12
2.2.2	Convolutional Codes	12
2.2.3	AMR Mode 12.2 kbit/s Error Correction	12
2.3	Models for Transmission	13
2.3.1	Analogue Schemes	15
2.3.2	Digital Schemes	16
2.4	Quality Evaluation	17
2.4.1	Perceptual Evaluation of Speech Quality (PESQ)	18
2.4.2	Perceptual Evaluation of Audio Quality (PEAQ)	18
3	Mixed Pseudo Analogue-Digital Speech & Audio Transmission	21
3.1	<i>Pseudo Analogue</i>	22
3.2	State-Of-The-Art of Analogue-Digital Transmission	22
3.2.1	Pseudo-Analog Speech Transmission in Mobile Radio Communication Systems	22

3.2.2	Hybrid Digital-Analog Coding	23
3.3	Principles of MAD Coding	24
3.3.1	Processing in the Digital Domain	25
3.3.2	Baseband Transmission Model	29
3.3.3	Transmission over Band Pass AWGN Channels	31
3.3.4	Transmission over Rayleigh Fading Channels	35
3.3.5	Scalability of the MAD Transmission Scheme	36
4	MAD Modulation Schemes	39
4.1	Digital Information	39
4.2	Pseudo Analogue Information	40
4.2.1	Mapping to the Unit Circle	41
4.2.2	Pulse Amplitude Modulation (PAM)	46
4.2.3	Quadrature Pulse Amplitude Modulation (QPAM)	47
4.2.4	Archimedes Spiral Mapping (ASM)	47
5	Information Theoretic Evaluation of MAD Transmission	59
5.1	Introduction to Information Theory	59
5.1.1	Entropy, Mutual Information	60
5.1.2	Channel, Channel Capacity	61
5.1.3	Channel Capacity of an AWGN Channel for Gaussian or Binary Input	61
5.1.4	Capacity of Multiplexed Channels	65
5.2	Rate Distortion Theory	68
5.3	Performance Limit	71
5.4	Information Theoretic Comparison to Digital Transmission	72
5.4.1	Simplified Transmission Model	73
5.4.2	Rate Split between LPC and Residual Signal	73
5.4.3	Analysis of the MAD Principle	75
5.4.4	Analysis of MAD using ASM Transmission	98
6	Application of MAD Transmission	115
6.1	MAD Speech Transmission	115
6.2	MAD Audio Transmission	115
6.2.1	MAD Wireless Microphones, Headsets or Hearing Aids with Wireless Audio Input	116
6.3	Further Studies & Prospects	117
6.3.1	Pseudo Analogue Channel Coding	117
6.3.2	Channel Adaptive MAD	117

7	Summary & Conclusions	119
7.1	MAD Transmission	120
7.2	Modulation for MAD Transmission	120
7.3	Information Theoretic Evaluation	121
7.4	Implementation	122
7.5	Application	122
A	The Spectral Flatness Measure	125
B	The Complementary Error Function	127
C	The Filter $H_s(z)$	129
D	Deutschsprachige Zusammenfassung	131
D.1	Grundlagen	131
D.2	Gemischt pseudoanalog-digitale Übertragung	132
D.3	Modulation bei MAD-Übertragung	134
D.4	Informationstheoretische Betrachtung	136
D.5	Anwendung	141
	List of Abbreviations & Mathematical Symbols	143
	Bibliography	149

1

Introduction

Today's speech, audio, and video coding and transmission systems are either analogue or digital, with a strong shift from analogue systems to digital systems during the last two decades.

Looking back in the history of communications, it is very interesting to note that the very first system, i.e. the telegraph, demonstrated by Joseph Henry in 1832 and Samuel F. B. Morse in 1838, was a truly digital system. The widespread Morse code, e.g., uses six discrete states: dot, dash, inter-character gap between dots and dashes, short gap between letters, medium gap between words and long gap between sentences.

Only later, with the invention of the telephone by Johann Philipp Reis around 1860 and with the first radio transmission by Alexander Popov, analogue communications started to arise. During most of the last century, analogue communications played the major role in personal communications (telephone) and broadcast radio and television.

Although the theoretical principles had already been formulated by Shannon in 1948 [86], it took until the end of the 20th century, when the *Groupe Spécial Mobile* decided for a digital pan-European mobile telephone system [70]. The change from analogue to digital started to affect most communication systems from Personal Mobile Radio (PMR), via emergency services radio (e.g., TETRA, Tetrapol), broadcast radio (DAB), broadcast TV (DVB), as well as storage and recording devices (CD, DVD, digital photography, etc). This time is often referred to as the *Digital Revolution*.

This contribution presents a novel transmission concept, that integrates both, analogue and digital communication links, for the benefit of low transmission bandwidth, low complexity, and high output quality. While an improvement of speech transmission in terms of complexity and usage of transmission bandwidth was the main driver for the invention of *Mixed Pseudo Analogue-Digital (MAD) Transmission*, as the new system is termed, the principles are generalized for audio signals and are valid for other types of signals, as well. Another application, where the

mixing of analogue and digital parts is already commonplace, would be hardware design, where analogue and digital circuits are placed on the same chip [89].

In the following, the main aspects of analogue, digital, and mixed transmission will be revisited.

1.1 Analogue Transmission

Analogue transmission is characterized by the fact that the information signal as well as the modulated signal are continuous in time and amplitude. For transmission, the analogue signals modulate a carrier signal. Modulating means to vary the properties of the carrier signal. Amplitude Modulation (AM) means alteration of the amplitude of the carrier signal, Phase Modulation (PM) alteration of the phase and Frequency Modulation (FM) alternation of the frequency.

Disadvantages of analogue systems are the sensitivity with respect to noise and the propagation loss. Noise sensitivity can be reduced within FM by using a higher bandwidth B for transmission and thus, exchanging bandwidth for signal-to-noise ratio $\frac{S}{N}$ according to Shannon's famous formula for the channel capacity C

$$C = B \cdot \log\left(1 + \frac{S}{N}\right). \quad (1.1)$$

Another reason for FM systems to be most widespread is, that with constant transmission power, the receivers can be very cheap and still eliminate amplitude variations which occur during transmission.

1.2 Digital Transmission

In contrast to analogue signals, digital signals are discrete in time and amplitude. Sampling the frequency band limited analogue signal at discrete intervals in time does not introduce any loss of information, as long as the sampling theorem [46] is fulfilled. Quantization of the amplitude to discrete levels, however, inevitably and irreversibly changes the signal, which has two implications:

- a) The quantization must be fine enough, not to notably degrade the signal
- b) Contrasting the (theoretically) infinite resolution of an analogue signal, the resolution of a digital signal is always limited.

The conversion from analogue to digital representation is illustrated in Figure 1.1.

The main advantage of digital transmission stems from the limited number of possible amplitudes (quantization levels), which can be represented digitally, allowing for certain degradations not only to be detected, but also to be corrected. Thus, a regenerator can re-establish an exact copy of the original signal before re-transmission, allowing for principally error-free transmission, if the applied channel coding is strong enough. Strong channel coding, however, comes at the expense of increased complexity, transmission rate (due to the added redundancy), and delay.

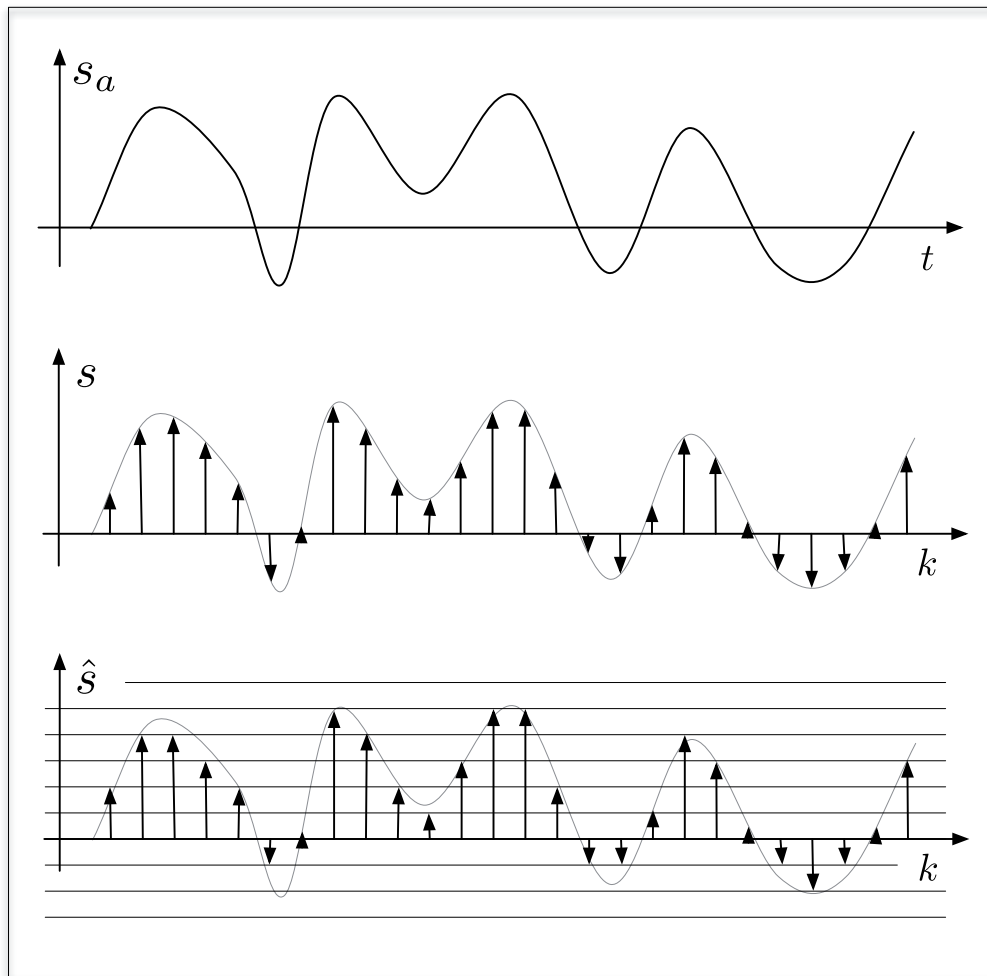


Figure 1.1: A/D conversion. Top: Analogue signal. Middle: Sampling. Bottom: Quantization.

1.3 Mixed Transmission

Shamai, Verdú, and Zamir have shown that, if source and channel are Gaussian and the analogue channel is transmitted in Single Sideband, its bandwidth / power can be used as efficiently as if a completely digital system were designed from scratch [52]. Analogue speech, audio, and video transmission systems, though, suffer badly from high transmission noise, while digital systems can completely recover the quantized source signal as long as the channel coding applied is strong enough and the received energy per bit is sufficient.

With increasing channel SNR (Signal-to-Noise Ratio), however, as soon as all bits are decoded correctly, the output quality of a digital system remains constant. The output quality is limited by the source coder and quantizer design. For analogue systems, the minimum bandwidth B'_{ana} required for the transmission of a speech or audio signal equals the audio bandwidth B'_{audio} , see e.g. [68]. Digital systems usually require a significantly higher bandwidth than analogue systems.

Generally speaking, the advantages and disadvantages of both worlds can be summarized as follows:

- Analogue Transmission
 - + Low channel bandwidth
 - Highly sensitive to additive noise
- Digital Transmission
 - + Robust channel coding against transmission errors
 - higher bandwidth
 - higher complexity
 - limited maximum quality defined by the speech/audio codec design

Mixed Pseudo Analogue-Digital (MAD) Transmission aims at combining the advantages of both paradigms with the main target being the need of low transmission bandwidth. Especially, MAD utilizes the fact that after linear prediction the normalized residual of speech and audio signals has a Gaussian distribution which is optimal for transmission over an AWGN channel.

The remainder of this thesis is structured as follows: Chapter 2 introduces the fundamentals. Chapter 3 gives a brief review of existing analogue-digital systems and presents the proposed Mixed Pseudo Analogue-Digital (MAD) transmission system in detail. Modulation schemes which can be used in MAD transmission are investigated in chapter 4. Chapter 5 gives a thorough theoretical motivation of the MAD transmission system and analyzes the optimal use of Archimedes Spiral Mapping (ASM) for modulation of the pseudo analogue residual. Applications of MAD are studied in Chapter 6 as well as providing a brief outlook to possible future work. In Chapter 7, the main concepts and conclusions of this work are summarized.

2

Fundamentals

Figure 2.1 gives an overview of a general transmission scheme. Within this chapter, relevant techniques in the blocks source coding, channel coding, and modulation shall be revisited to facilitate the understanding of the concepts described in the following chapters.

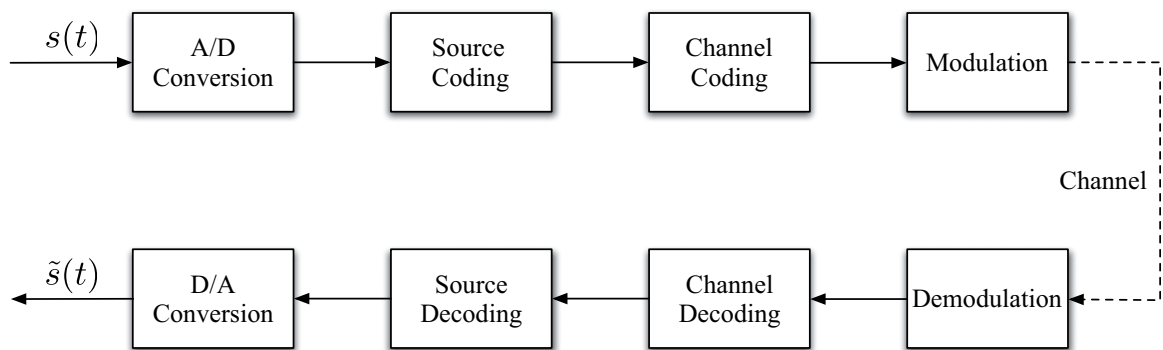


Figure 2.1: General scheme of digital transmission of analogue signals.

2.1 Source Coding

The source signal, speech and audio are in focus here, usually is an analogue signal that needs to be digitized for further processing. With the uncoded PCM (Pulse Code Modulation) samples having a relatively high bit rate, the point of source coding is to reduce the rate with lossless or lossy coding schemes before transmission. Source coding in general can be carried out in time as well as in frequency domain [72]. Here, time domain schemes, especially *Linear Predictive Coding (LPC)* will be regarded.

2.1.1 Audio Bandwidth

According to the sampling theorem [46] the largest possible frequency within a time discrete signal with sampling frequency f_s is

$$f_{max} = \frac{f_s}{2}. \quad (2.1)$$

Current fixed and mobile communication systems operate with narrow audio bandwidth limited to 300 - 3400 Hz, often referred to as telephone bandwidth in literature. The introduction of a wider audio bandwidth of 50 - 7000 Hz provides substantially improved speech quality, naturalness (low frequencies) and intelligibility (higher frequencies).

Within this contribution the following sampling rates are regarded:

- $f_s = 8$ kHz for narrowband speech, according to an audio bandwidth of 300 - 3400 Hz.
- $f_s = 16$ kHz for wideband speech or audio signals according to an audio bandwidth of 50 - 7000 Hz.
- $f_s = 32$ kHz for audio signals according to an audio bandwidth of up to 16000 Hz.
- $f_s = 48$ kHz for high quality audio signals according to an audio bandwidth of up to 24000 Hz.

Spectrograms of narrowband speech, wideband speech, and 32 kHz (of the same speech signal) are illustrated in Figure 2.2.

2.1.2 Linear Prediction

Linear Prediction (LP), e.g. [27, 74], is very effective to code speech signals and is used in almost all current speech coding standards as an autoregressive filter to model the vocal tract. The principle idea of linear predictive coding is to exploit correlation immanent to the input signal. For short-term block adaptive linear prediction, a windowed segment of the input signal is analyzed in order to obtain the coefficients of a filter $a_1 \dots a_N$ (LP filter order N) which minimize the energy of the difference between original signal $s(k)$ and residual signal $r(k)$, see Figure 2.3. The transfer function of the general linear prediction analysis filter is

$$1 - A(z) = 1 - \sum_{i=1}^N a_i \cdot z^{-i}. \quad (2.2)$$

It exploits the short-term correlation in form of the dependency of the actual sample $s(k)$ from the past N samples:

$$s(k) = \tilde{s}_{LP}(k) + r(k) = \sum_{i=1}^N a_i s(k-i) + r(k) \quad (2.3)$$

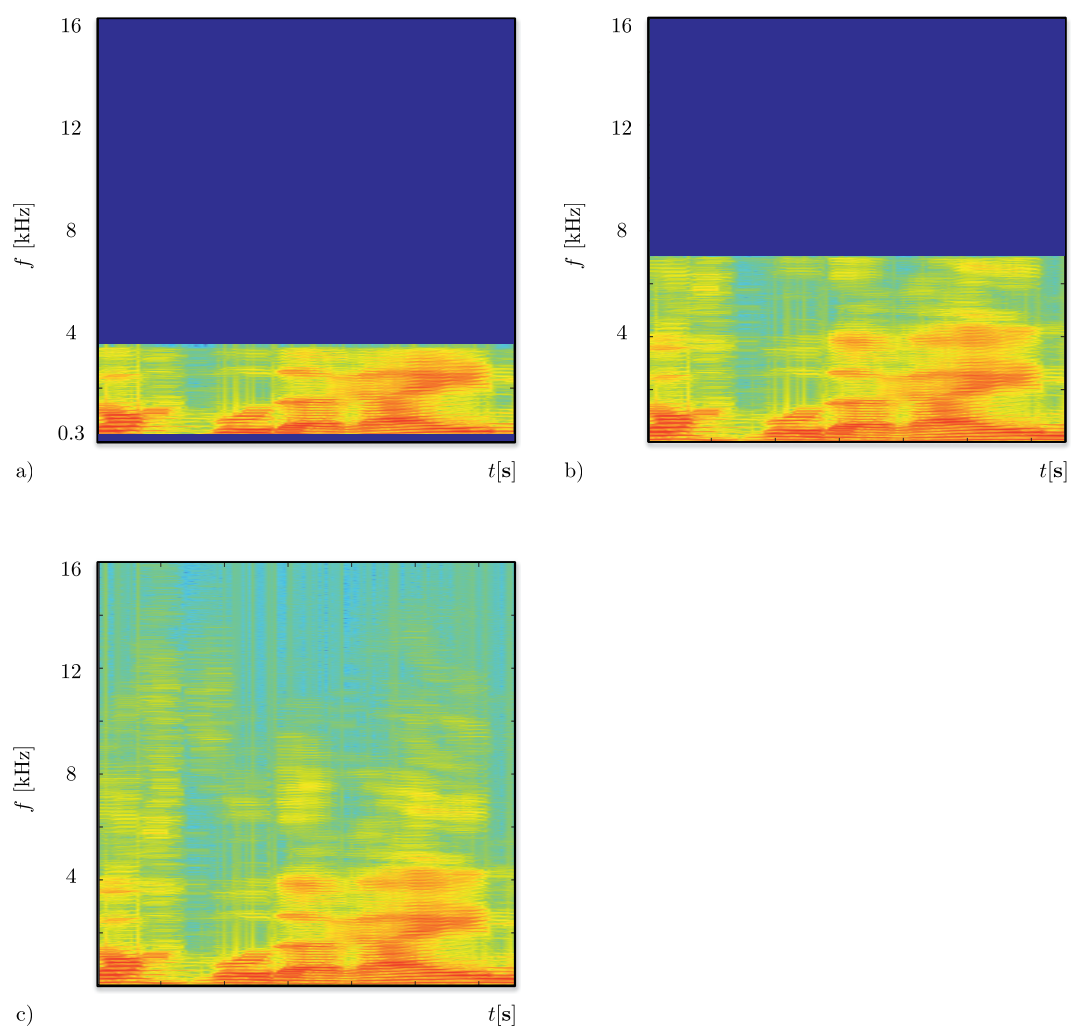


Figure 2.2: Audio bandwidth of a) narrowband speech, b) wideband speech, and c) 32 kHz speech.

with the LP estimate $\tilde{s}_{LP}(k)$ and the decorrelated residual signal (prediction error) $r(k)$.

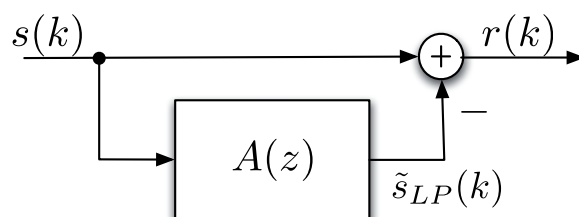


Figure 2.3: Principle of Linear Prediction.

Different algorithms exist to calculate the optimal filter coefficients. Most widespread is the autocorrelation method in combination with the Levinson Durbin algorithm [72].

The quantization of LPC parameters has been studied in detail, e.g., in [32, 74].

2.1.3 Residual Coding

In addition to the LP parameters, a quantized version $\hat{r}(k)$ of the prediction residual $r(k)$ needs to be transmitted. Figure 2.4 shows the generalized scheme of linear predictive transmission with a gain to normalize the power of the residual and an arbitrary quantizer to describe the shape of the normalized residual signal.

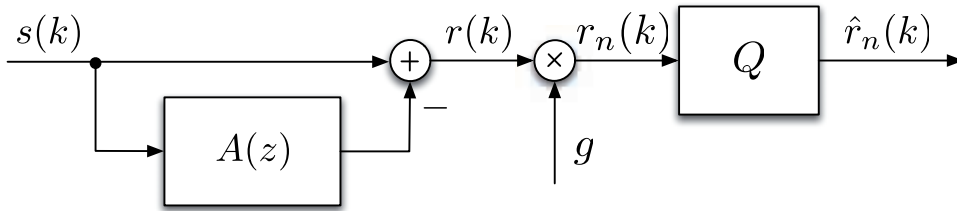


Figure 2.4: Principle of linear predictive transmission. LP coefficients, gains, and quantizer indices need to be transmitted.

Different approaches for quantization and transmission of the residual are known in literature, e.g., [72, 78, 91]. With the residual of a speech signal, usually the quantization is carried out in two steps. First, Long Term Prediction (LTP) is applied and then the LTP residual is quantized. The LTP filter

$$H_{LTP}(z) = 1 - P_{LTP}(z) = 1 - \beta \cdot z^{-\tau}, \quad (2.4)$$

with the gain β and pitch lag τ removes the pitch structure in voiced speech. Figure 2.5 shows the combined LP and LTP transmission scheme.

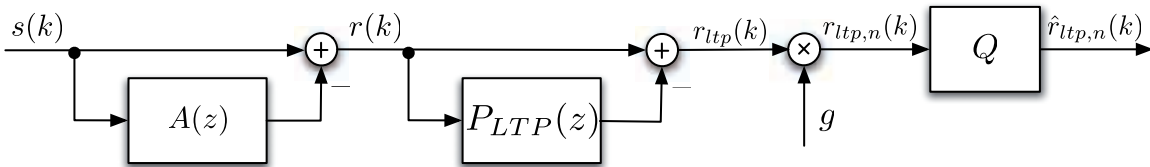


Figure 2.5: Principle of linear predictive transmission with long term prediction.

The most common approaches to transmit the residual are Residual-Excited Linear Prediction (RELP) and Code Excited Linear Prediction (CELP). They will shortly be introduced in the following.

REL P

Residual-Excited Linear Prediction (REL P) has first been introduced in [16, 58]. For transmission, the residual signal $r_{ltp}(k)$ of the long term predictor is low-pass filtered and down-sampled. At the receiver, the excitation is formed by up-sampling the received low-pass residual without applying the usual anti-aliasing filter, see Figure 2.6.

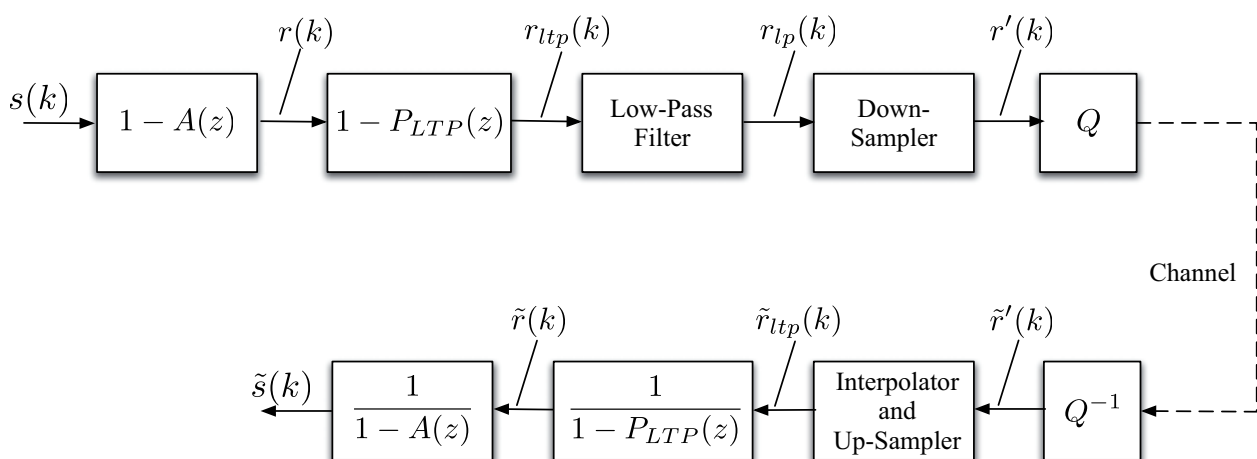


Figure 2.6: Principle of Residual-Excited Linear Prediction with Long Term Prediction.

As an effect, the higher frequency bands are mirrored copies of the lower band. Following the assumption that the LTP residual is spectrally flat, this is a sensible estimate of the true excitation signal. While REL P based speech codecs do not reach the perceptual quality of current CELP codecs, they have a significantly lower computational complexity. The original speech codec of the successful GSM mobile telephone system (GSM Full Rate) is a REL P derivative [24]. The GSM Full Rate speech codec has a bit rate of 13 kbit/s. 1.8 kbit/s are used to transmit the LP filter coefficients, the 11.2 kbit/s assigned to the residual split into 1.8 kbit/s for the LTP, 1.2 kbit/s for gains and 8.2 kbit/s for the quantizer indices q_i .

Code Excited Linear Prediction

Code Excited Linear Prediction (CELP) uses two codebooks to model the LTP synthesis filter (Adaptive Codebook, ACB), and the residual at the output of the LTP analysis filter (Stochastic, or Fixed Codebook, FCB), respectively. To determine the optimal excitation sequence, the weighted mean square error between the original signal and the synthesized sequences corresponding to the different codebook entries is evaluated. Due to the structure which incorporates the synthesis part within the analysis part, the term *Analysis-by-Synthesis Principle* is used. The weighting filter usually follows

$$W(z) = \frac{1 - A(z/\gamma_1)}{1 - A(z/\gamma_2)} \quad (2.5)$$

with γ_1 and γ_2 depending on the respective codec and transmission rate. A block diagram of the CELP codec is given in Figure 2.8.

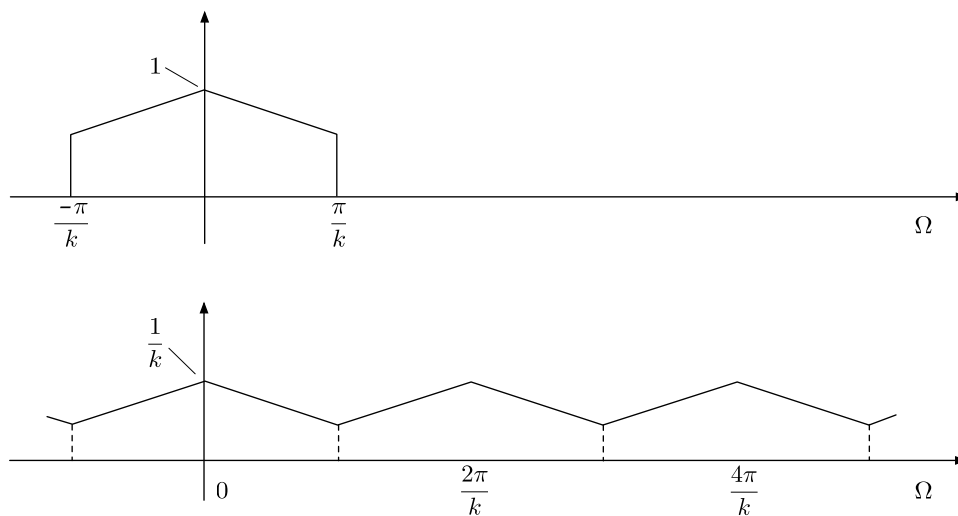


Figure 2.7: Generation of the excitation signal in the RELP decoder.

The Analysis-by-Synthesis Principle is the main reason for CELP codecs to have a much higher computational complexity than, e.g., RELP codecs. In their original description, Schroeder and Atal [21] conclude “The coding procedure is computationally very expensive; it took 125 sec of Cray-1 CPU¹ time to process 1 sec of the speech signal”. As an example, while the original GSM Full Rate speech codec [99] - based on the RELP principle - has a complexity of about 3.0 wMOPS (weighted Million Operations Per Second) [70], the later GSM Enhanced Full Rate speech codec [11, 101] has a complexity of up to 18.1 wMOPS [100] making use of structural codebooks and refined search algorithms which only need a fraction of the computational complexity needed for a full search of all possible codebook entries.

The GSM Enhanced Full Rate speech codec has a bit rate of 12.2 kbit/s. 1.9 kbit/s for the LP filter coefficients, 2.3 kbit/s for the LTP, 1.0 kbit/s for gains and 7 kbit/s for the FCB indices q_i . It is important to note that in the Enhanced Full Rate speech codec most of the bit rate and complexity is used for transmission of the residual, while the LP part only needs 2.75 wMOPS (of 18.1 wMOPS).

¹Central Processing Unit

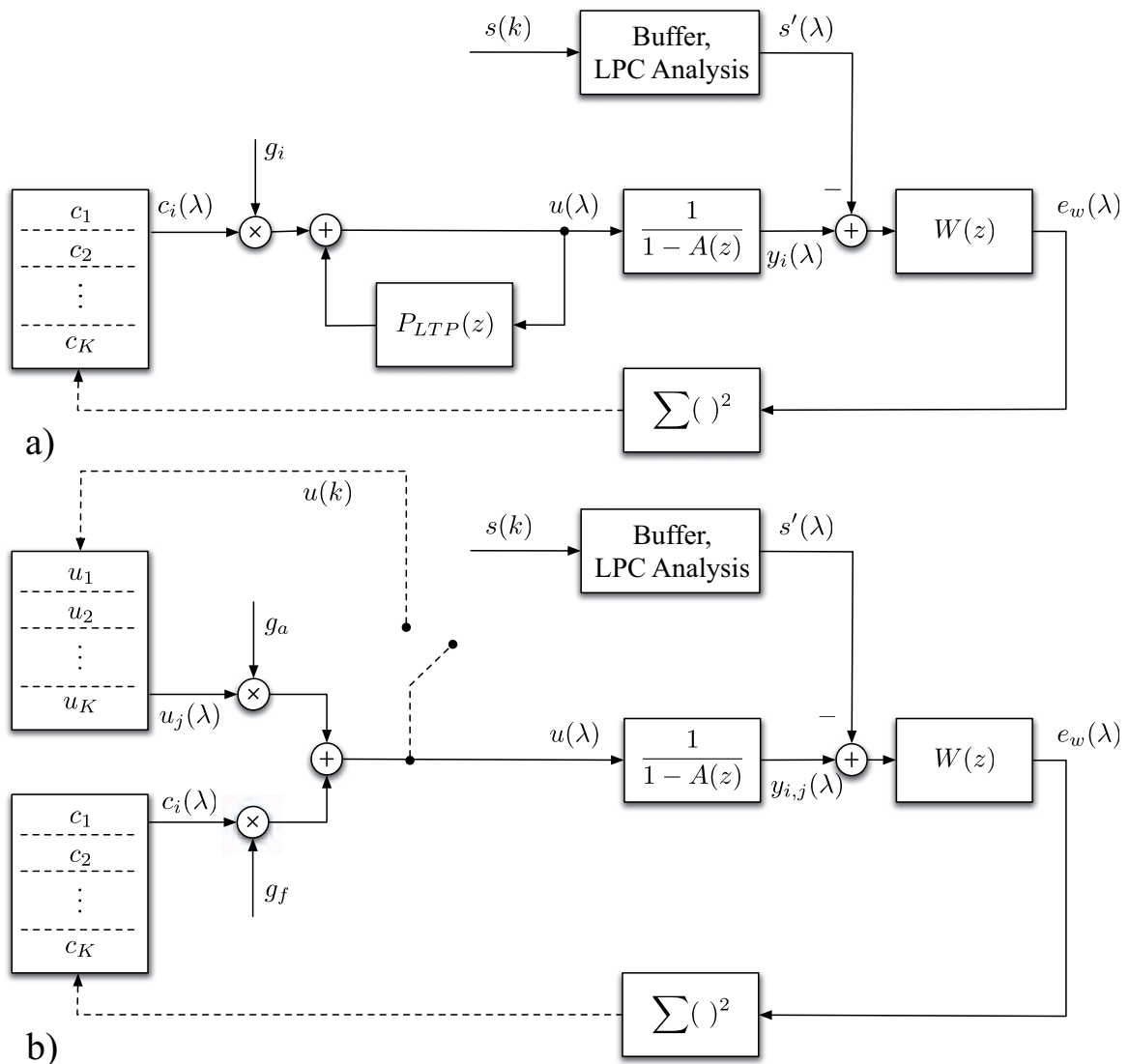


Figure 2.8: CELP encoder with LTP. a) Conventional realization of the LTP synthesis filter. b) Realization of the LTP loop by means of an adaptive codebook (ACB).

MAD transmission, as introduced in Section 3.3, offers a novel approach for transmitting the residual as normalized pseudo analogue samples. Not only does this scheme reduce the coding complexity and symbol rate (and thus the transmission bandwidth), it is also suited for audio signals as well as for speech signals due to the absence of an LTP filter.

2.2 Channel Coding

Channel coding, or error control coding, improves the communications performance if the transmitted signal is degraded by noise, interference, or fading. Channel

coding comprises both, error detection and error correction [87]. Known schemes are block codes, convolutional codes or Turbo codes.

2.2.1 Block Codes

In the case of block codes, a fixed block of k source bits (also called information bits) is transformed into a larger block of n bits called code bits. The $(n - k)$ bits which the encoder adds are called redundancy bits, parity bits, or check bits. They allow the decoder to calculate an error syndrome, which indicates whether an error occurred and if so, where.

2.2.2 Convolutional Codes

Convolutional codes do not operate on blocks of samples. Instead, they use a shift register to convolve each message bit m_k with $K - 1$ bits from the past to form the coded bit b_c . The parameter K is called the constraint length. A typical convolutional encoder is depicted in Figure 2.9.

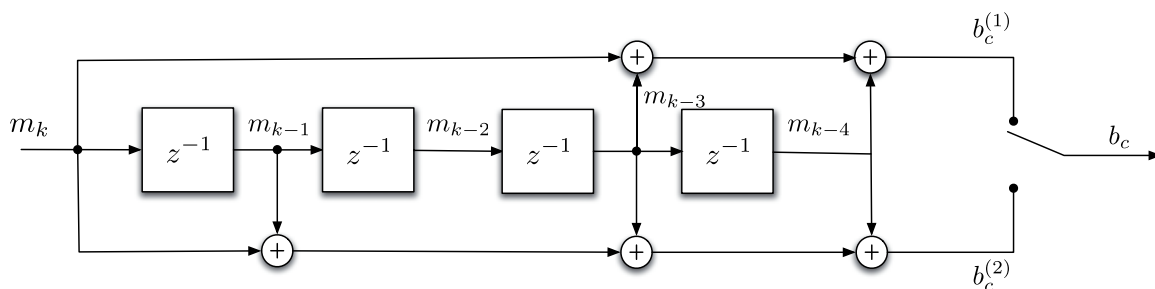


Figure 2.9: Convolutional encoder as used in GSM.

On the receiver side, the Viterbi Algorithm [60] is used to determine the most likely transmitted sequence. A list of codes of rate $\frac{k}{n} = \frac{1}{2}$, $K = 3$ to 9, and $\frac{k}{n} = \frac{1}{3}$, $K = 3$ to 8 which have the maximum free distance was compiled by Odenwalder [80].

2.2.3 AMR Mode 12.2 kbit/s Error Correction

The AMR mode 12.2 kbit/s speech encoder outputs a block of 244 information bits every 20 ms. The bits are re-ordered corresponding to their individual relevance for the speech decoder before channel coding is applied². The 78 Least Significant Bits (LSB, termed class 2 bits) will not receive any channel coding.

CRC-Check

An 8-bit Cyclic Redundancy Check (CRC) code is applied to the first 65 information

²The technique of selecting the channel coding by relevance of the respective bits is termed *Unequal Error Protection (UEP)* in literature. The idea of UEP has been extended to modulation schemes by T. Brüggem in [64].

bits for error detection. 4 bits, which are the most important bits within class 2, are repeated twice. The 182 bits of class 1 (which will receive channel coding) are again split into 50 most important bits (termed class 1a) and 132 bits of class 1b. The bits of class 1a are protected by 3 parity bits for error detection according to a shortened cyclic (53,50,2) block code, using the generator polynomial

$$g(D) = D^3 + D + 1. \quad (2.6)$$

Finally 4 zeros are added to allow for termination of the following convolutional code.

Convolutional Encoder

The class 1 bits are encoded with the 1/2 rate convolutional code defined by the polynomials

$$G_0 = 1 + D^3 + D^4 \quad (2.7)$$

and

$$G_1 = 1 + D + D^3 + D^4. \quad (2.8)$$

This encoder is the one depicted in Figure 2.9.

2.3 Models for Transmission

For transmission, the (possibly channel coded) source signal must be modulated to a carrier signal. The signal which is used for modulation is often referred to as baseband signal, as its spectral components have lower frequencies than those of the modulated signal. Using a sinewave as a carrier signal, the parameters which can be altered for modulation are:

- Amplitude A ,
- frequency f ,
- phase angle φ .

Using pulses as carrier, the parameters are:

- Amplitude A ,
- cycle duration T_p ,
- phase angle φ ,
- pulse duration T_i .

For the sake of a fair comparison of the MAD transmission scheme with a typical digital speech codec, for both approaches a (complex valued) baseband channel with additive noise and multiplicative fading is assumed. Therefore, only amplitude and phase modulation of a single carrier with constant frequency will be considered. The resulting term for the modulated signal $u(t)$ is

$$u(t) = A(t) \cdot \cos(\omega_0 t + \varphi(t)) \quad (2.9)$$

with constant carrier frequency ω_0 .

This term can also be written as

$$u(t) = A(t) \cdot \cos(\omega_0 t + \varphi(t)) \quad (2.10)$$

$$= A(t) \cdot \cos(\varphi(t)) \cdot \cos(\omega_0 t) - A(t) \cdot \sin(\varphi(t)) \cdot \sin(\omega_0 t) \quad (2.11)$$

$$= x_I(t) \cdot \cos(\omega_0 t) - x_Q(t) \cdot \sin(\omega_0 t). \quad (2.12)$$

x_I and x_Q are called inphase and quadrature component, respectively. The general modulation scheme is referred to as quadrature modulation.

For modulation, $K \geq 1$ bits (digital modulation) or values (analogue modulation) r_λ of duration T_b are mapped to a complex symbol $d(k)$ of duration $T = K \cdot T_b$ (signal space mapping)

$$d(k) = d'(k) + jd''(k). \quad (2.13)$$

The complex symbols $d(k)$ are then filtered with the pulse shaping filter G to obtain the inphase and quadrature components, see Figure 2.10.

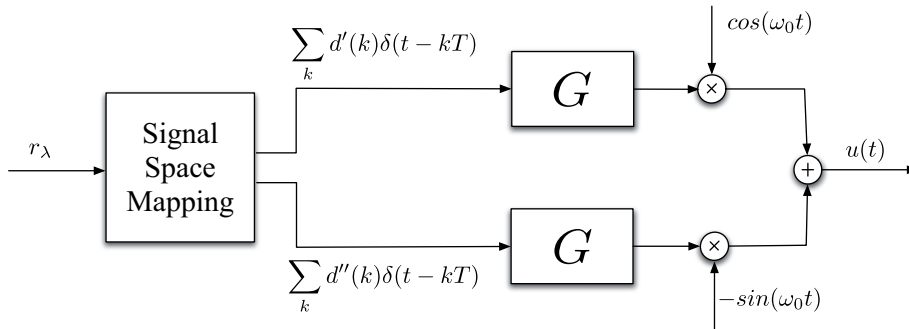


Figure 2.10: Quadrature modulation with pulse shaping filter.

The inphase and quadrature components can be combined in a complex signal $e_T(t)$, see Figure 2.11.

$$e_T(t) = x_I(t) + jx_Q(t) \quad (2.14)$$

The complex signal $e_T(t)$ is known as complex envelope of the transmitter signal $u(t)$.

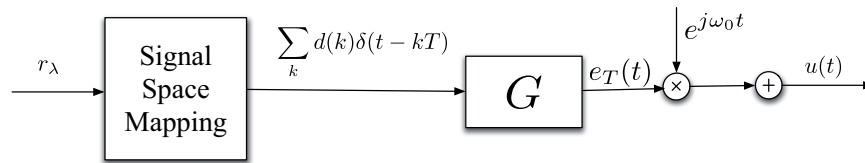


Figure 2.11: Quadrature modulation with pulse shaping filter, complex notation.

2.3.1 Analogue Schemes

Amplitude Modulation

The modulation of the amplitude of the carrier can be expressed as

$$u_{AM}(t) = A(t) \cos(\omega_0 t) \quad (2.15)$$

with the initial phase φ_0 set to zero. $A(t)$ can either be directly proportional to the signal $r_a(t)$ or be added to a constant according to

$$A(t) = A_0 + r_a(t) \quad (2.16)$$

with A_0 being the amplitude of the unmodulated carrier.

Figure 2.12 shows exemplary spectra of a baseband signal $R(\omega)$ and the respective modulated signal $R_{AM}(\omega)$ for $A_0 = 0$.

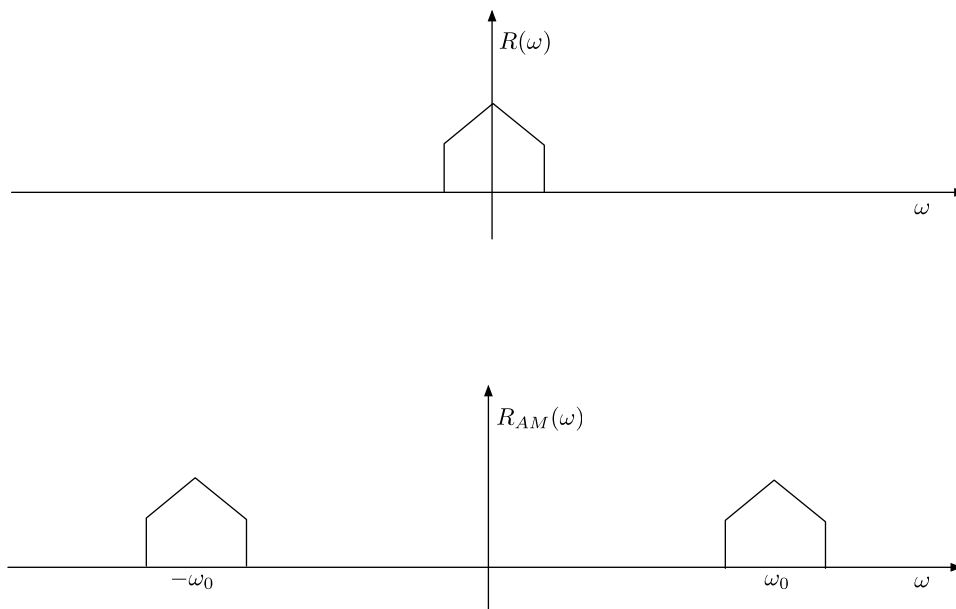


Figure 2.12: Signal spectrum $R(\omega)$ and AM spectrum $R_{AM}(\omega)$.

As the information is held in both sidebands, this kind of AM modulation is called double-sideband AM. Double-sideband AM has a transmission bandwidth double as wide as that of the source signal. AM is continuous in time and amplitude.

Pulse-Amplitude Modulation

PAM is the time discrete version of AM. In the general quadrature modulation scheme, samples of the signal $r(k)$ are mapped to the complex symbols $d(k)$.

$$\begin{aligned} d'(k) &= A_0 + r(k) \\ d''(k) &= 0 \end{aligned} \tag{2.17}$$

or in case of quadrature PAM

$$\begin{aligned} d'(k') &= A_0 + r(k) \\ d''(k') &= A_0 + r(k + 1). \end{aligned} \tag{2.18}$$

2.3.2 Digital Schemes

The digital schemes regarded are the most widespread, Amplitude Shift Keying (ASK), and Phase Shift Keying (PSK).

Amplitude Shift Keying

While PAM is the time discrete version of AM, ASK is not only time discrete, but it also uses discrete amplitudes $\hat{A}(k)$. In case of binary $\hat{A}(k)$ with elements $\hat{A}(k) \in \{0, 1\}$ this scheme is also called on-off-keying. Generalized for M -ary signals (signals with M discrete levels) $r(k)$, ASK can be regarded as PAM with quantized amplitudes, compare Figure 2.13.

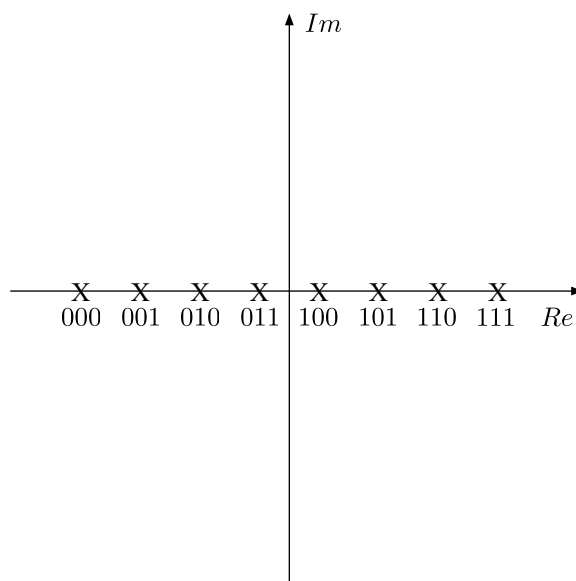


Figure 2.13: Signal constellation of 8-ASK.

Phase Shift Keying

In PSK the phase of the carrier signal is altered according to the discrete source signal $s(k)$. The signal space mapping ensures that allowed signal points have maximum distance from each other, see Figure 4.1 for the signal sets of Binary PSK (BPSK), Figure 4.2 for Quadrature PSK (QPSK), and Figure 2.14 for 8-PSK. Depending on the application, the mapping of symbols to the signal set can be different, Gray mapping, e.g., aims at minimizing the number of bit errors if a neighbouring point in the signal space is decoded.

Modulation used for MAD transmission is studied in more detail in Chapter 4.

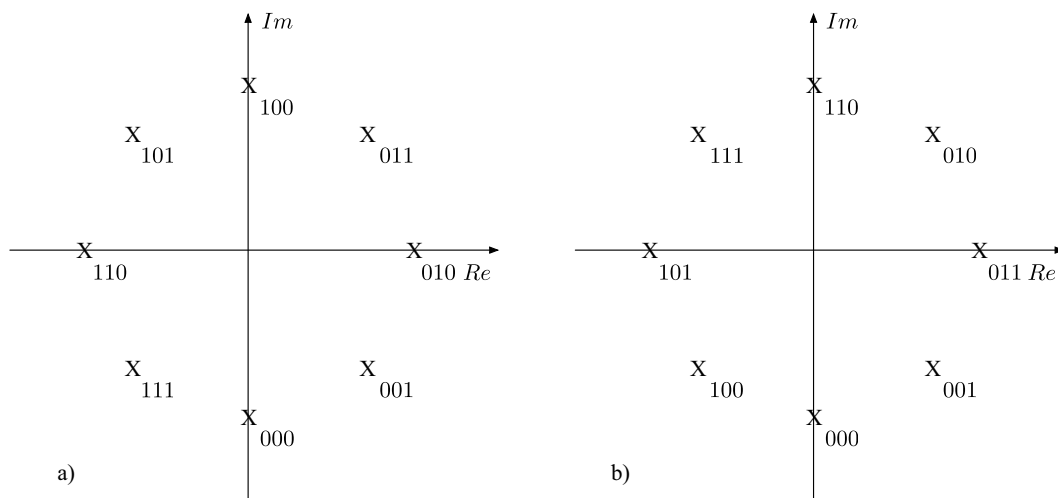


Figure 2.14: Signal constellation of 8-PSK. a) regular (natural binary) mapping, b) Gray mapping.

2.4 Quality Evaluation

The subjectively perceived quality of speech and audio signals should be the design goal for any codec or transmission system which delivers a signal to the human ear. Unfortunately, automated measuring of the perceptual quality is not straight forward, which is why mostly instrumental quality measures such as the SNR, MSE³, or the Itakura-Saito distance [38], are used in codec design.

To evaluate the perceptual quality of speech signals, standardized listening tests have been developed [108] to get an average impression amongst numerous test subjects. The absolute quality is categorized between 1 (very bad) and 5 (excellent) [109]. The average result is termed Mean Opinion Score (MOS).

As listening tests are time consuming and expensive, instrumented measures to estimate the outcome of a listening test have recently been developed that incorporate numerous features of the signal. The measures briefly introduced here are intrusive, that is the undisturbed original is needed for comparison to the distorted signal.

³Mean Square Error

2.4.1 Perceptual Evaluation of Speech Quality (PESQ)

Targeting the estimation of speech quality, the PESQ measure has been developed in 2001 [20] and standardized by the International Telecommunication Union (ITU) as ITU-T P.862.2 [107]. The structure is shown in Figure 2.15. The auditory transform is a psychoacoustic model which maps the signal into a representation of perceived loudness in time and frequency (Bark spectrum). Disturbance processing incorporates non-linear averaging, masking, and asymmetry. Following the understanding that localized errors dominate perception, PESQ integrates disturbance over several time-frequency scales using a method designed to take account of the distribution of error in time and amplitude. The estimated speech quality is mapped to a scale of 1 (very bad) to 5 (excellent) termed PESQ-MOS. PESQ is used for narrowband speech (Narrowband PESQ) and wideband speech (PESQ).

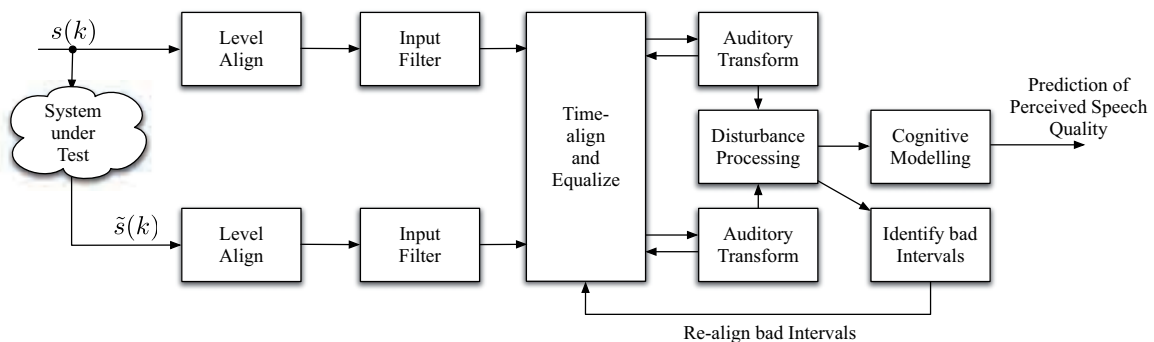


Figure 2.15: Detailed signal flow of PESQ.

2.4.2 Perceptual Evaluation of Audio Quality (PEAQ)

The PEAQ measure has been introduced in [56]. It is based on the combination of multiple objective quality measures (Objective Difference Grade: ODG) to predict the subjectively perceived audio quality (Subjective Difference Grade: SDG) mapped to a scale of -4 (very annoying impairment) to 0 (imperceptible impairment).

PEAQ is an ITU recommendation featuring two variations. The basic version is intended to be fast enough for real-time monitoring, while the advanced version is computationally more demanding but gives more reliable results. The structure of both the basic version and the advanced version is shown in Figure 2.16. Original and degraded signal are transformed into a time-frequency representation by the psychoacoustic model. Then a task-specific model of auditory cognition reduces these data to a number of scalar variables, some of which are mapped to the desired quality measurement. P. Kabal pointed out that the PEAQ standard leaves room for different interpretations in [76]. For the evaluation of audio signals his recommendations concerning the standard have been respected in this thesis.

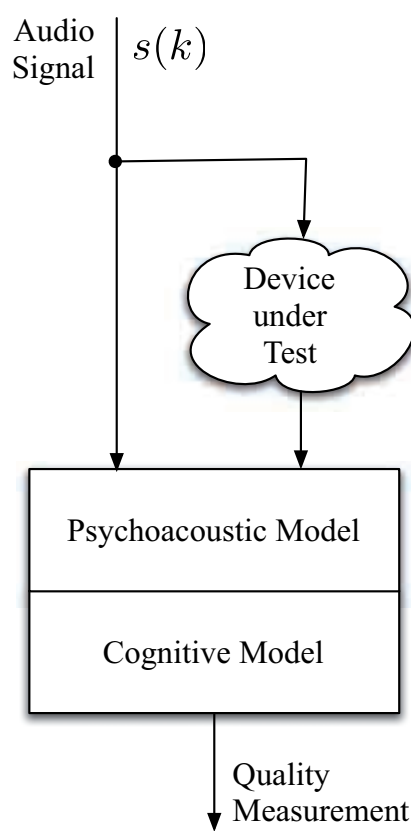


Figure 2.16: General signal flow of PEAQ.

3

Mixed Pseudo Analogue-Digital Speech & Audio Transmission

Mixed Pseudo Analogue-Digital Speech and Audio Transmission is a hybrid scheme based on the principles shown in Figure 3.1. For each frame, it requires

- a digital channel for transmitting the spectral envelope and short term energy:
 - prediction coefficients a_i
 - gain factors g^1
- and a pseudo analogue channel for transmitting normalized discrete-time samples

$$r_n(k) = r(k) \cdot g \tag{3.1}$$

of the prediction residual $r(k)$.

MAD transmission is very efficient with respect to the required transmission bandwidth and with respect to computational complexity. It allows to exploit the mechanisms of linear predictive (LP) coding and noise shaping to produce high quality speech [7] and audio [8].

In the following a definition of *Pseudo Analogue* will be given, previous work will briefly be reviewed, and the novel MAD transmission concept will be detailed.

¹Please note, that (for a good reason which will be explained in the following) the definition of these gain factors differs from that of other speech coding standards.

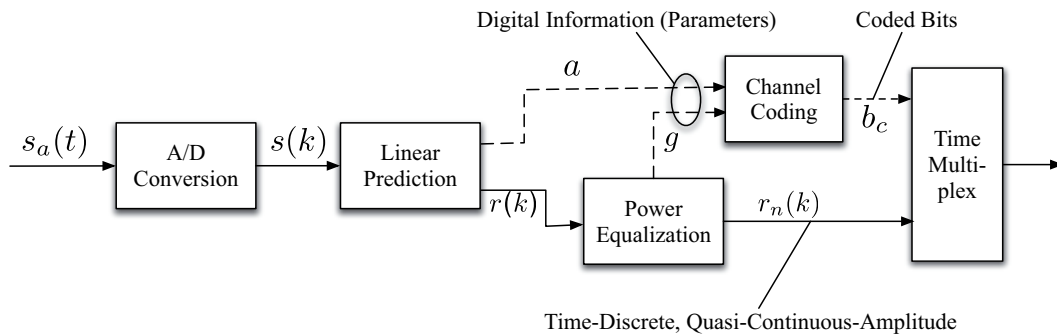


Figure 3.1: Principle of Mixed Pseudo Analogue-Digital Transmission.

3.1 Pseudo Analogue

Strictly speaking, analogue signals are continuous in time and amplitude. The systems described in the following use a sampled representation of these analogue signals following Nyquist's sampling theorem [81]. While this representation is discrete in time, the amplitude is not quantized beyond the resolution of the simulation environment, thus being *quasi-continuous*. In the following, the time discrete, quasi-continuous amplitude signals will be referred to as *pseudo analogue*².

3.2 State-Of-The-Art of Analogue-Digital Transmission

In literature two approaches exist which combine the advantages of digital and analogue transmission.

An early scheme which is closely related to the MAD concept as proposed in this thesis has been studied for ADPCM (Adaptive Differential Pulse Code Modulation) by T. Miki et. al. [44]. It will be introduced in Section 3.2.1.

In [54] a digital channel is used for transmitting the Vector Quantizer (VQ) codebook index of the quantized version of a vector of input samples and an analogue channel (time-discrete, continuous amplitude) to transmit the quantization error. Thus, the receiver gets a quantized (digital) representation of the signal and additionally a refinement signal with continuous amplitude. This approach is examined in Section 3.2.2.

3.2.1 Pseudo-Analog Speech Transmission in Mobile Radio Communication Systems

In 1990, T. Miki, C.-E. W. Sundberg, and N. Seshadri described a first low complexity pseudo analogue transmission system at the IEEE International Symposium

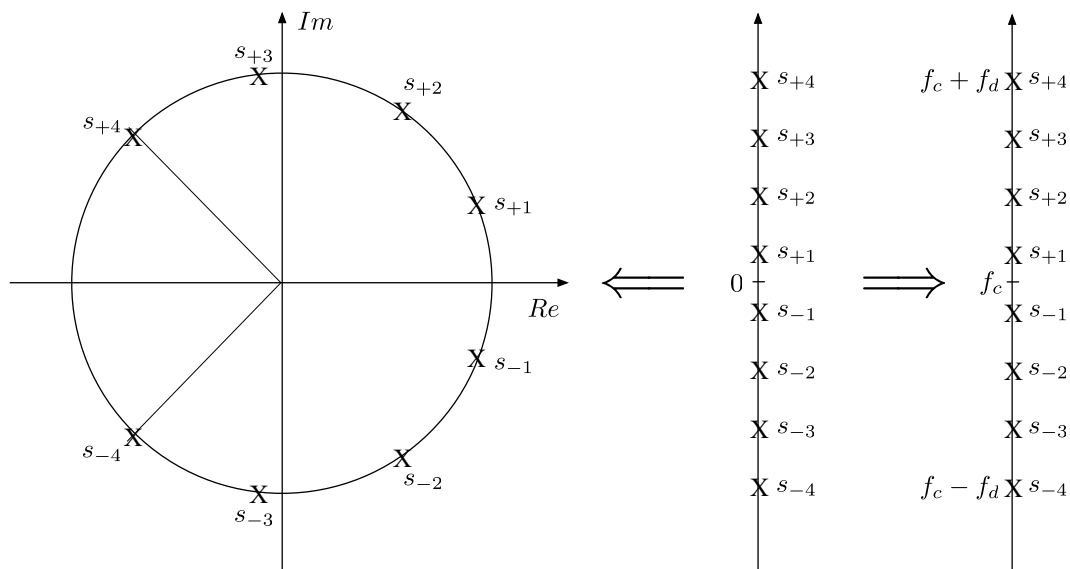


Figure 3.2: Signal mapping for skewed multilevel DPSK and FSK.

on Information Theory [17]. In 1993 this description was published in the IEEE Transactions on Vehicular Technology [44].

Their idea was to combine a speech coder based on ADPCM with a multilevel digital modulation scheme such as M -ary Differential Phase Shift Keying (M-DPSK) or M -ary Frequency Shift Keying (M-FSK) to transmit the residual which is quantized with large M . Mainly, a special case of M-DPSK, called skewed DPSK, is used. Skewed DPSK uses a mapping onto a segment of the unit circle, which is depicted in Figure 3.2 and analyzed in more detail in [15].

The MAD transmission system is a much more general approach which uses higher order prediction with noise shaping to produce the pseudo analogue information and includes digital transmission of side information which is protected by channel coding. The early proposal by Miki can be interpreted as a very special case of the MAD approach.

3.2.2 Hybrid Digital-Analog Coding

Another approach to combine digital and analogue coding, known as *Hybrid Digital-Analog Coding (HDA)*, was proposed by N. Phamdo and U. Mittal in [19, 48], and generalized by M. Skoglund, N. Phamdo, F. Alajaji, Y. Wang, and T. Linder in [22, 25, 54, 55].

The general idea of HDA coding is shown in Figure 3.3. The quantization error of a digital encoder (CELP in [19, 48], a general vector quantizer in [22, 54, 55]) is transmitted over the pseudo analogue channel with either linear, or nonlinear analogue coding, the latter being related to the work of A. Fuldseth and T. Ramstad

²While generally using the British form, “analogue”, the American form, “analog”, is used when quoted from papers written in American English.

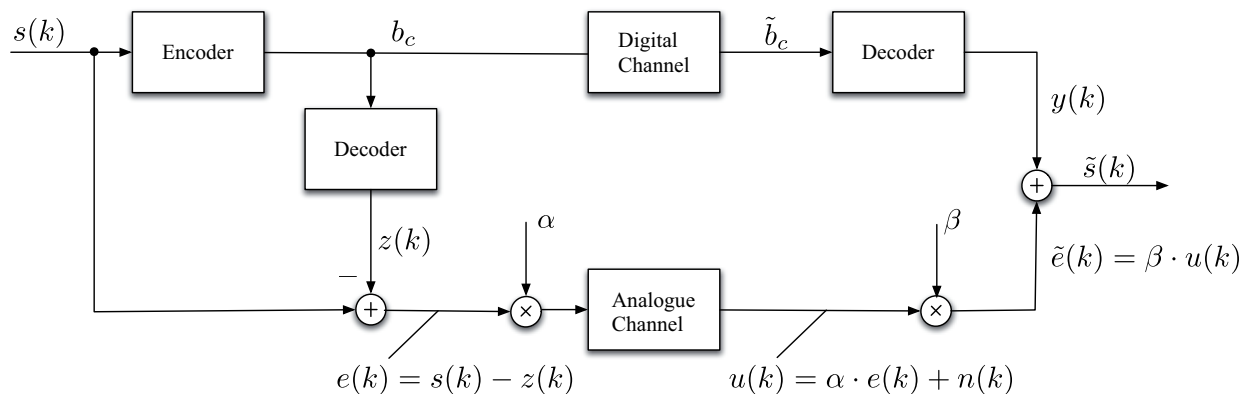


Figure 3.3: Hybrid Digital-Analog coding according to [19, 48] with analogue transmission of the digital coding error.

[5], which we also refer to when analyzing the Archimedes Spiral Mapping in Section 4.2.4.

With the analogue part of the HDA system being formed by the quantization error of a general digital quantizer (the CELP speech codec can also be regarded as such, see [67]), the output of the digital channel can be used as signal representation in low channel SNR conditions, while the analogue refinement helps improving the output for better channels. This can be regarded as 2-stage coding, which as such is inferior to comparable coding in one optimal stage [67]. MAD, in contrast, does not code in two stages, but rather it chooses systematically which part of the information has to be sent via the analogue and which part via the digital channel. In the MAD approach the digital information and the analogue one are needed both. However, the combined information utilizes the increased capacity of the AWGN³ Channel with Gaussian input (see Chapter 5).

3.3 Principles of MAD Coding

The general concept of MAD transmission is shown in Figure 3.1. It is based on Linear Prediction (see Section 2.1.2) with digital transmission of the LP parameters. The unquantized residual of the linear predictor is normalized to unit average power and transmitted as pseudo analogue samples (thus omitting the quantizer Q of Figure 2.4). Normalized samples $r_n(k)$ and coded bits b_c (digital parameters after channel coding) are transmitted sequentially over the baseband channel.

A complete baseband model of the Mixed Pseudo Analogue-Digital transmission system is given in Figures 3.4 (transmitter) and 3.5 (receiver) and will be detailed in the following. Narrowband speech sampled with $f_s = 8$ kHz, wideband speech sampled with $f_s = 16$ kHz, and audio signals sampled with $f_s = 16$ kHz, $f_s = 32$ kHz, and $f_s = 48$ kHz are considered. Application to audio signals with different sampling rates is viable, since the design of MAD transmission is not based on a

³Additive White Gaussian Noise

model of speech production. The objective of MAD transmission is to maximize the subjective quality while minimizing the required transmission bandwidth and coding complexity.

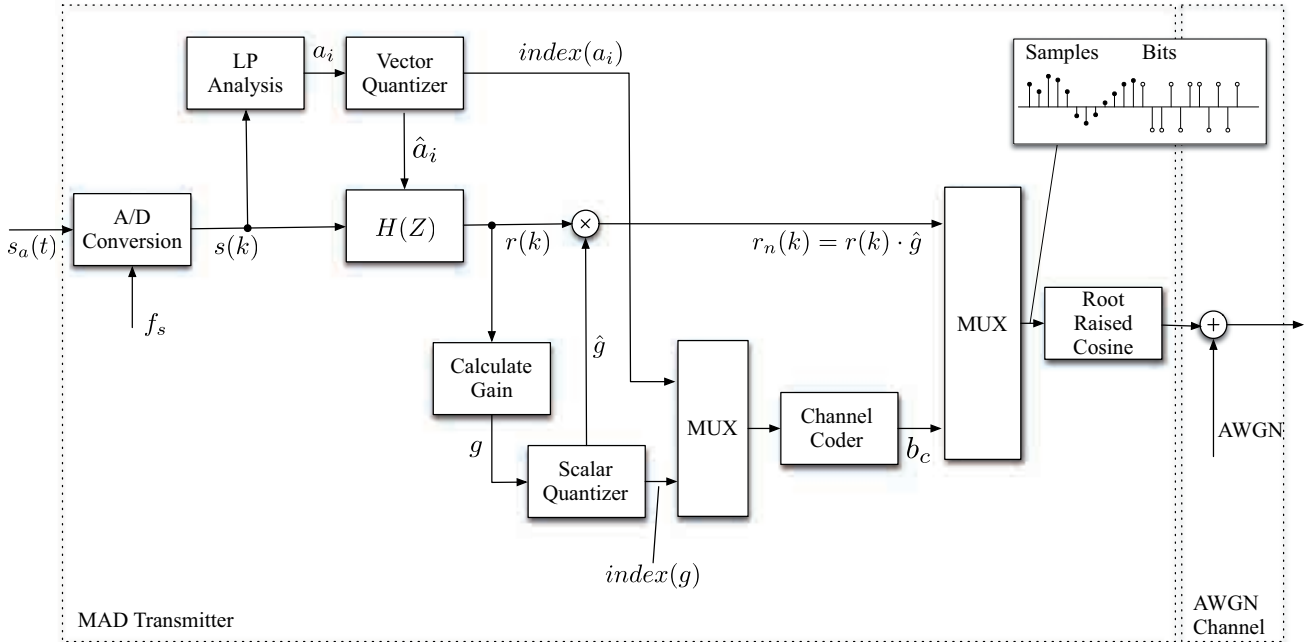


Figure 3.4: Mixed Pseudo Analogue-Digital Speech Transmission: Transmitter and AWGN channel.

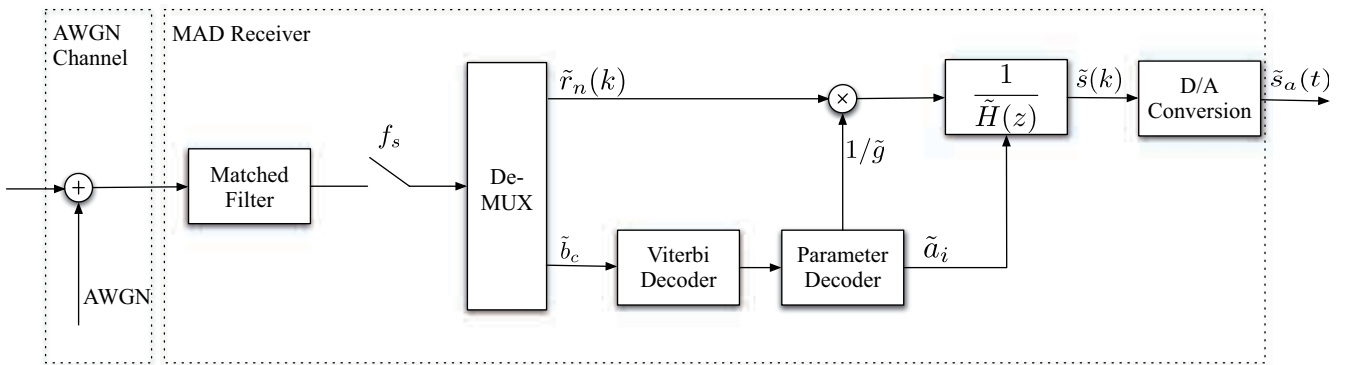


Figure 3.5: MAD Transmission: Receiver.

3.3.1 Processing in the Digital Domain

Linear Prediction

In the MAD system, the general linear prediction analysis filter

$$1 - A(z) = 1 - \sum_{i=1}^N a_i \cdot z^{-i} \tag{3.2}$$

is modified to vary the prediction. The MAD transmission system controls the modified prediction filter

$$H(z) = \frac{1 - A(z)}{1 - A(z/\gamma)} \quad (3.3)$$

by a factor of $\gamma = 0$ (full prediction) to $\gamma = 1$ (no prediction). Varying the prediction parameter γ implies varying the colouring of the audible noise at the receiver output, as the channel noise is filtered with the modified LP synthesis filter

$$G(z) = \frac{1 - A(z/\gamma)}{1 - A(z)}, \quad (3.4)$$

see Figure 3.6.

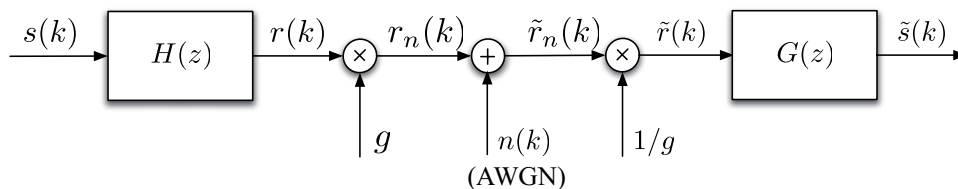


Figure 3.6: Equivalent model for the pseudo analogue part of the MAD transmission system.

This is also called noise shaping in the literature [51, 91].

While the audible noise is coloured with the spectral shape of the speech signal with full prediction

$$\gamma = 0; \quad \frac{1}{H(z)} = \frac{1}{1 - A(z)}, \quad (3.5)$$

the audible noise remains white without prediction

$$\gamma = 1; \quad \frac{1}{H(z)} = 1. \quad (3.6)$$

Figure 3.7 shows the measured perceptual speech quality (Perceptual Evaluation of Speech Quality [107]) over γ for two exemplary channel SNR. The simulation results indicate that for narrowband speech $\gamma \approx 0.5$ yields the best quality regardless of the channel SNR. For wideband speech the best perceptual quality is found for $\gamma \approx 0.6$ to 0.7 , see Figure 3.8. A similar effect has been studied by H. Krüger in [13], who modifies the calculation of the LP coefficients in the frequency domain to achieve noise shaping. H. Krüger's modifications will be used in Chapter 5 to facilitate the theoretical analysis of the MAD principle.

In the MAD speech transmission system, the LP filter coefficients a_i are quantized with a Vector Quantizer, as shown below, and the gain factor g described in Section 3.3.1 is quantized with a scalar quantizer (Q). The quantizer codebook

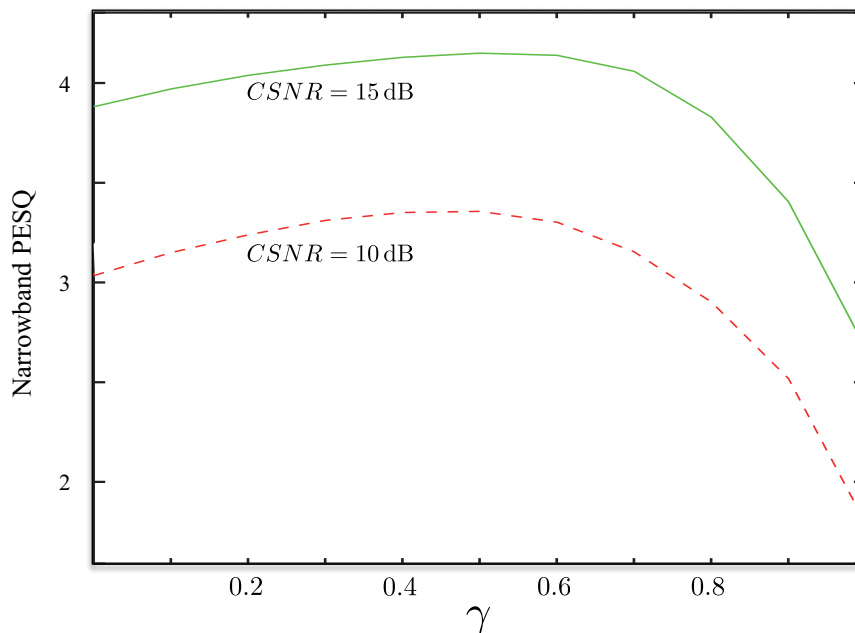


Figure 3.7: Influence of prediction strength γ on the subjective speech quality measure PESQ, narrowband case. For narrowband speech an LP filter of order 10 is applied every 5 ms.

indices of these quantizers form the digital part of the transmission. For this part channel coding as described in section 3.3.1 is applied.

For narrowband input speech (300 Hz - 3.4 kHz audio bandwidth, 8 kHz sampling rate) routines from the narrowband Adaptive Multirate (AMR-NB) speech codec mode 12.2 kbit/s are used. Two sets of LP filter coefficients of order 10 are calculated per 20 ms speechframe and then jointly quantized using SMQ⁴ with the original AMR-NB quantization codebooks [95]. The codebook index from AMR-NB requires 38 bits per 20 ms speechframe.

Wideband input speech (70 Hz - 7 kHz audio bandwidth, 16 kHz sampling frequency) requires an LP filter of order 16. For every 20 ms speechframe one set of LP filter coefficients is calculated and quantized using a combination of Split Vector Quantization (SVQ) and Multi-Stage Vector Quantization (MSVQ) with the original wideband Adaptive Multirate (AMR-WB) quantization codebooks [96]. The codebook index from AMR-WB requires 46 bits per 20 ms speechframe.

For audio signals sampled at any sampling rate, MAD uses the same LP filter of order 16 and the same vector quantizer as does MAD for wideband speech (MAD-WB). While comparison of quantized and unquantized LP coefficients clearly shows that the codebooks stemming from the AMR wideband speech codec are not trained for audio signals, the strength of the MAD principle is already evident in this suboptimal configuration. Incorporating carefully designed quantizer codebooks trained for MAD with audio and speech signals, will enhance the perceived quality even further.

⁴Split Matrix Quantization

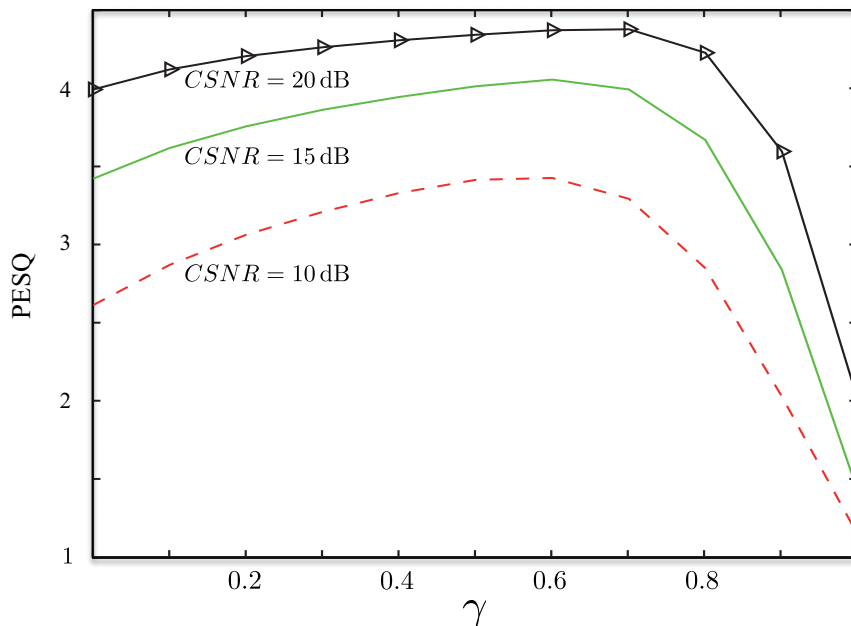


Figure 3.8: Influence of prediction strength γ on the subjective speech quality measure PESQ, wideband case. For narrowband speech an LP filter of order 16 is applied every 5 ms.

Use of a Long Term Prediction (LTP) filter is not feasible in MAD transmission, as MAD is applied to audio as well as to speech signals.

Power Equalization

If different transmission systems are to be compared, besides the bandwidth, the mean output power of the transmitters should be the same. To reduce the dynamics of the residual signal $r(k)$ and to ensure equal average energy per transmitted symbol, for each 5 ms subframe (N_s samples) of the residual signal $r(k)$, a gain

$$g = \sqrt{N_s / \sum r(k)^2} \quad (3.7)$$

is calculated. Multiplying $r(k)$ by the quantized gain factor \hat{g} in each subframe results in normalized, continuous-amplitude samples

$$r_n(k) = r(k) \cdot \hat{g} \quad (3.8)$$

with an average power of 1, which is equivalent to the digital transmission of the symbols 1 and -1 , respectively. The gains g are quantized with a scalar 5-bit Lloyd-Max quantizer and transmitted together with the coefficients a_i (compare Figure 3.4). The quantized gains \hat{g} and coefficients \hat{a}_i form the digital information of the MAD transmission scheme.

Channel Coding for LPC and Gains

To protect the quantized LP coefficients \hat{a}_i and gains \hat{g} , from transmission errors, a rate 1/2 convolutional channel code [87] is applied. The polynomials are

$$G_0 = 1 + D^3 + D^4 \quad (3.9)$$

and

$$G_1 = 1 + D + D^3 + D^4, \quad (3.10)$$

i.e., the same channel code as used for the GSM system with full-rate speech coding [98], which acts as a reference in Section 3.3.3. The output \mathbf{b}_c of the channel coder has the rate R_{digi} . At the receiver side of both systems a hard-decision Viterbi decoder [87] is used in all cases. This scheme was chosen for reasons of low complexity and comparability.

3.3.2 Baseband Transmission Model

In a first approach, BPSK modulation is used with a symbol rate

$$R_{digi} = \frac{1}{T}, \quad (3.11)$$

T being the time per bit, and a Root Raised Cosine (RRC) pulse shaping filter with a roll-off factor

$$\alpha = 0.5. \quad (3.12)$$

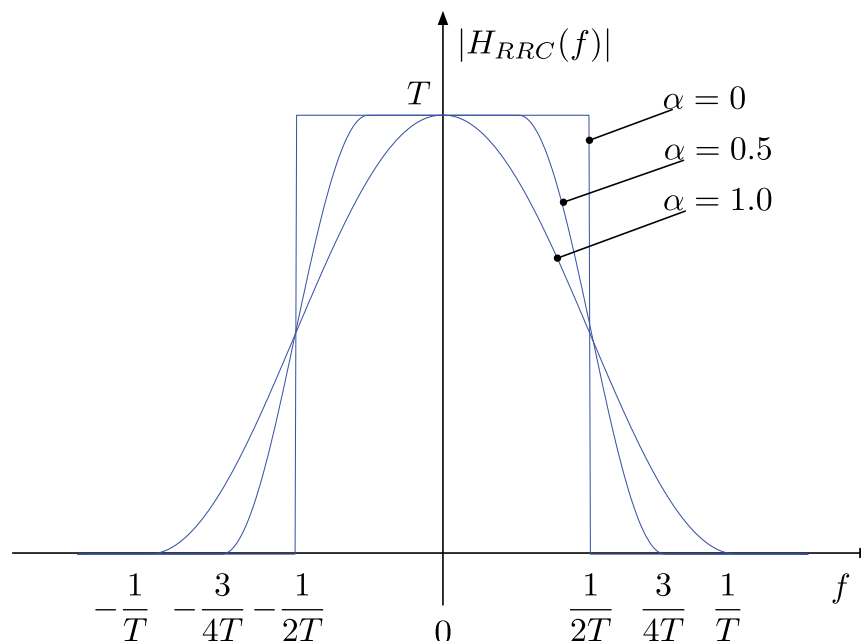


Figure 3.9: Frequency response $|H_{RRC}(f)|$ of the Root Raised Cosine filter with different roll-off factor α .

The required bandwidth for band pass transmission [81] is

$$B'_{digi} = R_{digi}(1 + \alpha). \quad (3.13)$$

In the signal space, BPSK is considered for transmission of the digital information. The unquantized, quasi-continuous residual r_n is transmitted with Pulse Amplitude Modulation. The transmission of analogue and digital parts is investigated in the baseband. To prevent inter-symbol interference, the multiplexed analogue and digital pulses are shaped with the same RRC filter (roll-off factor $\alpha = 0.5$).

The residual signal is not quantized (which means substantial savings in terms of complexity). Instead, the normalized, time-discrete, continuous-amplitude samples

$$r_n = \hat{g} \cdot r \quad (3.14)$$

are directly fed to the Root Raised Cosine filter in addition (time multiplex) to the digital data and transmitted over the AWGN or Rayleigh channel. Thus, at the receiver, the decoded signal is affected by channel noise, but not by quantization noise. For the combined signal, the two-sided low pass bandwidth [68], which is relevant for band pass transmission, equals

$$\begin{aligned} B' &= B'_{ana} + B'_{digi_{BPSK}} \\ &= (1 + \alpha) \cdot (R_{ana} + R_{digi}) \\ &= 1.5(R_{ana} + R_{digi}) \end{aligned} \quad (3.15)$$

with the analogue sample rate R_{ana} and the digital bit rate R_{digi} .

Throughout this contribution, the SNR on the channel will be measured as E_s/N_0 , with E_s being the energy per coded bit or average energy per sample, respectively.

3.3.3 Transmission over Band Pass AWGN Channels

Narrowband MAD and AMR

To evaluate the MAD principle, it was compared to the GSM Adaptive Multirate (AMR-NB) codec [95] mode 12.2 kbit/s (GSM Enhanced Full-rate) operating at 22.8 kbit/s including channel coding [98]. In the simulation, the AMR-NB bitstream was fed to the same pulse shaping filter as described above. Transmission over an AWGN channel with spectral density $\frac{N_0}{2}$ is assumed. In addition to the identical convolutional code, no further error concealment was used in both cases. The required bandwidth of the AMR-NB speech codec is

$$B'_{AMR_{NB}} = 1.5 \text{ /bit} \cdot 22.8 \text{ kbit/s} = 34.2 \text{ kHz.} \quad (3.16)$$

Narrowband MAD transmission needs 38 bits/20 ms to quantize the LP coefficients of order 10 with modules from the AMR-NB narrowband speech codec, 20 bits/20 ms for the gains (4 subframes times 5 bits), and 4 bits/20 ms for termination of the convolutional code, adding up to

$$\begin{aligned} R_{digi_{NB}} &= (38 + 20 + 4) \frac{\text{bits}}{\text{frame}} \cdot 50 \frac{\text{frames}}{\text{s}} \cdot 2 \\ &= 6.2 \text{ kbit/s} \end{aligned} \quad (3.17)$$

after channel coding. With the sampling rate

$$f_{s_{NB}} = 8000 \text{ Hz} \quad (3.18)$$

the bandwidth needed for the residual signal equals

$$B'_{ana_{NB}} = 1.5 \cdot f_{s_{NB}} = 12 \text{ kHz} \quad (3.19)$$

and the bandwidth needed for the digital part equals

$$B'_{digi_{NB}} = 1.5 \cdot R_{digi_{NB}} = 9.3 \text{ kHz.} \quad (3.20)$$

Due to the time multiplexed transmission of normalized samples $r_n(k)$ and coded bits \mathbf{b}_c , the total bandwidth is

$$\begin{aligned} B'_{MAD_{NB}} &= B'_{ana_{NB}} + B'_{digi_{NB}} \\ &= 1.5 \cdot (f_{s_{NB}} + R_{digi_{NB}}) \\ &= 21.3 \text{ kHz.} \end{aligned} \quad (3.21)$$

The black solid line in Figure 3.10 shows the measured wideband PESQ values for different E_s/N_0 with E_s the energy per coded bit or average energy per sample, respectively. Wideband PESQ has been chosen to be able to compare the AMR-NB to both narrowband and wideband MAD transmission in section 3.3.3. The original wideband speech signal (clean speech, 1.3 million samples, mixed male and

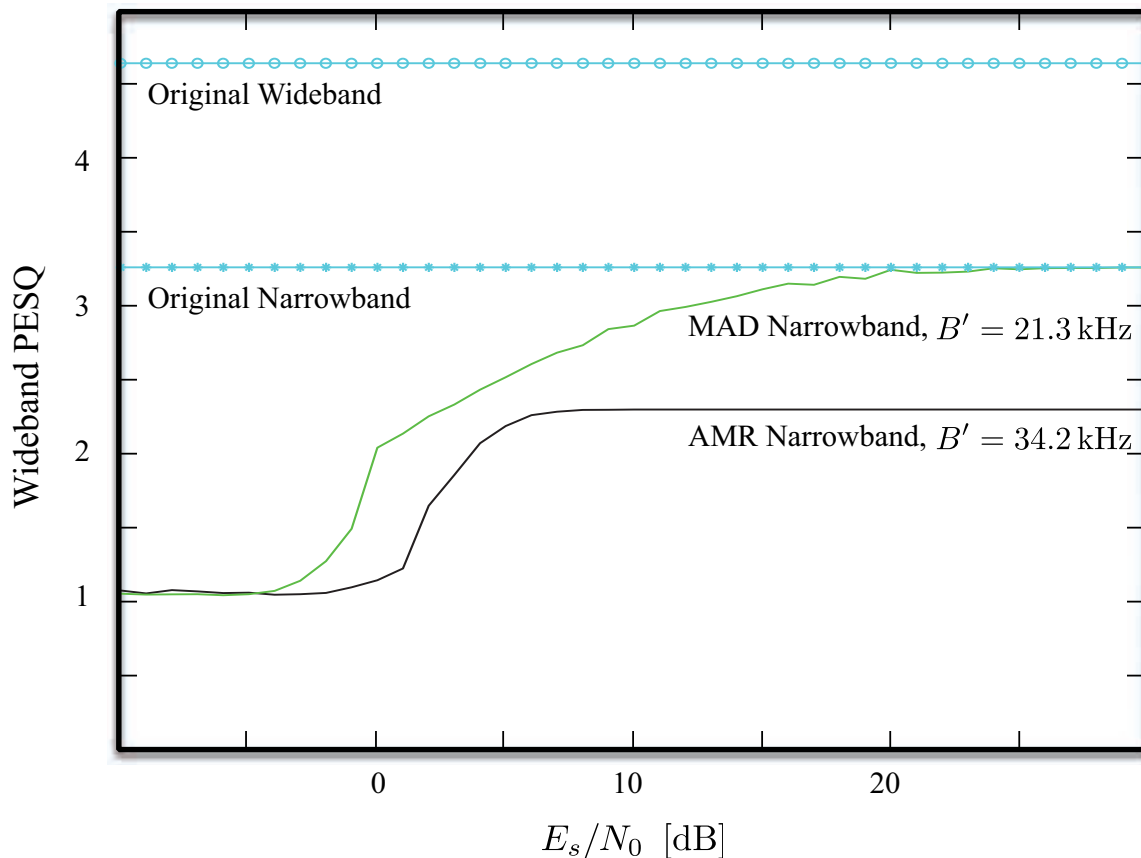


Figure 3.10: Comparison of AMR-NB and narrowband MAD coding. E_s : energy per coded bit or average energy per sample, respectively.

female speakers, German, English, Spanish, and Russian language) scores 4.65 on that scale, the narrowband reference signal is a 4 kHz low-pass filtered version of the wideband signal, scoring 3.26. With error free transmission, the digital narrowband AMR-NB speech codec scores 2.29 on the average.

Figure 3.10 shows the measured wideband PESQ values of narrowband MAD transmission for different E_s/N_0 . Besides a reduction in bandwidth by about 38%, the MAD transmission scheme also has a significantly lower computational complexity than a CELP scheme as used, e.g., in the AMR-NB speech codec (see Section 5.4.3). This stems from the fact that there are no open loop pitch, adaptive, and stochastic codebook searches in MAD. Using MAD transmission, the speech quality rises with improving channel conditions until truly transparent speech transmission is reached. With falling E_s/N_0 , MAD degrades gracefully up to the point when the digital information is corrupted and wrong LP coefficients are decoded. This threshold effect, however, starts at lower E_s/N_0 than with the digital system.

Wideband Coding

If wideband speech (7 kHz audio bandwidth, 16 kHz sampling frequency) is available at the transmitter, the MAD transmission scheme allows for wideband coding with increased sampling rate and LP filter order. The transmission bandwidth also

remains well in the same region as the bandwidth required for narrowband AMR-NB transmission.

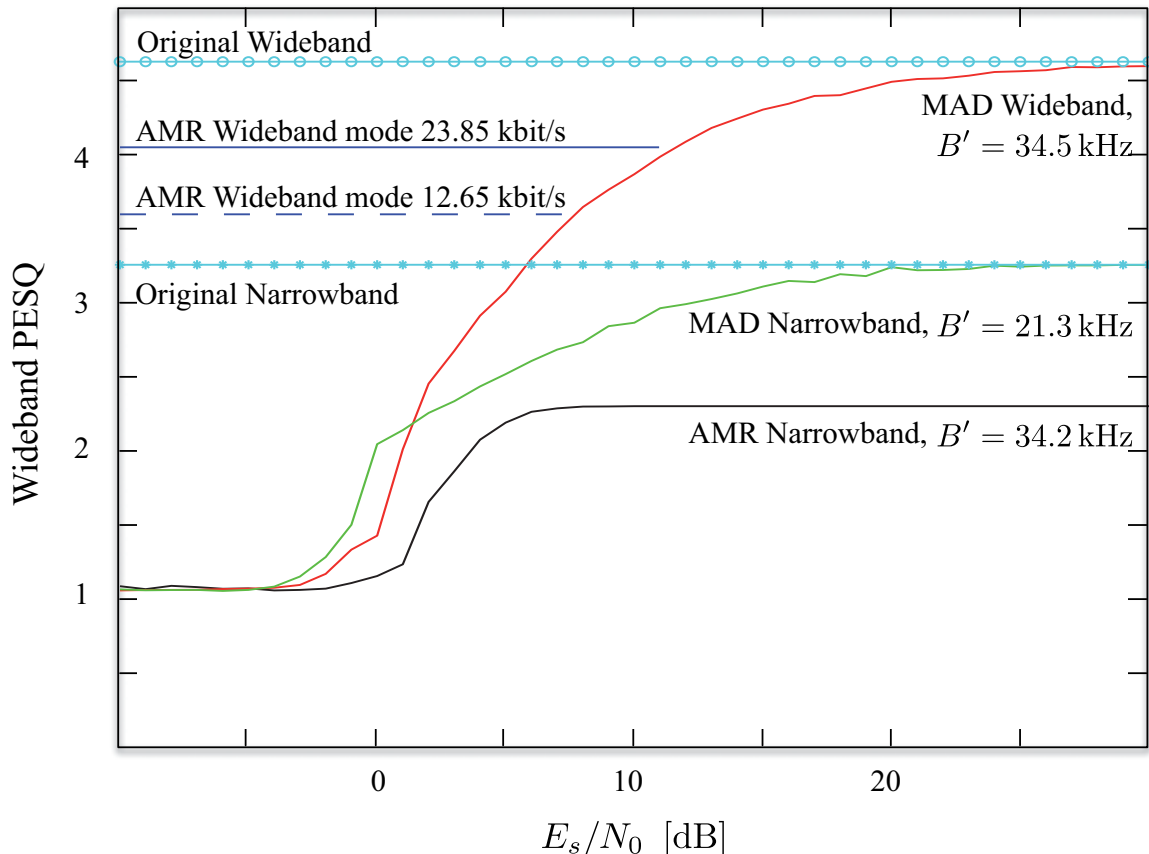


Figure 3.11: Comparison of AMR-NB and MAD coding. E_s : energy per coded bit or average energy per sample, respectively.

Wideband MAD transmission differs from narrowband MAD transmission only with respect to the input sample rate of the speech signal and linear prediction order. Quantization of the LP coefficients of order 16 is carried out with modules from the AMR-WB wideband speech codec [96]. Wideband transmission requires 46 bits/20 ms for the LP coefficients. Thus

$$\begin{aligned} R_{\text{digi}_{WB}} &= (46 + 20 + 4) \frac{\text{bits}}{\text{frame}} \cdot 50 \frac{\text{frames}}{\text{s}} \cdot 2 \\ &= 7 \text{ kbit/s} \end{aligned}$$

after channel coding. With this and a sampling frequency

$$f_{s_{WB}} = 16000 \text{ Hz}, \quad (3.22)$$

the bandwidth used for the analogue residual becomes

$$B'_{\text{ana}_{WB}} = 1.5 \cdot f_{s_{WB}} = 24 \text{ kHz} \quad (3.23)$$

and the bandwidth needed for the digital part is

$$B'_{digi_{WB}} = 1.5 \cdot R_{digi_{WB}} = 10.5 \text{ kHz.} \quad (3.24)$$

The total bandwidth can finally be obtained:

$$\begin{aligned} B'_{MAD_{WB}} &= B'_{ana_{WB}} + B'_{digi_{WB}} \\ &= 1.5 \cdot (f_{s_{WB}} + R_{digi_{WB}}) \\ &= 34.5 \text{ kHz.} \end{aligned}$$

Figure 3.11 shows the gain in speech quality using wideband coding. The computational complexity of wideband MAD transmission differs only with respect to the sampling frequency from that of narrowband MAD. It is still substantially below the one of a narrowband CELP codec, compare Section 5.4.3. To give another reference, the maximum speech quality of the AMR-WB speech codec mode 12.65 kbit/s and mode 23.85 kbit/s are shown on the left of Figure 3.11.

MAD Audio Coding

MAD audio transmission is based on the same modules for linear prediction of order 16 and power equalization as MAD-WB. The only difference is the possibility to select a different sampling rate for the input signal (and thus for the transmission of the residual signal). Choosing, e.g., $f_{s_{Audio}} = 32 \text{ kHz}$ results in:

$$\begin{aligned} R_{digi_{Audio32}} &= (46 + 20 + 4) \frac{\text{bits}}{\text{frame}} \cdot 50 \frac{\text{frames}}{\text{s}} \cdot 2 \\ &= 7 \text{ kbit/s} \end{aligned}$$

after channel coding. The bandwidth used for the analogue residual is

$$B'_{ana_{Audio32}} = 1.5 \cdot f'_{s_{Audio}} = 48 \text{ kHz} \quad (3.25)$$

and the bandwidth needed for the digital part remains

$$B'_{digi_{Audio32}} = 1.5 \cdot R_{digi_{Audio32}} = 10.5 \text{ kHz.} \quad (3.26)$$

The total bandwidth is

$$\begin{aligned} B'_{MAD_{Audio32}} &= B'_{ana_{Audio32}} + B'_{digi_{Audio32}} \\ &= 1.5 \cdot (f_{s_{Audio}} + R_{digi_{Audio32}}) \\ &= 58.5 \text{ kHz.} \end{aligned}$$

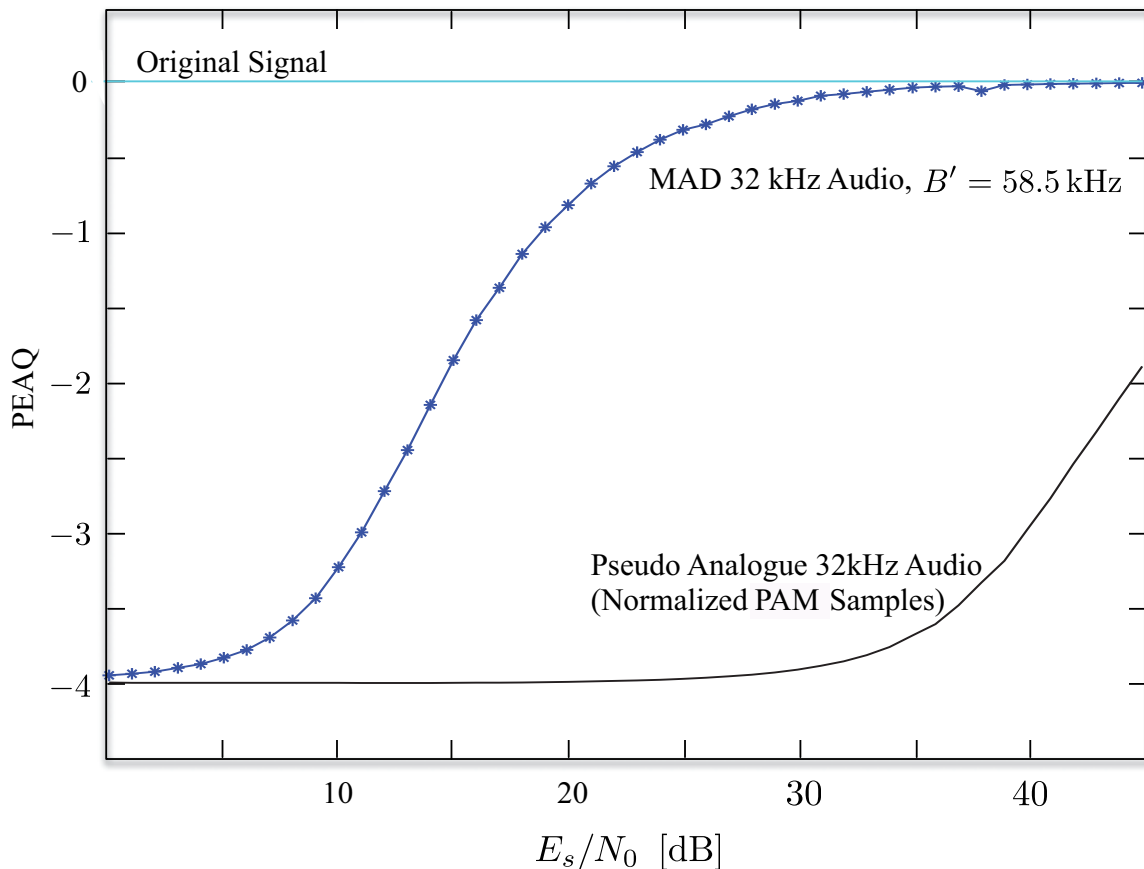


Figure 3.12: MAD Audio Transmission. E_s : energy per coded bit or average energy per sample, respectively.

Figure 3.12 shows the Perceptual Estimation of Audio Quality [104] for audio signals transmitted with the MAD system in comparison to normalized PAM transmission without noise shaping. The benefit of MAD transmission with audio signals of further sampling rates was rated similarly in informal listening tests carried out at our institute.

3.3.4 Transmission over Rayleigh Fading Channels

To demonstrate the behaviour of MAD transmission in Rayleigh fading (flat fading) scenarios, the AWGN channel has been extended to incorporate a Rayleigh fading factor α_k as shown in Figure 3.13. It is assumed that fading bursts are resolved by appropriate interleaving for the digital information (α_k uncorrelated) and that the attenuation factors α_k are known to the receiver from adequate estimation. This assumption is reasonable, as for the pseudo analogue samples no interleaving is necessary and thus with Equation (3.8) at the receiver follows

$$\frac{1}{N_s} \sum r_n^2(k) \alpha_k^2 = \alpha_k^2, \quad (3.27)$$

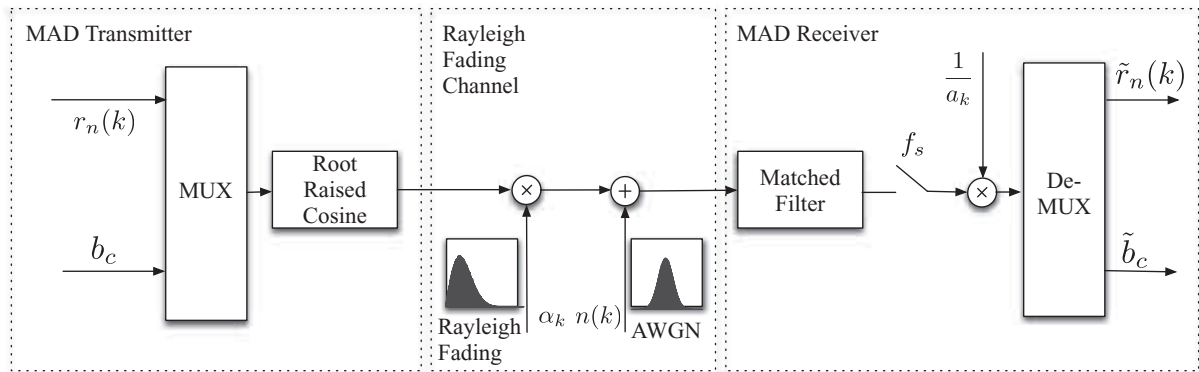


Figure 3.13: MAD Transmission: Model of the Rayleigh fading channel.

with N_s corresponding to the number of samples in a subframe of 5 ms length.

Figure 3.14 shows the performance of narrowband and wideband MAD coding, and that of the AMR-NB transmitting over an AWGN channel with Rayleigh fading. It can be seen that, while the degradation is somewhat higher due to the fading losses, the general behaviour is not changed and MAD transmission still offers competitive performance with most channel conditions.

3.3.5 Scalability of the MAD Transmission Scheme

As indicated in Section 3.3.1, MAD uses the same LP predictor for wideband speech and audio signals. The only difference is that audio signals might be sampled with various sampling rates. While sampling rates of 32 kHz, 44.1 kHz, or 48 kHz are most widespread, MAD transmission can be adapted to virtually any sampling rate of the input signal. For optimum results, however, the VQ which quantizes the LP coefficients should be trained to the respective set of input signals.

Additionally, MAD allows for seamless adaptation to different channel conditions within the general transmission framework, see, e.g., Section 4.2.4.

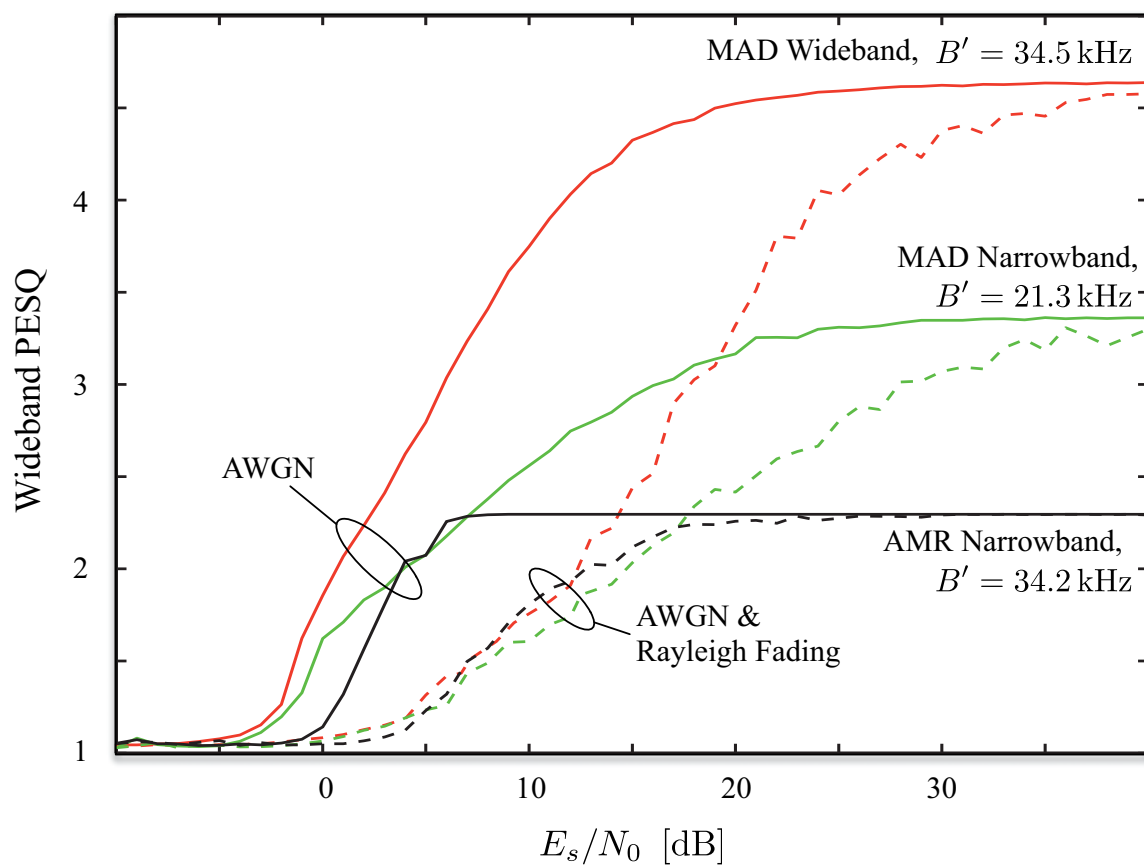


Figure 3.14: Comparison of AMR-NB and MAD transmission over Rayleigh fading channels. E_s : energy per coded bit or average energy per sample, respectively.

4

MAD Modulation Schemes

This chapter presents different modulation schemes which can be used for MAD transmission. Generally, these modulation schemes fall into two categories: schemes with constant magnitude of the complex envelope and those with multi-level modulation. Multi-level schemes in general require more expensive demodulators.

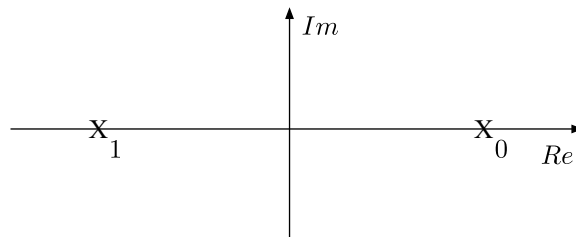


Figure 4.1: Signal constellation of BPSK.

4.1 Digital Information

Digital modulation schemes with constant power include Phase Shift Keying and Frequency Shift Keying among others. For MAD transmission, BPSK and QPSK are used. They have been introduced in Section 2.3.2. The respective signal sets can be found in Figures 4.1 and 4.2.

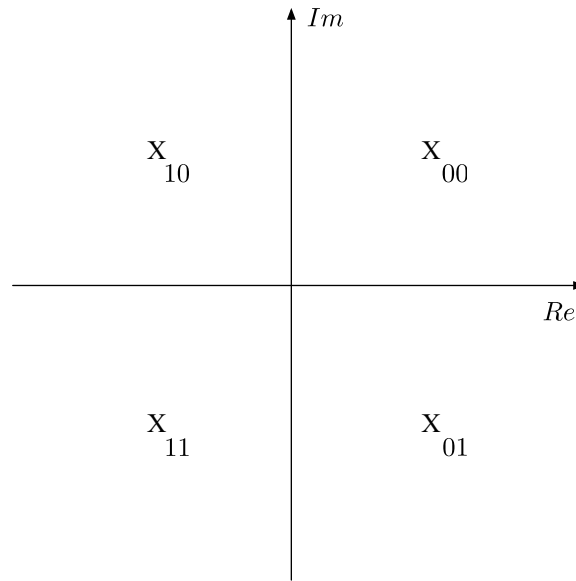


Figure 4.2: Signal constellation of QPSK.

Schemes with variable envelope like Amplitude Shift Keying (ASK, see Figure 2.13), or Quadrature Amplitude Modulation (QAM) are well known and thoroughly covered in literature, e.g., [81].

4.2 Pseudo Analogue Information

To transmit the pseudo analogue information, modulation schemes that allow for discrete-time transmission of continuous values have to be used.

Within the following sections, possible mappings to the complex signal space are introduced. Early work on optimal mappings has been carried out by Shannon [53], Kotel'nikov [79], Wozencraft [90], and Sakrison [83]. The geometric interpretation of the modulation schemes [43] used in the following sections is of special interest to understand the design principles.

Considering a scalar input r constrained to a finite range $\pm A$ and assuming a mapping of the line representing the values of r onto a locus in the 2-dimensional plane with coordinates h_1 and h_2 as illustrated in Figure 4.3, the received signal \tilde{r}' is a point in the plane that differs from the transmitted point r' of the locus by two-dimensional Gaussian noise. The task of the demodulator now is to estimate which value on the locus was sent given a received point in the plane. The simplest rule (which is not necessarily optimum) is to select the point on the locus which is closest to the received point. The question now is how to draw the locus. If the noise is small enough not to decode one major segment of the locus, if another was transmitted, and if the representation of the noise components is chosen as one tangent to the locus and one perpendicular to the locus at the estimated value, the effective noise will be the component projected to the locus (the tangent). The

error in the estimate of r , however, only depends on how much the estimated value \tilde{r}_{est} differs from the transmitted signal r . If r was uniformly mapped to a locus of length

$$L = 2A \cdot m, \quad (4.1)$$

then the error in \tilde{r}_{est} normalized to the regular scale of r will be $\frac{1}{m}$ in the size of the noise component.

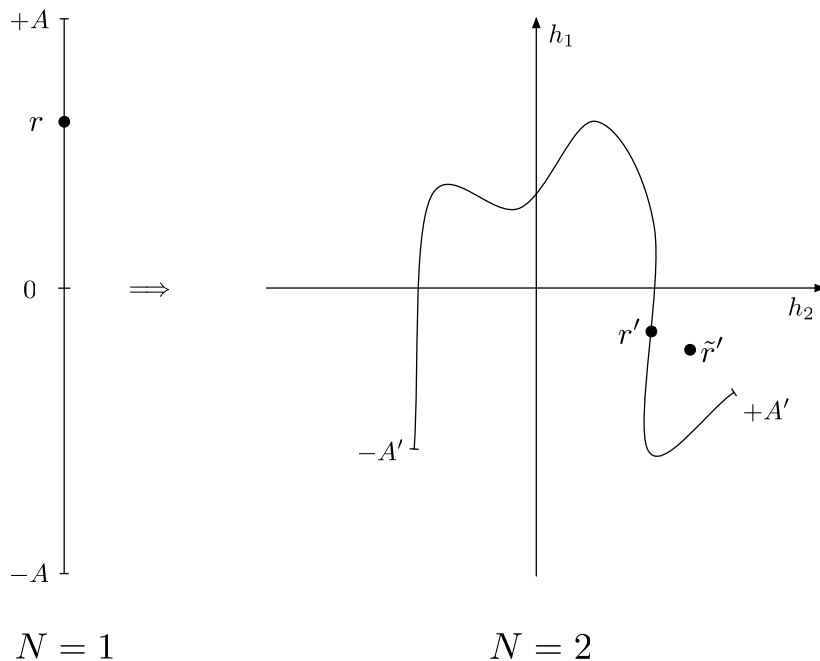


Figure 4.3: Geometrical approach.

Even with constrained average power in the two dimensions, such that

$$P_s = E\{h_1^2(r') + h_2^2(r')\} = E\{r^2\} \quad (4.2)$$

the normalized error in the estimate can be arbitrarily small if the length L of the locus is sufficiently long. The problem here is that with limited power P_s the locus cannot be lengthened without reducing the distance between locations representing largely different values of r . This fact will be looked at in Sections 4.2.4 and 5.4.4.

4.2.1 Mapping to the Unit Circle

First the modulation schemes with constant amplitude shall be regarded. The constraint of having unit energy means that all valid points in the signal space fall onto the unit circle. Thus, the maximum length L of the locus in the signal space is

$$L_{UnitCircle} = 2\pi. \quad (4.3)$$

In the following, different possibilities of mapping pseudo analogue samples to the unit circle will be examined.

Direct Mapping to the Complete Circle

Figure 4.4 shows a direct mapping to the unit circle. This mapping corresponds to pseudo analogue phase modulation.

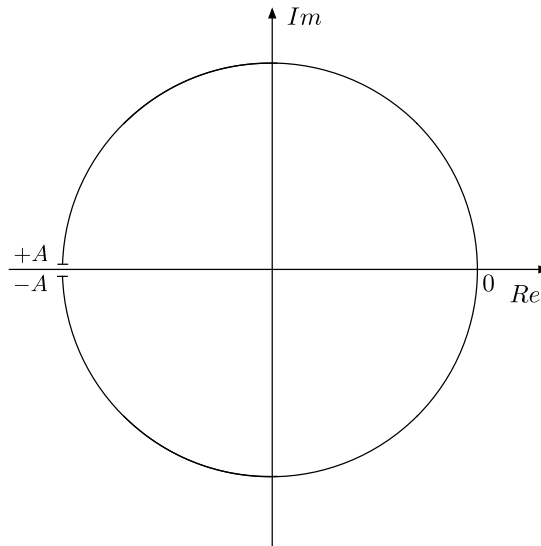


Figure 4.4: Mapping of r_n to the complete unit circle. Note that $+A$ and $-A$ are neighbouring points on the circle.

While it achieves a length of

$$L_{CompleteCircle} = 2\pi, \quad (4.4)$$

this mapping has the drawback of positive and negative maxima being at neighbouring positions with virtually no distance in between, which means that in the case of noise there is a 50% chance of $+A$ being decoded as $-A$.

2-Way Mapping to the Complete Circle

To circumvent the problem of positive and negative maxima being too close to each other, $r_n(k)$ could be quantized to $\hat{r}_n(k)$ with q bits and mapped to the unit circle as depicted in Figure 4.5. Here, $+A$ and $-A$ are separated by the farthest possible distance. The least significant bit determines whether the upper or lower half circle is used for transmission. Therefore, this mapping is called *2-Way Mapping* (2-WM).

Usually, q will be chosen to be $q = 8$, $q = 16$, or even higher for the transmission of pseudo analogue samples. Thus, this mapping can be regarded as M -ary PSK (MPSK) with M equal to

$$M = 2^q \quad (4.5)$$

using a special ordering. In contrast to Gray mapping, the goal of 2-WM is not to reduce the bit error probability but rather the distortion of $\hat{r}_n(k)$ due to the

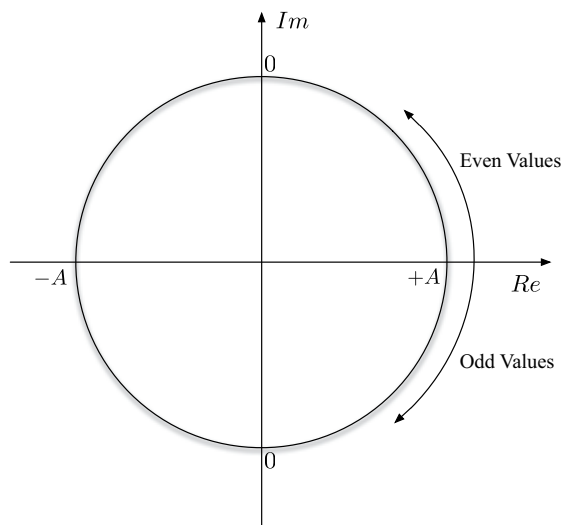


Figure 4.5: 2-way mapping of \hat{r}_n to the unit circle. While $+A$ and $-A$ are separated by the furthest possible distance, effectively the signal space is only used by half.

transmission noise. However, while the complete unit circle seems to be covered, 2-WM just maps the linear axis twice. The mappings are equal and do not add up.

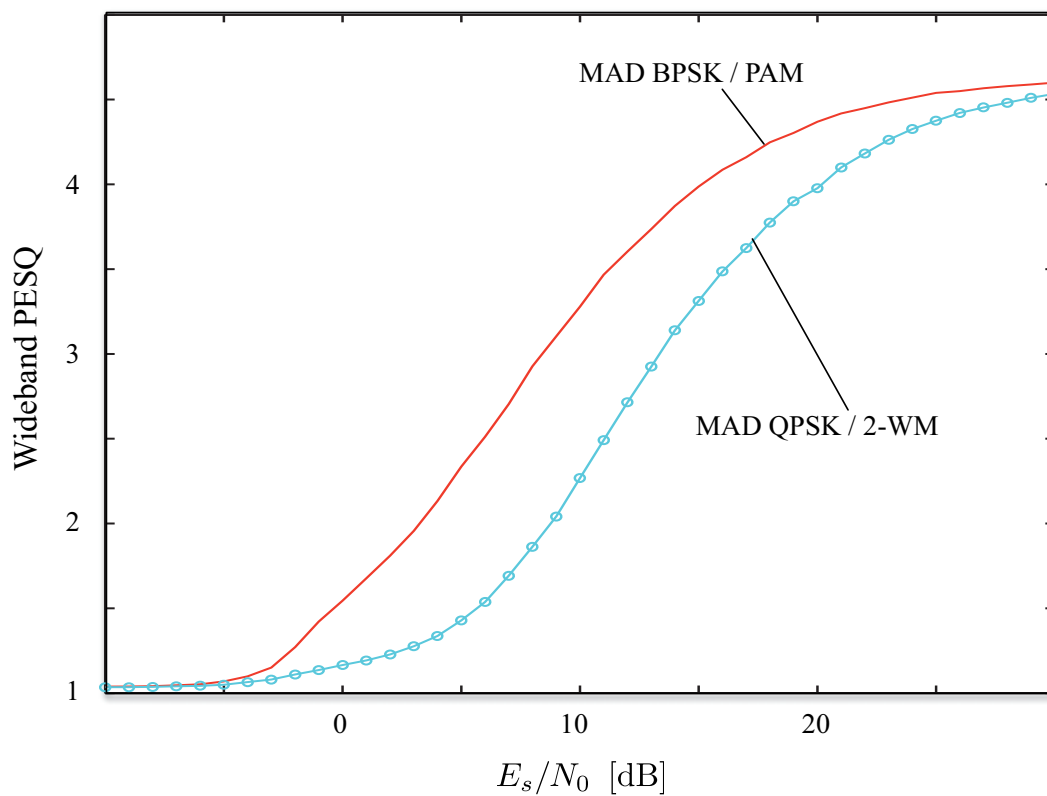


Figure 4.6: MAD Transmission: Speech quality for 2-way mapping of \hat{r}_n .

Thus, 2-WM has an effective length of

$$L_{2-WM} = \pi \quad (4.6)$$

which is lower than the effective length of PAM.

Figure 4.6 shows the speech quality of MAD transmission using 2-WM in comparison to the MAD system introduced in Section 3.3.3. As would be expected, MAD transmission using 2-WM in the pseudo analogue part has a lower output quality than MAD transmission using PAM.

Skewed Phase Shift keying

Based on Miki's pseudo analog transmission system [17, 44], see 3.2.1, Lin, Wang, and Sundberg have analyzed skewed PSK (SPSK) in more detail in [15]. Skewed PSK addresses the problem of $+A$ and $-A$ being too close together in the "Mapping to the Complete Circle" by separating these end points on the locus by an angle of φ as depicted in Figure 4.7.

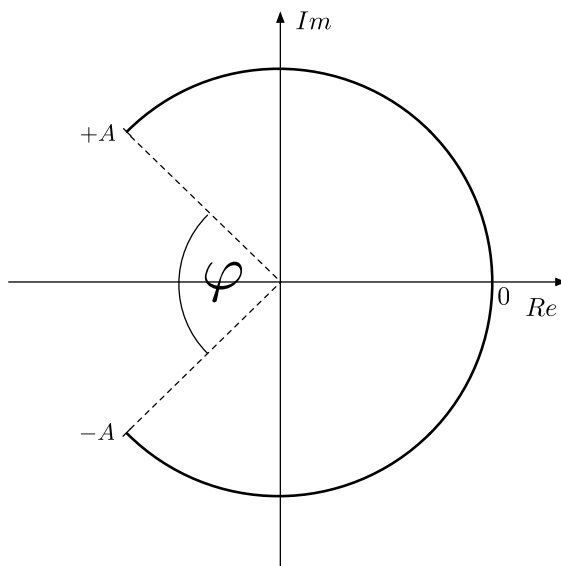


Figure 4.7: Skewed Phase Shift Keying.

This scheme works well for low resolutions, such as, e.g., 4 bits per sample, where optimal skewing angles can be calculated and verified by simulations (see Figure 4.8). For increased resolution of the quantization, e.g., for 8 bits per sample as shown in Figure 4.9, no distinct optimal skewing angle is observed and the performance is dictated primarily by the proximity of neighbouring points. The reason for this is that the normalized residual has a Gaussian distribution (see Section 5.4.3) resulting in the maxima having a significantly lower probability than smaller values. Thus, the frequent errors in small values contribute more to the degradation than rare sign errors at the maxima.

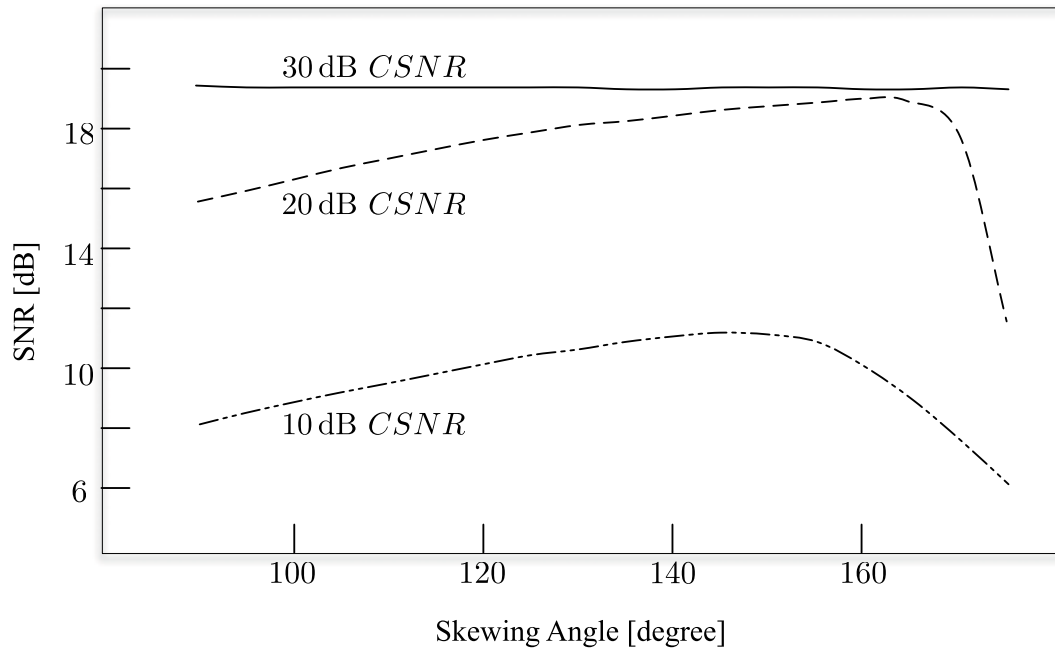


Figure 4.8: Performance of 4-bit PCM/SPSK transmission in Gaussian channels. SNR at the receiver output over Channel-SNR (CSNR). Figure taken from [15]

The length L of the locus in skewed PSK depends on the skewing angle:

$$L_{SkewedPSK} = 2\pi - \varphi. \quad (4.7)$$

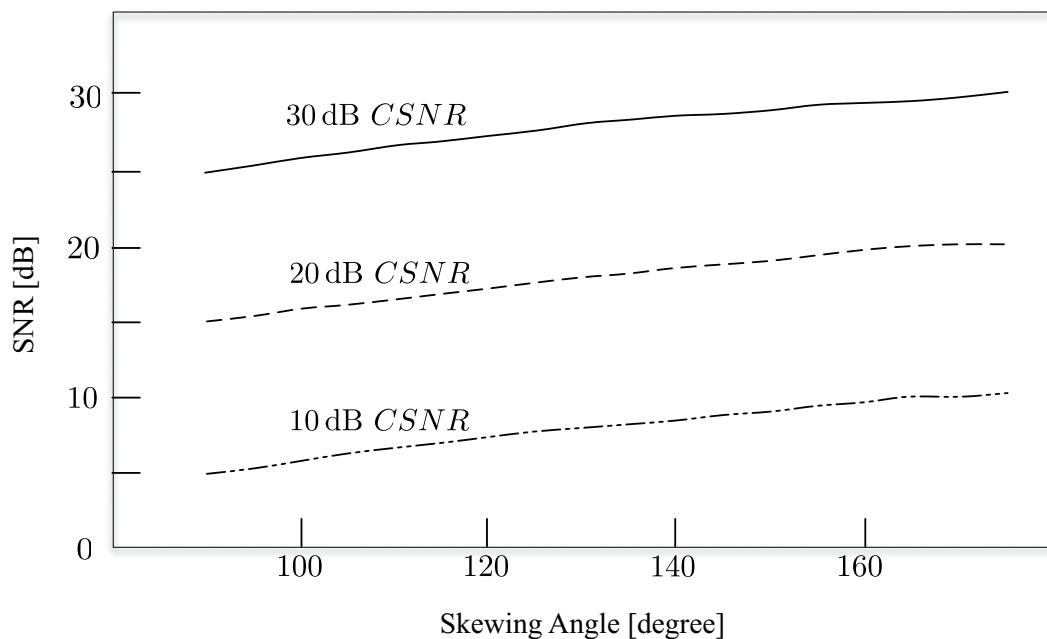


Figure 4.9: Performance of 8-bit PCM/SPSK transmission in Gaussian channels. SNR at the receiver output over Channel-SNR (CSNR). Figure taken from [15]

What all modulation schemes, introduced in Section 4.2.1, have in common is, that for unbounded sources, either a companding function, which maps the real line onto the length of the locus needs to be used, or clipping is necessary. Suitable companding functions are regarded, e.g., in Sakrison's monograph [84].

4.2.2 Pulse Amplitude Modulation (PAM)

As has been detailed above, the length L of the locus determines the output SNR of the modulation schemes in weak noise conditions (just as the Euklidian distance determines the error probability in digital modulation schemes). If the amplitude of the modulation system is not limited, Pulse Amplitude Modulation¹ will allow for mapping the samples directly to the real axis of the signal space following Equation (2.17) (see Figure 4.10).

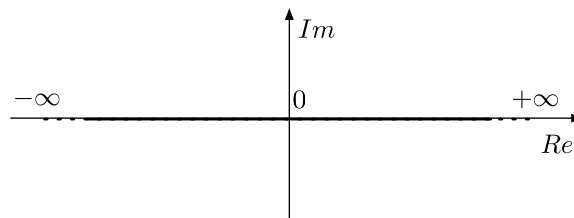


Figure 4.10: Pulse Amplitude Modulation.

If the samples have unit average energy, this modulation will operate at the same average transmit power as the systems introduced in Section 4.2.1. Goblick, Tufts, and Berger have shown in [29, 35, 62] that for a system transmitting samples of a memoryless Gaussian source over an AWGN channel at Nyquist rate

$$\frac{1}{T} = 2B, \quad (4.8)$$

PAM reaches the least mean square error theoretically attainable with any communication system whatsoever.

In Chapter 3.3 PAM was used for transmission of the normalized pseudo analogue residual signal $r_n(k)$.

¹In this text Pulse Amplitude Modulation refers to the modulation of a signal with continuous amplitude, while Amplitude Shift Keying is used to express the digital variant of PAM with M amplitude levels. Or, in other words, PAM resembles M -ASK with $M \rightarrow \infty$.

4.2.3 Quadrature Pulse Amplitude Modulation (QPAM)

Combining two samples of the normalized pseudo analogue residual signal $r_n(k)$ as in Equation (2.18) for transmission in inphase and quadrature component, respectively, (corresponding to BPSK and QPSK in the digital case) yields Quadrature Pulse Amplitude Modulation (QPAM). MAD transmission using QPSK/QPAM performs equivalently to MAD using BPSK/PAM as described in Chapter 3.3 with the only difference being that the transmission bandwidth is reduced by a factor of two.

4.2.4 Archimedes Spiral Mapping (ASM)

Mappings from a message space of dimension M to a signal space of dimension N are referred to as *Shannon-Kotel'nikov* mappings in literature [4] to honor Claude Shannon who presented a geometrical view in his 1949 paper [53] and V.A. Kotel'nikov, who had already developed a theory for bandwidth expanding modulation systems in his doctoral dissertation [79] dating back to 1947.

If the dimension N of the signal space is greater than the dimension M of the message space, the mapping is a bandwidth expansion mapping, which can be regarded as analogue error control coding, since it adds redundancy and directly maps the message into the signal space without involving a codeword as does conventional digital coding. The opposite case of $M > N$ can be viewed as joint source-channel coding, as the message is compressed while the effects of both, compression errors and channel noise are minimized, provided that the mapping is properly designed, as Cai and Modestino point out in [2].

Examples of $M = 1 : N = 2$ mappings that have been suggested in [2, 3, 57] are shown in Figures 4.11 to 4.15.

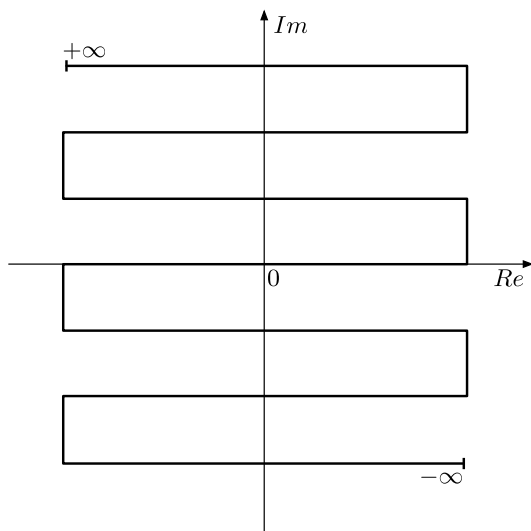


Figure 4.11: A continuous Shannon mapping curve [2].

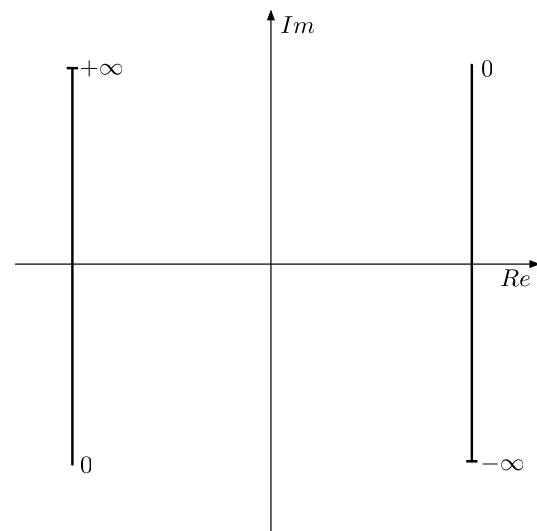


Figure 4.12: A 2-split Shannon mapping curve [57].

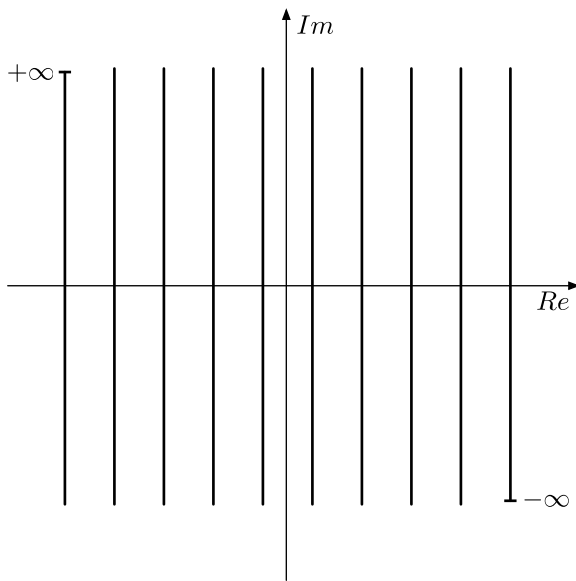


Figure 4.13: An m -split Shannon mapping curve [57]. ($m = 10$)

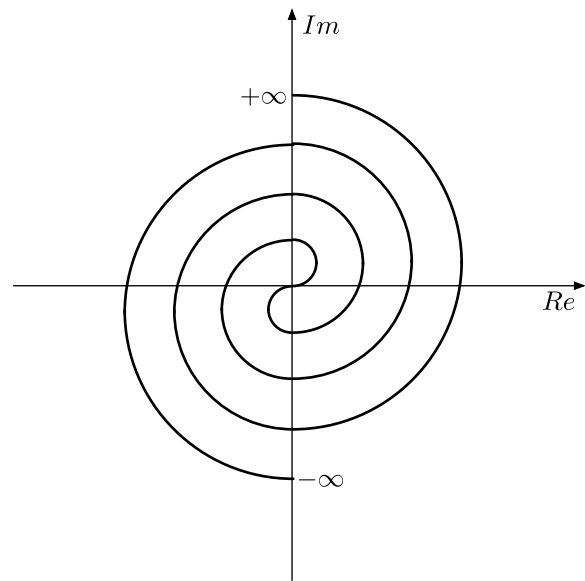


Figure 4.14: A Shannon mapping curve made from half circles [57].

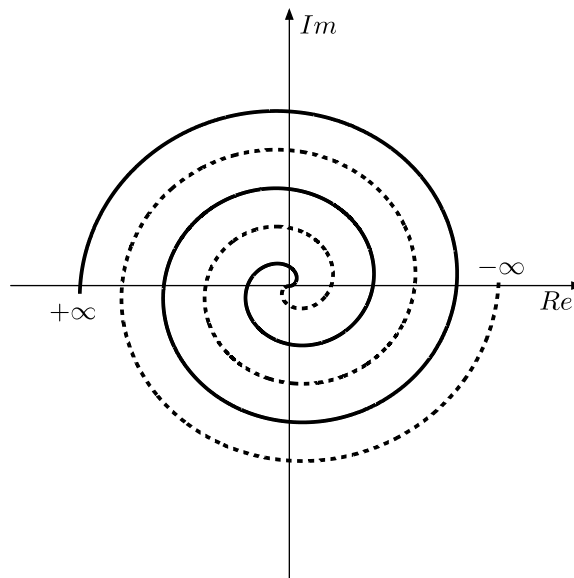


Figure 4.15: Archimedes Spiral Mapping as defined in Section 4.2.4. Here, in contrast to the curve from Figure 4.14, the radius grows steadily with the angle.

MAD transmission using Archimedes Spiral Mapping (ASM) for the pseudo analogue residual was proposed in [9] and will be summarized in the following. This mapping has been chosen as it offers the longest possible locus with constrained power.

Baseband Transmission Model for QPSK/ASM

Regarded is the MAD transmission system as generally introduced in Chapter 3.3, with a complex baseband transmission model as depicted in Figures 4.16 (transmission) and 4.17 (receiver).

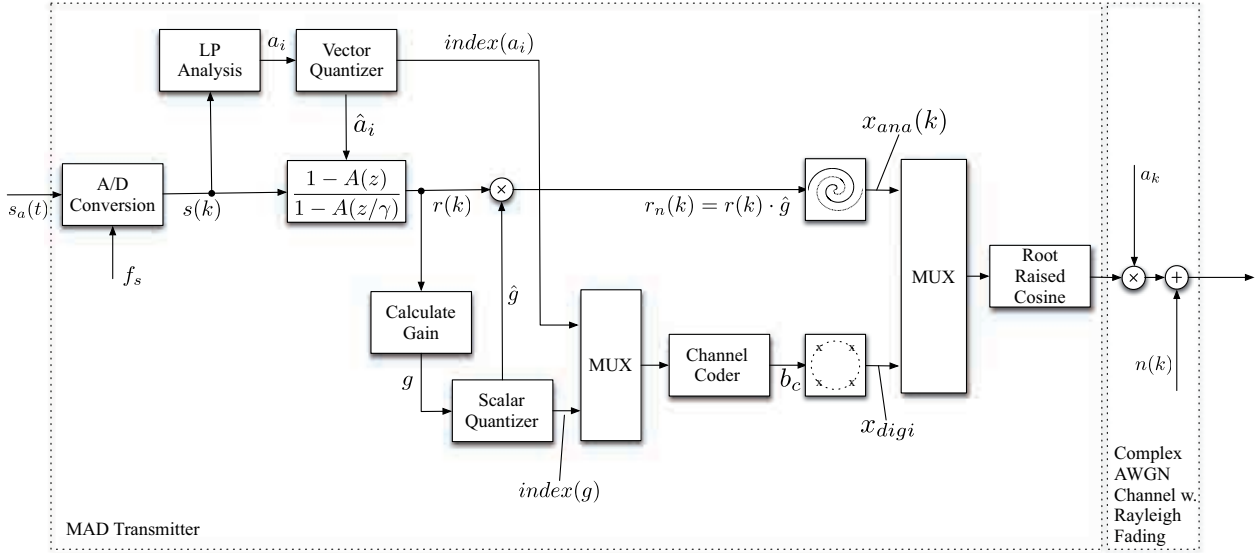


Figure 4.16: 2-dimensional MAD transmission: Transmitter and channel.

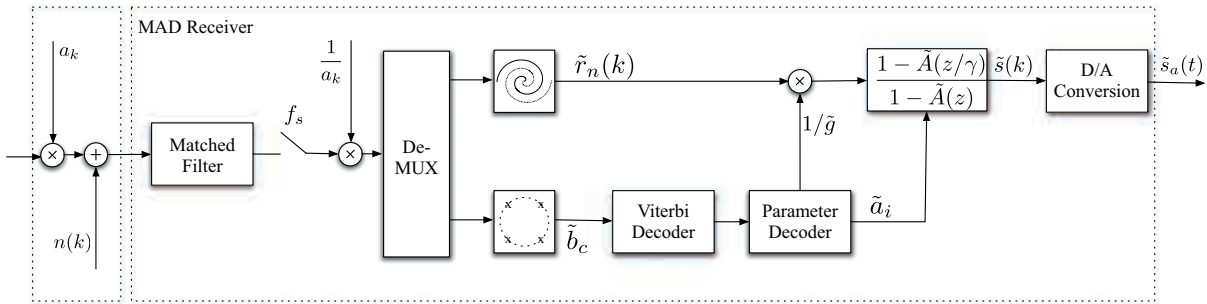


Figure 4.17: 2-dimensional MAD transmission: Receiver.

While for transmission of the digital information QPSK is used (see Section 2.3.2), the normalized, time-discrete, continuous-amplitude samples

$$r_n(k) = \hat{g} \cdot r(k) \tag{4.9}$$

are mapped to an Archimedes Spiral (see Figures 4.15 and 4.18). The Archimedes Spiral is defined in polar coordinates as

$$\varphi_{Ar}(r_n) = \begin{cases} \frac{r_n}{c} & \text{for } r_n \geq 0 \\ \frac{|r_n|}{c} + \pi & \text{for } r_n < 0 \end{cases} \tag{4.10}$$

with angle φ_{Ar} , radius $|r_n|$ and a real constant c that defines the tightness of the spiral [63]. Taking (4.10) into account, the complex signal samples transmitted equal

$$x_{ana}(k) = r_n(k) \cdot \exp(j \frac{|r_n(k)|}{c}). \quad (4.11)$$

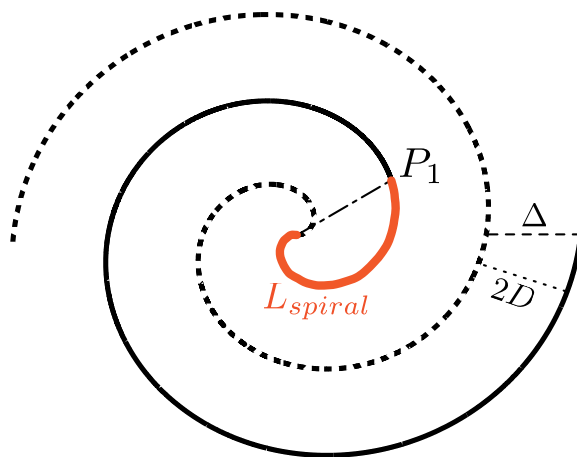


Figure 4.18: Archimedes Spiral. $r_n \geq 0$ solid, $r_n < 0$ dashed. L_{spiral} denotes the length of the locus from the center to P_1 . Δ is the radial distance between two branches while $2D$ represents the shortest distance between two branches.

ASM uses exactly the same transmit power as the PAM described above, regardless of the constant c , as the amplitudes r_n of PAM and ASM are equal. The transmission bandwidth of PAM and ASM also are identical, as duration and pulse shaping remain equal while the transmitted signal is altered in phase only. The complex signal is fed to the Root Raised Cosine filter in addition (time multiplex) to the digital QPSK signal \mathbf{x}_{digi} and transmitted over the AWGN channel. Again, instead of quantization noise there is channel noise. The required two-sided low pass bandwidth for the complex signal equals

$$B' = B'_{ana} + B'_{digiQPSK} = (1+\alpha) \cdot (R_{ana} + 0.5R_{digi}) = 1.5(R_{ana} + 0.5R_{digi}) \quad (4.12)$$

with R_{ana} the analogue sample rate, R_{digi} the digital bit rate, and QPSK transmitting 2 bit per cycle.

Properties of the Archimedes Spiral

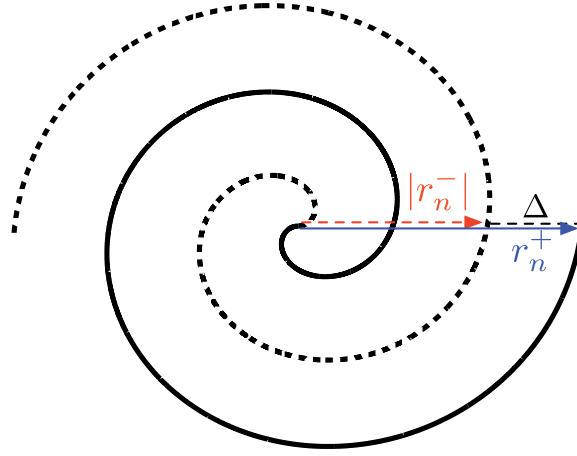


Figure 4.19: Archimedes Spiral. $r_n^+ \geq 0$ solid, $r_n^- < 0$ dashed. Δ is the radial distance between two branches.

The radial distance Δ between the two parts of the spiral representing positive and negative values r (solid and dashed line in Figures 4.19) is constant. From the definitions in Equations (4.11) and (4.10) follows for a positive sample r_n^+ and a negative sample r_n^- at the same angle $\varphi_{Ar}(r_n^+) = \varphi_{Ar}(r_n^-) = \varphi_{Ar}$ follows

$$\frac{r_n^+}{c} = \frac{|r_n^-|}{c} + \pi \quad (4.13)$$

thus

$$r_n^+ = |r_n^-| + \pi c \quad (4.14)$$

and

$$\Delta = r_n^+ - |r_n^-| = \pi c. \quad (4.15)$$

Thus, the smaller c the tighter the spiral.

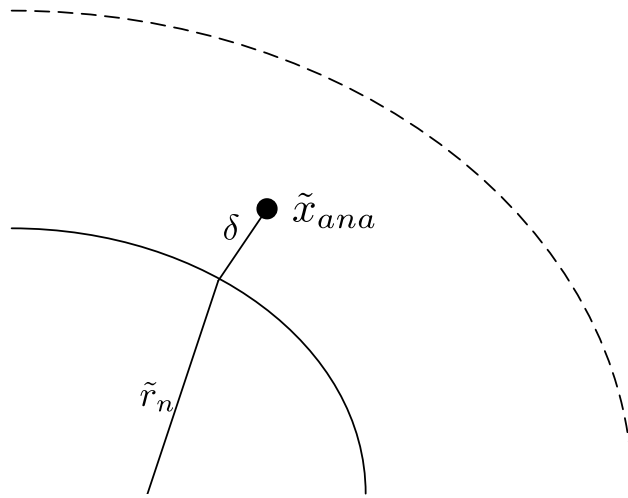


Figure 4.20: Archimedes Spiral. Decoding of a received symbol \tilde{x}_{ana} into \tilde{r}_n .

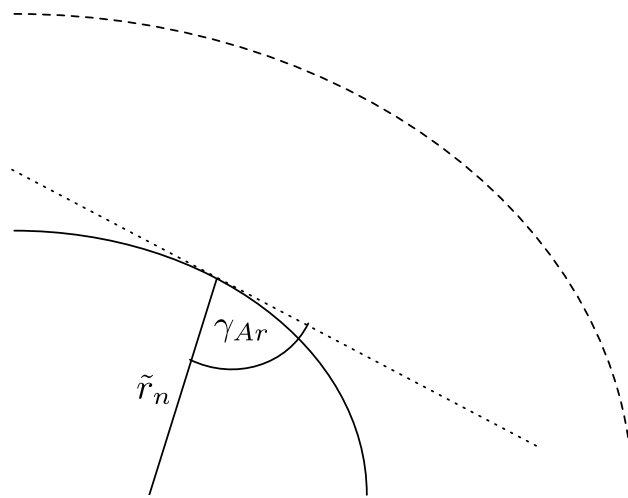


Figure 4.21: Archimedes Spiral. Angle $\gamma_{Ar} < \frac{\pi}{2}$ between a radius and the tangent of the spiral.

To calculate the length L_{spiral} of the locus, the differential length

$$\frac{dL}{d\varphi} = \sqrt{\left(\frac{dr_n}{d\varphi}\right)^2 + r_n^2} \quad (4.16)$$

can be integrated [85, 88]:

$$L_{spiral} = \int_{\varphi_1}^{\varphi_0} \sqrt{\left(\frac{dr_n}{d\varphi}\right)^2 + r_n^2} d\varphi \quad (4.17)$$

$$= \int_{\varphi_1}^{\varphi_0} \sqrt{c^2 + c^2 \varphi^2} d\varphi \quad (4.18)$$

$$= c \cdot \int_{\varphi_1}^{\varphi_0} \sqrt{1 + \varphi^2} d\varphi \quad (4.19)$$

$$= c \cdot \left[\frac{\varphi}{2} \sqrt{1 + \varphi^2} + \frac{1}{2} \ln(\varphi + \sqrt{1 + \varphi^2}) \right]_{\varphi_0}^{\varphi_1} \quad (4.20)$$

which results in a total length from the origin ($\varphi_0 = 0$) to a point P_1 at the angle φ_1 of

$$L_{spiral} = c \left[\frac{\varphi_1}{2} \sqrt{1 + \varphi_1^2} + \frac{1}{2} \ln(\varphi_1 + \sqrt{1 + \varphi_1^2}) \right] \geq |r_n(P_1)|. \quad (4.21)$$

Figure 4.22 shows L_{spiral} over c for a normalized radius $r_n = 1$.

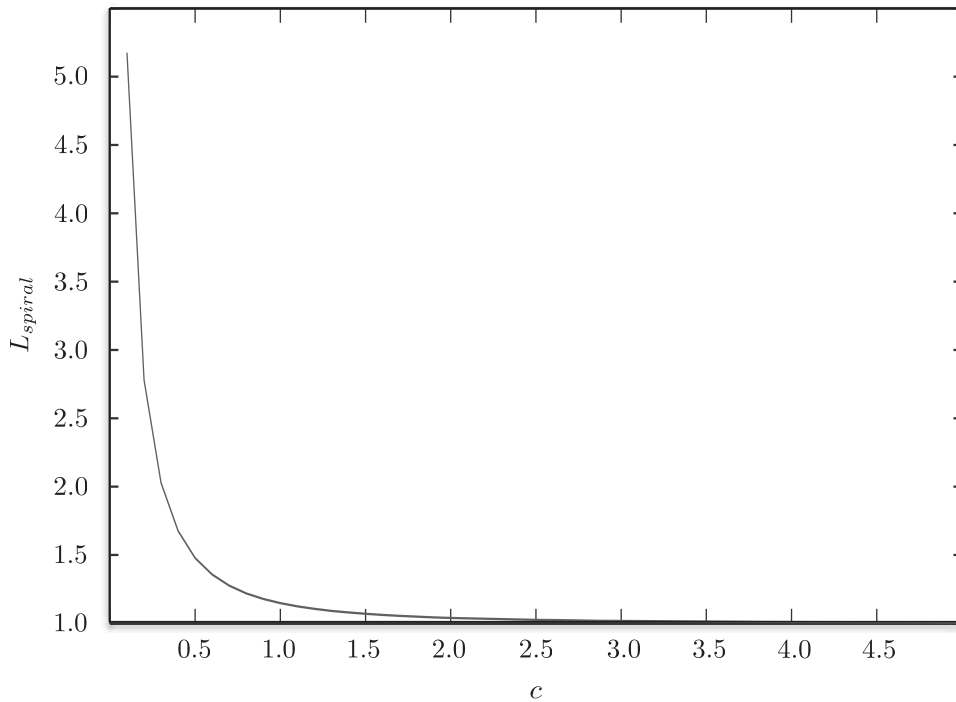


Figure 4.22: Length L_{spiral} over the tightness factor c for a normalized radius of $r_n = 1$.

The angle γ_{Ar} between a radius and the tangent of the spiral at this radius is [69, 85]

$$\gamma_{Ar} = \arctan(\varphi_{Ar}) = \arctan(r_{Ar}/c), \quad (4.22)$$

see Figure 4.21.

To decode the received complex signal samples \tilde{x}_{ana} , the receiver finds the closest distance δ to the locus of the Archimedes Spiral and decodes the corresponding radius \tilde{r}_n (see Figure 4.20).

Decoding of a distorted signal results in two different possible errors:

- I A smaller displacement of \tilde{x}_{ana} on the correct branch of the spiral (weak noise case), or
- II a much larger displacement of \tilde{x}_{ana} due to selection of a wrong branch (strong noise case, threshold effect).

If only errors of type I occur, the mean displacement of ASM

$$dis_{ASM} = |r_n - \tilde{r}_{n,ASM}| \quad (4.23)$$

due to the added noise will be lower than that of PAM

$$dis_{ASM} \leq dis_{PAM} = |r_n - \tilde{r}_{n,PAM}|, \quad (4.24)$$

as the length d_{spiral} of the spiral is greater than the PAM amplitude r_n and thus the noise is compressed by the mapping from the Archimedes Spiral to \tilde{r}_n , as described above.

As the branches of the spiral are equally spaced at the distance $2D$ (see Figure 4.18), the probability $P_e(D)$ of choosing an incorrect branch in the outer branches, i.e. $\varphi_{Ar} > \pi$, is for errors of type II:

$$P_e(D) = 1 - \operatorname{erf}\left(\frac{D}{\sqrt{2}\sigma_n}\right) = \operatorname{erfc}\left(\frac{D}{\sqrt{2}\sigma_n}\right) \quad (4.25)$$

with noise power σ_n^2 . For the Definition of $\operatorname{erf}(\cdot)$ and $\operatorname{erfc}(\cdot)$ refer to Appendix B.

Figures 4.23 to 4.25 show the effect of AWGN with 10dB SNR on the transmitted spiral for different values of c . The choice of the tightness factor c in presence of noise is looked at in more detail in Section 5.4.4.

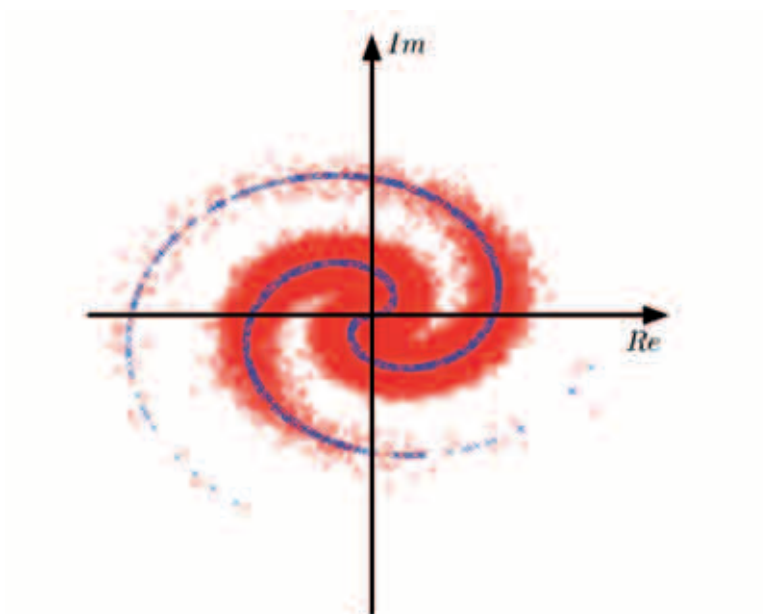


Figure 4.23: ASM with SNR = 10 dB and $c = 0.5$.

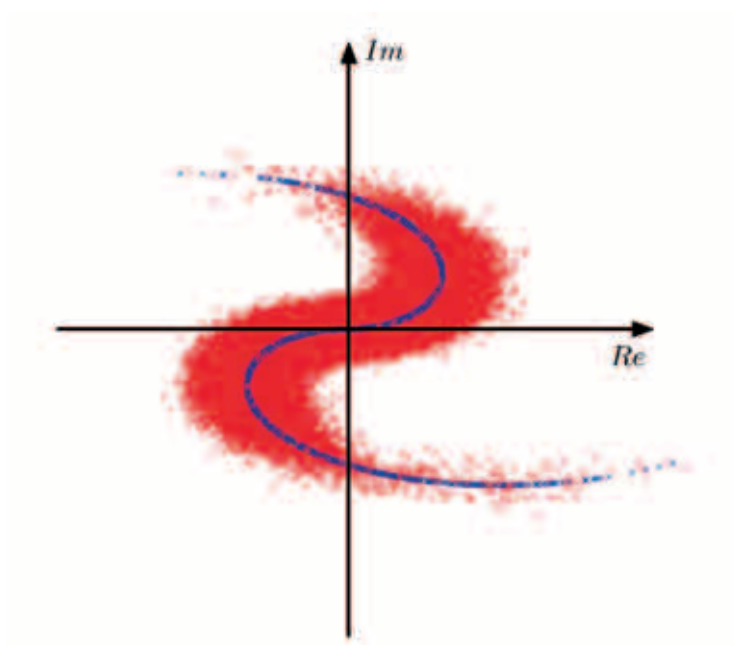


Figure 4.24: ASM with SNR = 10 dB and $c = 1.5$.

MAD Transmission with QPSK/ASM

Using 2-dimensional QPSK instead of BPSK for transmission of the digital information does not change the SNR at the receiver output. However, with the symbol rate reduced by a factor of two, the required bandwidth is cut in half. This means that in the case of wideband MAD Transmission the channel bandwidth, required for digital transmission of LP coefficients and gains, can be reduced from

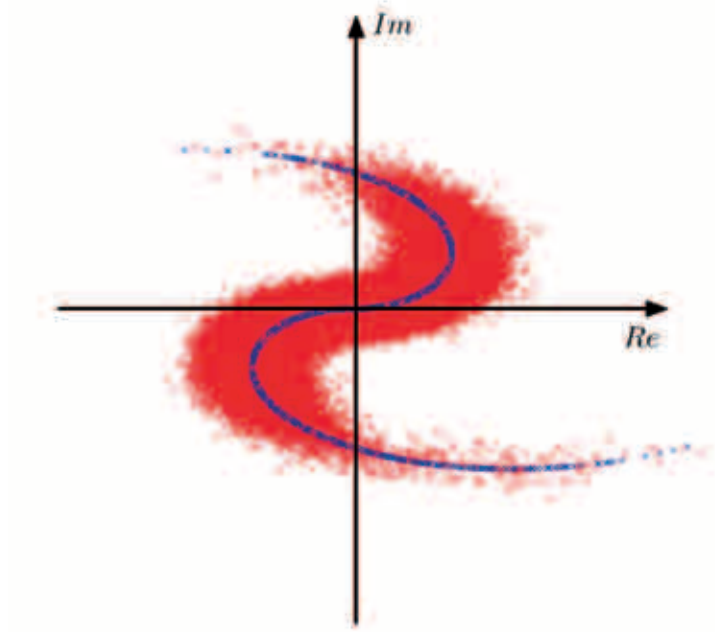


Figure 4.25: ASM with SNR = 10 dB and $c = 2.5$.

$$B'_{digi_{BPSK}} = 10.5 \text{ kHz} \quad (4.26)$$

to

$$B'_{digi_{QPSK}} = 5.25 \text{ kHz}. \quad (4.27)$$

Mapping the time-discrete quasi-continuous-amplitude samples of the residual signal to the Archimedes Spiral does not affect the required bandwidth as only the phase is changed and not the transmission rate. Thus, the complete channel bandwidth required for MAD wideband speech transmission is reduced from

$$B'_{BPSK} = (24 + 10.5) \text{ kHz} = 34.5 \text{ kHz} \quad (4.28)$$

to

$$B'_{QPSK} = (24 + 5.25) \text{ kHz} = 29.25 \text{ kHz}. \quad (4.29)$$

with $B'_{ana} = 1.5 \cdot 16 \text{ kHz} = 24 \text{ kHz}$.

The effect of mapping the residual to an Archimedes Spiral can best be studied in Figure 4.26 which shows the output SNR of the speech signal over the channel E_s/N_0 for different constants c . While for good channels c may be smaller (and thus the spiral tighter, which again means L is longer) to have a higher gain in SNR, it must be increased with the noise power.

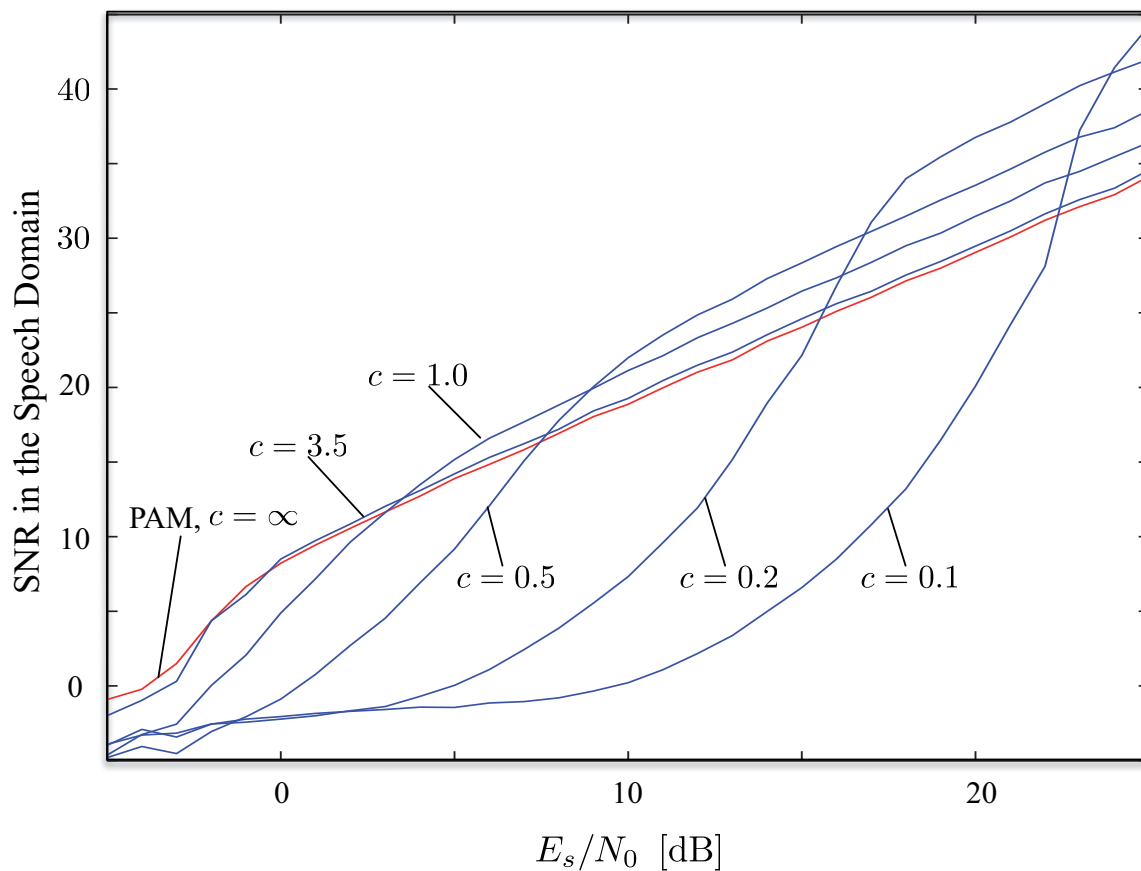


Figure 4.26: Effect of ASM.

Figure 4.27 shows the speech quality achieved with the new modulation scheme in comparison to MAD coding with BPSK/PAM.

A more thorough theoretical review on ASM will be given in Section 5.4.4.

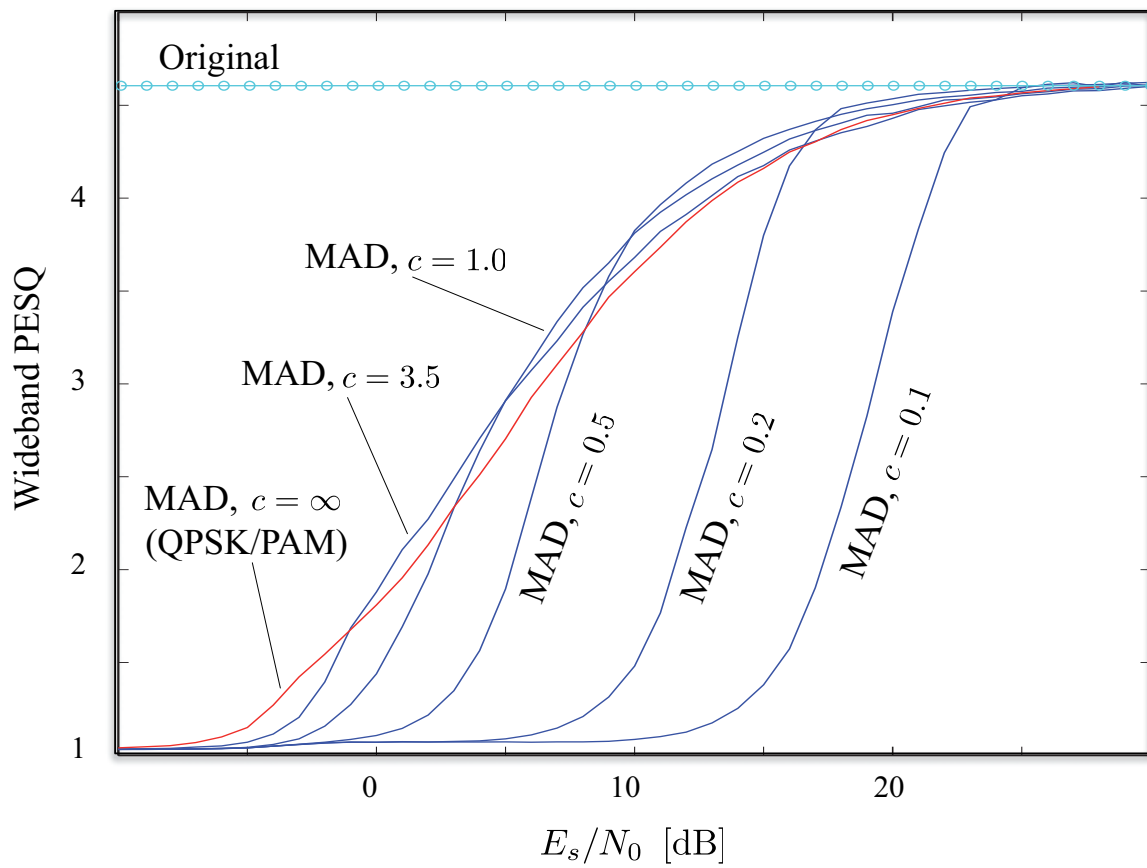


Figure 4.27: Evaluation of MAD speech coding with QPSK/ASM.

5

Information Theoretic Evaluation of MAD Transmission

This chapter presents the information theoretic background of the MAD concept which has been described in Chapters 3 and 4. A few basic concepts of information theory and rate distortion theory will be introduced first.

5.1 Introduction to Information Theory

Credited mostly to Claude E. Shannon and his famous paper [86], information theory tries to analyze and explain the principles and bounds of description and transmission of information. Raymond W. Yeung sketches an exemplary scenario [94]: “What is the maximum amount of meaningful information which can be conveyed on one page of facsimile - limited by the finite resolution of the fax machine and possible transmission errors, supported by the fact that a reader would still be able to read a text if a number of letters can no longer be deciphered. Information theory provides an answer to this and similar questions.”

In the following, a discrete random variable X with alphabet \mathcal{X} and probability

$$P_X(x) = \Pr \{X = x\}, \quad x \in \mathcal{X} \tag{5.1}$$

is considered.

5.1.1 Entropy, Mutual Information

To measure the average uncertainty of a random variable X with probability $P(x)$ ¹ the entropy is defined to be (e.g., [66])

$$H(X) = - \sum_{x \in \mathcal{X}} P(x) \cdot \text{ld}(P(x)). \quad (5.2)$$

The entropy represents the amount of information inherent in the random variable.

Extending this definition to a pair of random variables the joint entropy is obtained as

$$H(X, Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x, y) \cdot \text{ld}(P(x, y)) \quad (5.3)$$

with joint probability $P(x, y)$.

With the entropy $H(X)$ being the uncertainty of a single random variable, a conditional entropy

$$\begin{aligned} H(X|Y) &= \sum_{y \in \mathcal{Y}} P(y) H(X|Y = y) & (5.4) \\ &= - \sum_{y \in \mathcal{Y}} P(y) \sum_{x \in \mathcal{X}} P(x|y) \cdot \text{ld}(P(x|y)) \\ &= - \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} P(x, y) \cdot \text{ld}(P(x|y)) \end{aligned}$$

can be defined as the entropy of one random variable X given another Y . This conditioning reduces the uncertainty of the first random variable by an amount $I(X; Y)$:

$$I(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x, y) \cdot \text{ld} \left(\frac{P(x, y)}{P(x)P(y)} \right) \quad (5.5)$$

$$= H(X) - H(X|Y) \quad (5.6)$$

$$= H(Y) - H(Y|X) \quad (5.7)$$

which is called the mutual information, that is the relative entropy (or *Kullback Leibler distance*) between $P(x, y)$ and $P(x)P(y)$. The relationship between entropy and mutual information is illustrated in Figure 5.1.

¹For brevity of the notation the subscript is skipped if the meaning is unambiguous.

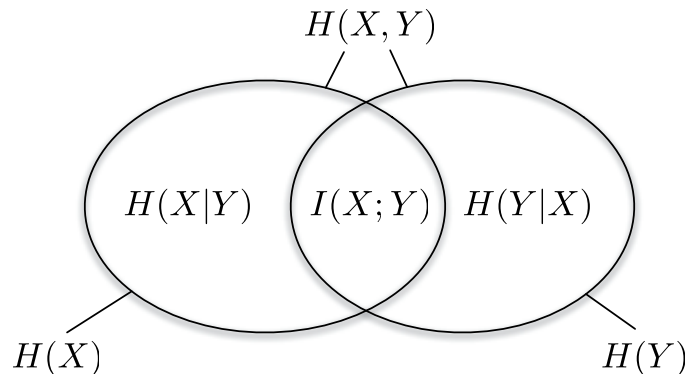


Figure 5.1: Relationship between entropy and mutual information

5.1.2 Channel, Channel Capacity

Considering the transport of information we follow [66]. Given is a discrete memoryless channel with an input alphabet \mathcal{X} , an output alphabet \mathcal{Y} , and a probability transition matrix $\mathbf{P}(y|x)$ that defines the probability of an output y following an input x , independent of the previous channel inputs and outputs.

The channel capacity in *bits per sample* of this channel is defined as

$$C' = \max_{P(x)} I(X; Y), \quad (5.8)$$

where the maximum is taken over all possible input distributions $p(x)$.

Assuming a bit rate $1/T$ one can rewrite (5.8) as

$$C = \max_{P(x)} \frac{I(X; Y)}{T} \quad (5.9)$$

in *bits per second*.

5.1.3 Channel Capacity of an AWGN Channel for Gaussian or Binary Input

Gaussian Input

Regarding an AWGN channel as depicted in Figure 5.2 with the noise $n(t)$ being statistically independent of the signal x , the conditional entropy $H(Y|X)$ can be calculated as (e.g., [92])

$$\begin{aligned} H(Y|X) &= - \int \int p_X(x)p(y|x) \cdot \text{ld}(p(y|x)) \, dx \, dy \\ &= - \int \int p_X(x)p_N(n) \cdot \text{ld}(p_N(n)) \, dx \, dn \\ &= - \int p_X(x) \, dx \int p_N(n) \cdot \text{ld}(p_N(n)) \, dn \\ &= H(N) \end{aligned} \quad (5.10)$$

with

$$p(y|x) = p_N(n = y - x), \quad (5.11)$$

$p_N(n = y - x)$ being the probability mass function of the noise signal n , due to $y = x + n$.

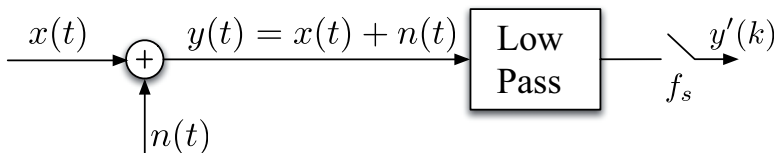


Figure 5.2: Simplified model of the AWGN channel with $B = B_{ana}$.

As the channel considered is band limited with bandwidth $B = B_{ana}$, an equivalent symbol rate (sampling rate) can be expressed to be

$$f_s = 2B_{ana} = 1/T_{ana} \quad (5.12)$$

and thus

$$\frac{I(X;Y)}{T} = 2B_{ana}(H(Y) - H(N)) \quad (5.13)$$

Note that, following the relevant literature, the single-sided low pass bandwidth B is used for theoretical analysis, while the two-sided low pass bandwidth B' which is relevant for transmission (and thus for frequency assignment of the regulatory bodies) is used in chapters 3 and 4. Also, for theoretical analysis, an ideal low pass filter (root raised cosine / matched filter with $\alpha = 0$) is assumed.

With the entropy of white Gaussian noise with variance σ_n^2 being ([66, 92])

$$H(N) = \frac{1}{2} \cdot \text{ld}(2\pi e\sigma_n^2) \quad (5.14)$$

and the following conditions

- $y(t)$ has limited power
- $H(Y)$ is maximized if $y(t) = x(t) + n(t)$ is Gaussian [66]
- $y(t)$ is Gaussian if $x(t)$ is Gaussian, too
- $\sigma_y^2 = \sigma_x^2 + \sigma_n^2$ as x and n are not correlated

the maximum mutual information (which represents the capacity of the band limited AWGN channel) is

$$\begin{aligned}
\frac{I(X;Y)}{T} &= 2B_{ana}(H(Y) - H(N)) & (5.15) \\
&= 2B_{ana} \left(\frac{1}{2} \cdot \text{ld} (2\pi e(\sigma_x^2 + \sigma_n^2)) - \frac{1}{2} \cdot \text{ld}(2\pi e\sigma_n^2) \right) \\
&= 2B_{ana} \cdot \frac{1}{2} \cdot \text{ld} \left(\frac{2\pi e(\sigma_x^2 + \sigma_n^2)}{2\pi e\sigma_n^2} \right) \\
&= 2B_{ana} \cdot \frac{1}{2} \cdot \text{ld} \left(1 + \frac{S}{N_{ana}} \right) \\
&= B_{ana} \cdot \text{ld} \left(1 + \frac{S}{N_{ana}} \right) \\
&= B_{ana} \cdot \text{ld} \left(1 + \frac{E_S \cdot T_{ana}}{T_{ana} \cdot N_0} \right) \\
&= B_{ana} \cdot \text{ld} \left(1 + \frac{E_S}{N_0} \right) \\
&= C_{Gauss} \text{ in bits per second} & (5.16)
\end{aligned}$$

or

$$C'_{Gauss} = \frac{1}{2} \cdot \text{ld} \left(1 + \frac{S}{N} \right) \text{ in bits per source letter.} \quad (5.17)$$

Binary Input

Considering a binary input signal

$$x_b(t) \in \{x_1, x_2\} = \{-1, +1\} \quad (5.18)$$

with symbol rate

$$f_s = 2B_{digi} = \frac{1}{T_{digi}}, \quad (5.19)$$

additive white Gaussian noise $n(t)$ with normalized noise power

$$\sigma_n^2 = N_{digi} \cdot T_{digi} = \frac{N_0}{2} \quad (5.20)$$

and equal probability of both elements

$$P(x_1 = -1) = P(x_2 = +1) = \frac{1}{2} \quad (5.21)$$

leads to

$$H(X) = 1 \quad (5.22)$$

and

$$H(X|Y) = - \sum_{i=1}^2 \int_{-\infty}^{\infty} p(x_i, y) \cdot \text{ld}(P(x_i|y)) \, dy \quad (5.23)$$

$$= - \sum_{i=1}^2 P(x_i) \int_{-\infty}^{\infty} p(y|x_i) \cdot \text{ld}(P(x_i|y)) \, dy. \quad (5.24)$$

(e.g., [92]). The channel capacity for binary transmission is derived from Equations (5.7) and (5.21):

$$C_{bin} = 2B_{digi} \cdot \left[1 + \sum_{i=1}^2 \frac{1}{2} \cdot \int_{-\infty}^{\infty} p(y|x_i) \cdot \text{ld}(P(x_i|y)) \, dy \right] \quad (5.25)$$

The a-posteriori probabilities follow as

$$P(x|y) = \left(1 + \left(\frac{p(y|x = -1)}{p(y|x = +1)} \right)^x \right)^{-1} \quad (5.26)$$

$$= \left(1 + e^{\frac{-2y \cdot x}{\sigma_n^2}} \right)^{-1} \quad (5.27)$$

due to

$$p(y|x) = \frac{1}{\sqrt{2\pi}\sigma_n} \cdot e^{-\frac{(y-x)^2}{2\sigma_n^2}} \quad (5.28)$$

in combination with (5.20).

Assuming a symmetrical channel, that is $p(y|x) = p(-y|-x)$, and doubling the result for, e.g., $x = -1$, the channel capacity becomes

$$C_{bin} = 2B_{digi} \cdot \left[1 - \frac{1}{\sqrt{2\pi}\sigma_n} \cdot \int_{-\infty}^{\infty} e^{-\frac{(y+1)^2}{2\pi\sigma_n^2}} \cdot \text{ld} \left(1 + e^{\frac{2y}{\sigma_n^2}} \right) \, dy \right]. \quad (5.29)$$

or

$$C'_{bin} = \left[1 - \frac{1}{\sqrt{2\pi}\sigma_n} \cdot \int_{-\infty}^{\infty} e^{-\frac{(y+1)^2}{2\pi\sigma_n^2}} \cdot \text{ld} \left(1 + e^{\frac{2y}{\sigma_n^2}} \right) \, dy \right]. \quad (5.30)$$

The integral has no closed solution, so it has to be solved numerically. Figure 5.3 shows the capacities C_{Gauss} and C_{bin} of an AWGN channel normalized to the Bandwidth B . As the maximum mutual information is achieved with Gaussian input, it follows that

$$C_{Gauss} > C_{bin} \quad (5.31)$$

for all SNR².

²It may be noted that while C_{Gauss} is the maximum achievable capacity for an additive white Gaussian noise channel, for other noise characteristics different input distributions will be optimum.

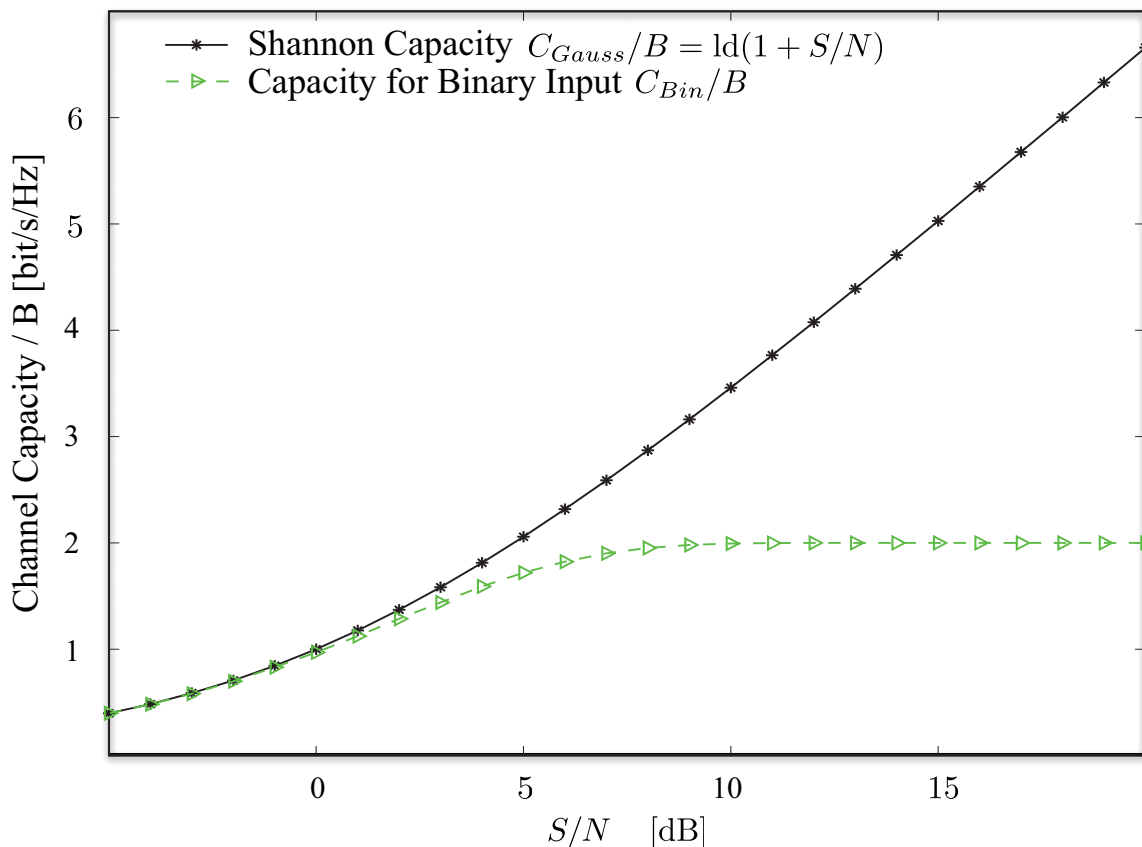


Figure 5.3: Capacities of an AWGN channel with Gaussian input C_{Gauss} , and binary input C_{bin} , respectively.

5.1.4 Capacity of Multiplexed Channels

The main characteristic of MAD transmission is the split of the information into the spectral envelope, which shall be transmitted over a digital channel, and the residual signal, which is transmitted over a pseudo analogue channel (for details see Section 5.4.1). The digital and the pseudo analogue transmission can either be carried out in different frequency bands (parallel channels), or they can be multiplexed in time.

Parallel Transmission

Assuming parallel transmission over band limited channels using non-overlapping frequency bands with sufficient channel spacing, the transmission of digital and of pseudo analogue information do not interfere. Thus the total transmission bandwidth for parallel transmission $B_{tot,p}$ is the sum of the bandwidth used for each channel:

$$B_{tot,p} = B_{digi} + B_{ana} \quad (5.32)$$

The same holds for the total capacity C_{tot} :

$$C_{tot,p} = C_{digi} + C_{ana} \quad (5.33)$$

If the digital information has a binary representation with probabilities $P(x_1)$ and $P(x_2)$ as given in (5.21), and the pseudo analogue input signal $r(k)$ is assumed to have a Gaussian pdf, (5.33) can be rewritten as

$$C_{tot,p} = C_{bin} + C_{Gauss} \quad (5.34)$$

and thus, with (5.32), (5.29), and (5.16):

$$\begin{aligned} \frac{C_{tot,p}}{B_{tot,p}} = & 2 \cdot \frac{B_{digi}}{B_{tot,p}} \cdot \left[1 - \frac{1}{\sqrt{2\pi}\sigma_n} \cdot \int_{-\infty}^{\infty} e^{-\frac{(y+1)^2}{2\pi\sigma_n^2}} \cdot \text{ld} \left(1 + e^{\frac{2y}{\sigma_n^2}} \right) dy \right] \\ & + \frac{B_{tot,p} - B_{digi}}{B_{tot,p}} \cdot \text{ld} \left(1 + \frac{E_S}{N_0} \right). \end{aligned} \quad (5.35)$$

Figures 5.4 to 5.6 show the behaviour of (5.35) for different signal-to-noise ratios.

Deriving $\frac{C_{tot,p}}{B_{tot,p}}$ yields

$$\frac{\partial \left\{ \frac{C_{tot,p}}{B_{tot,p}} \right\}}{\partial \left\{ \frac{B_{digi}}{B_{tot,p}} \right\}} = 2 \cdot \left[1 - \frac{1}{\sqrt{2\pi}\sigma_n} \cdot \int_{-\infty}^{\infty} e^{-\frac{(y+1)^2}{2\pi\sigma_n^2}} \cdot \text{ld} \left(1 + e^{\frac{2y}{\sigma_n^2}} \right) dy \right] - \text{ld} \left(1 + \frac{E_S}{N_0} \right). \quad (5.36)$$

And with (5.31)

$$\frac{\partial \left\{ \frac{C_{tot,p}}{B_{tot,p}} \right\}}{\partial \left\{ \frac{B_{digi}}{B_{tot,p}} \right\}} < 0. \quad (5.37)$$

There are two important characteristics to note:

- a) For the exclusive transmission of pseudo analogue or digital information only

$$C_{ana} > C_{digi} \quad (5.38)$$

for all channel conditions (due to (5.37)).

- b) With rising E_s/N_0 , C_{ana} will rise faster than C_{digi} (see Figure 5.3), thus the slope of C_{tot} will be steeper. This can be studied in Figures 5.4 to 5.6.

The effect of these results will be detailed in Section 5.4.3.

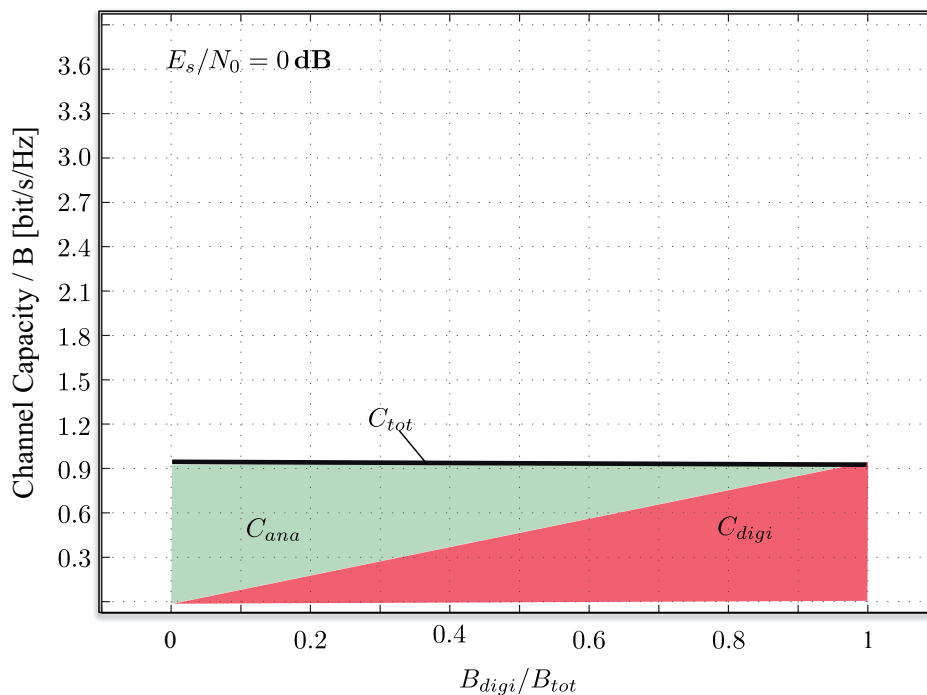


Figure 5.4: Total capacity of an AWGN channel with parallel Gaussian and binary input and $E_s/N_0 = 0$ dB.

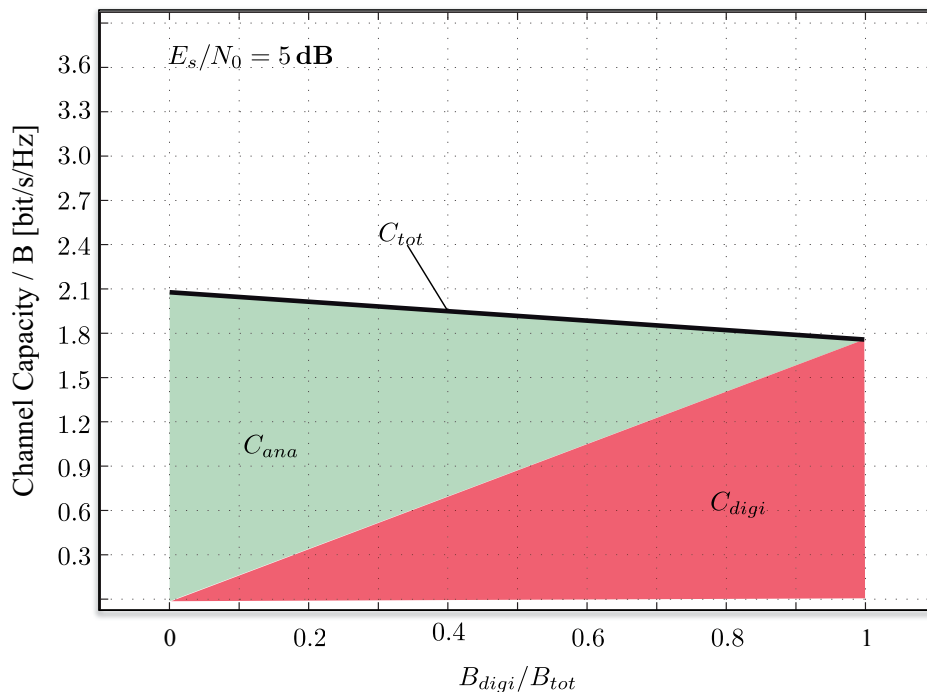


Figure 5.5: Total capacity of an AWGN channel with parallel Gaussian and binary input and $E_s/N_0 = 5$ dB.

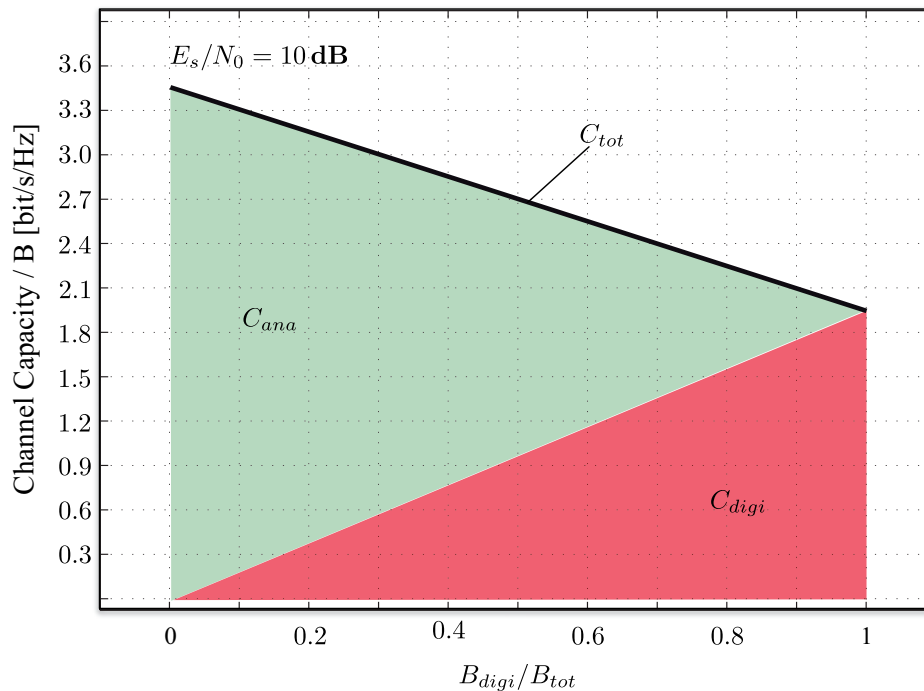


Figure 5.6: Total capacity of an AWGN channel with parallel Gaussian and binary input and $E_s/N_0 = 10$ dB.

Interleaved Transmission

As shown in [64] and [1] the interleaved transmission of digital and pseudo analogue information can be regarded as an alternating transmission and non-transmission on each of the channels, with the transmissions sharing the total time of transmission t_{tot} . Avoiding inter-symbol interference by, e.g., root raised cosine filters, the description of interleaved transmission is equivalent to that of parallel transmission over two (logical) channels.

Considering an AWGN channel and equal average power in digital and pseudo analogue transmission, multiplexing as suggested in Figure 3.4 is feasible. This multiplexing scheme, however, results in a fixed relation of the effective bandwidth used for digital transmission and bandwidth used for pseudo analogue transmission defined by the number of bits versus the number of samples within a given time interval. This effect is detailed Section 5.4.3.

5.2 Rate Distortion Theory

The previous sections have shown how the capacity of a transmission system depends on the SNR on the channel (see Figure 5.3). The minimum rate R for error-free representation of a source, however, is given by the entropy H of the source and is constant over all SNR.

Rate Distortion Theory answers the question which rate (quantization or transmission) results in which distortion (Rate-Distortion-Function, RDF) or - conversely - which distortion requires which rate (Distortion-Rate-Function, DRF).

To understand the idea of Rate-Distortion-Function or Distortion-Rate-Function a distortion function $d(x, y)$ needs to be defined, which measures the cost of representing the symbol x by the symbol y . Examples of common distortion functions are [66]:

- Hamming distortion

$$d(x, y) = \begin{cases} 0 & : x = y \\ 1 & : x \neq y \end{cases} \quad (5.39)$$

which results in a probability of error distortion, and

- Squared error distortion

$$d(x, y) = (x - y)^2 \quad (5.40)$$

which is the most popular distortion measure for analogue samples.

Further distortion measures are analyzed, e.g., in [38, 39]. Taking the expectation of (5.40) yields the MSE³ distortion:

$$d_{MSE} = E\{d(x, y)\} = E\{(x - y)^2\}. \quad (5.41)$$

Using the MSE distortion implies that there is a large punishment for one large error while multiple small errors are not that critical. This is not always true for the human auditory system. E.g., a time or phase shifted version of a speech signal cannot be distinguished from the original by the human listener despite of having a rather large MSE distortion. Thus, for better evaluation, often more complex measures such as the Itakura-Saito distance [38] or perceptual measures of speech quality like PESQ ([107], see 2.4.1) are used.

The RDF defines the minimum of all rates R with which a given distortion D , e.g., MSE distortion, can be achieved. In other words, the $R(D)$ is given by the minimum mutual information between x and y :

$$R(D) = \min_{p(y|x): E\{d(y,x)\} \leq D} I(X; Y) \quad (5.42)$$

The following properties hold for the RDF:

- $R(D)$ is non-increasing in D
- $R(D)$ is convex
- $R(D) = 0$ for $D \geq D_{max}$
- $R(0) \geq H(x)$

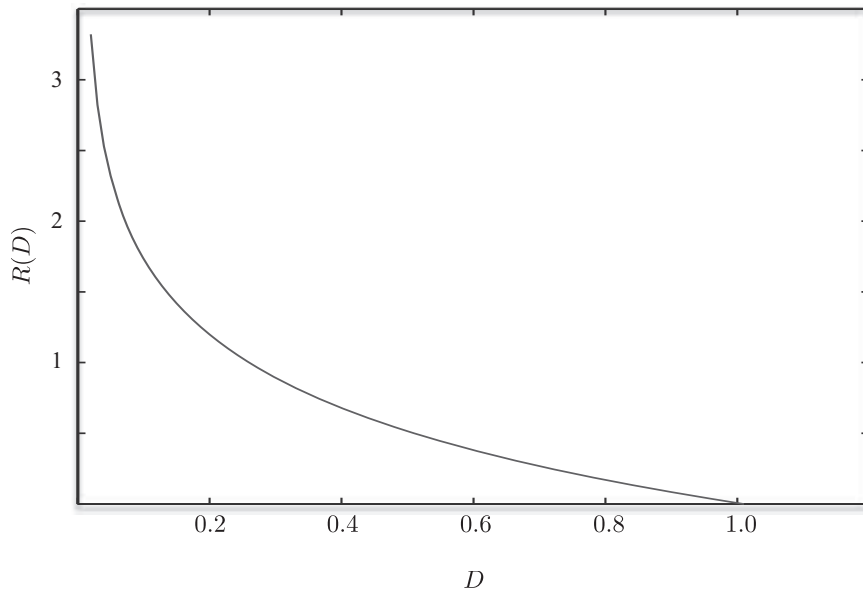


Figure 5.7: A Rate-Distortion Function $R(D)$.

An exemplary RDF is given in Figure 5.7.

The DRF on the other hand describes the minimum attainable distortion D for a given rate R . While the definition of RDF and DRF is straight forward, neither of them is generally simple to be calculated analytically and explicit calculations are in most cases impossible.

For a zero-mean, memoryless Gaussian source with variance σ_x^2 and MSE distortion the RDF is explicitly given (e.g. [62]):

$$R_G(D) = \begin{cases} \frac{1}{2} \log \frac{\sigma_x^2}{D} & : 0 \leq D \leq \sigma_x^2 \\ 0 & : D > \sigma_x^2 \end{cases} \quad (5.43)$$

The corresponding DRF can be calculated as

$$D_G(R) = \sigma_x^2 2^{-2R} \quad (5.44)$$

or in normalized form independent of the variance σ_x^2

$$10 \log \frac{D_G(R)}{\sigma_x^2} = -6.02 \frac{\text{dB}}{\text{bit}} \cdot R \quad [\text{dB}]. \quad (5.45)$$

For a more detailed elaboration the reader is referred to [62, 67, 94].

³Mean Square Error

5.3 Performance Limit

Equating the channel capacity C (Section 5.1) and rate distortion $R(D)$ (Section 5.2) the Optimum Performance Theoretically Attainable (OPTA) can be determined for Gaussian input:

$$\frac{1}{2} \cdot \text{ld} \frac{S}{D} = B \cdot T \cdot \text{ld} \left(1 + \frac{S}{N} \right) \quad (5.46)$$

The OPTA is an upper bound on the SNR for a certain transmission scenario.

The performance of a system can be measured in terms of the source Signal-to-Distortion Ratio (Inverse of (5.45)) for a given Channel SNR (CSNR). The Signal-to-Distortion Ratio (SDR) follows from (5.46)

$$SDR = \frac{S}{D} = \left(1 + \frac{S}{N} \right)^{2BT} \quad (5.47)$$

Goblick, Tufts, and Berger have shown in [29, 35, 62] that for a power constrained system transmitting samples of a memoryless Gaussian source over an AWGN channel at Nyquist rate

$$\frac{1}{T} = 2B, \text{ or } 2BT = 1, \quad (5.48)$$

PAM reaches the least mean square error theoretically attainable. With real $s(k)$, signal power $E_{s,real} = S \cdot T$ and noise power $N = 2 \cdot B \cdot \frac{N_0}{2} = \frac{N_0}{2T}$, S/N becomes

$$\frac{S}{N} = \frac{E\{s^2(k)\}}{E\{n^2(k)\}} = \frac{E_{s,real}}{\frac{\sigma_n^2}{2}} = \frac{2 \cdot E_{s,real}}{N_0} \quad (5.49)$$

With this, the SDR becomes

$$SDR_{PAM} = \frac{S}{D} = \left(1 + \frac{S}{N} \right) = \left(1 + \frac{2 \cdot E_{s,real}}{N_0} \right) \quad (5.50)$$

or, in dB, incorporating (5.17)

$$\begin{aligned} 10 \cdot \log_{10}(SDR_{PAM}) &= 10 \cdot \log_{10} \left(\frac{S}{D} \right) \\ &= 10 \cdot \log_{10}(2) \cdot \text{ld} \left(1 + \frac{S}{N} \right) \\ &= 6.02 \frac{\text{dB}}{\text{bit}} \cdot C'_{Gauss}. \end{aligned} \quad (5.51)$$

Considering 2-dimensional modulation, the real and imaginary parts of the noise, $\text{Re}\{n(k)\}$ and $\text{Im}\{n(k)\}$, are independantly Gaussian distributed with zero mean and variance

$$E\{\text{Re}\{n(k)\}^2\} = E\{\text{Im}\{n(k)\}^2\} \stackrel{!}{=} \frac{\sigma_n^2}{2}. \quad (5.52)$$

With complex transmit symbols $s(k)$ and coherent reception, S/N becomes

$$\frac{S}{N} = \frac{E\{s^2(k)\}}{E\{\text{Re}\{n(k)\}^2\} + E\{\text{Im}\{n(k)\}^2\}} = \frac{E_s}{\sigma_n^2} = \frac{E_s}{N_0}. \quad (5.53)$$

The difference to (5.49) stems from the fact, that the noise in the complex baseband is independant in real and imaginary part (see also [77]).

As with complex transmission two symbols can be transmitted at the same time, the rate doubles, so that $2BT' = 2$. The SDR follows to be

$$SDR = \frac{S}{D} = \left(1 + \frac{S}{N}\right)^2 = \left(1 + \frac{E_s}{N_0}\right)^2 \quad (5.54)$$

or, in dB,

$$10 \cdot \log_{10}(SDR) = 20 \cdot \log_{10}(2) \cdot \text{ld} \left(1 + \frac{S}{N}\right) = 6.02 \frac{\text{dB}}{\text{bit}} \cdot \text{ld} \left(1 + \frac{S}{N}\right). \quad (5.55)$$

Similarly, using (5.30), the SDR for a binary input signal is

$$10 \cdot \log_{10}(SDR) = 6.02 \frac{\text{dB}}{\text{bit}} \cdot C'_{bin}. \quad (5.56)$$

Figure 5.8 shows the optimum performance in terms of the SDR for the following scenarios:

- a) 2-dimensional modulation, OPTA for $2BT' = 2$.
- b) The AWGN channel with Gaussian input and C'_{Gauss} . This is the OPTA for $2BT = 1$.
- c) The AWGN channel with binary input and C'_{bin} from (5.30).

5.4 Information Theoretic Comparison to Digital Transmission

To explain the behaviour of MAD transmission more precisely, a simplified model of linear predictive speech and audio transmission, as detailed in the following, is regarded.

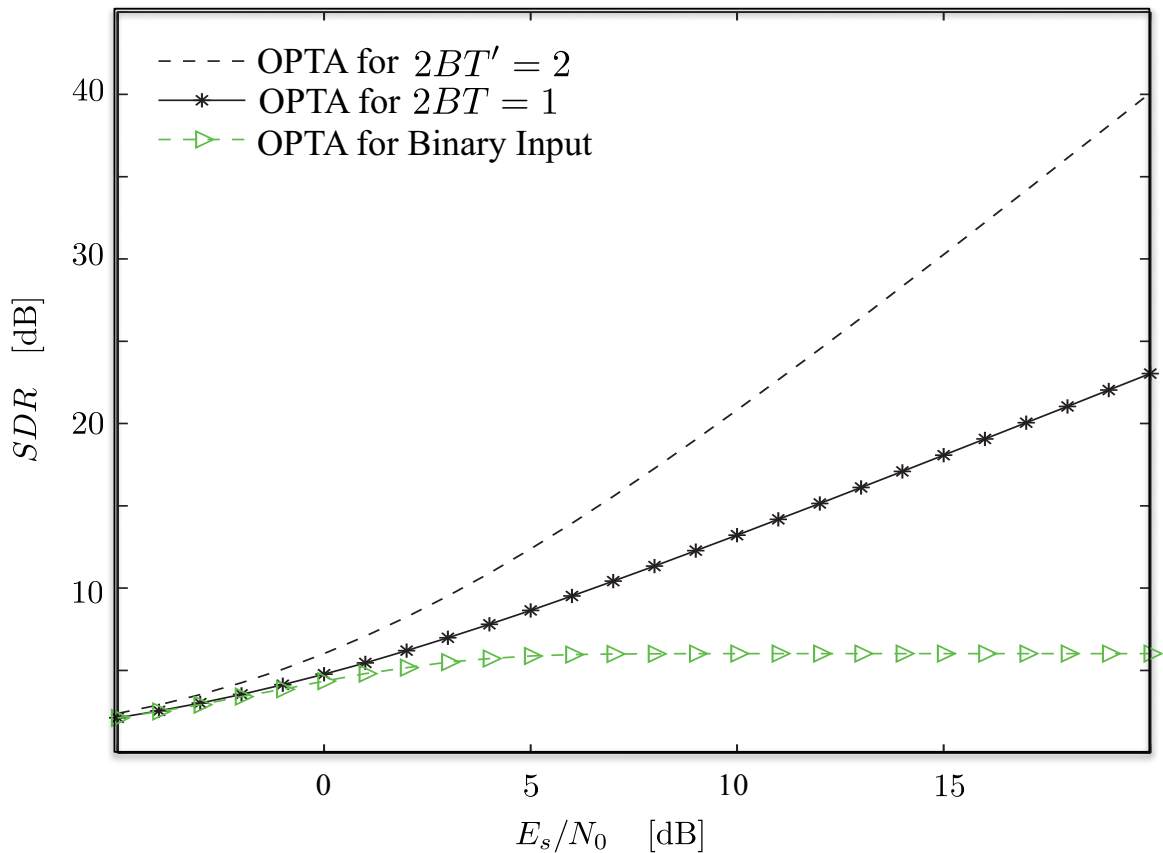


Figure 5.8: Comparison of OPTA for binary and Gaussian transmission over the AWGN channel.

5.4.1 Simplified Transmission Model

In linear predictive speech and audio coding, which is found in numerous coding standards [91], the speech signal is separated into the spectral envelope inherent in the LP filter coefficients, gains, and a residual signal. While the residual signal is usually transmitted digitally after quantization, it is transmitted as is in the MAD transmission system (see Chapter 3). To understand the effect of the MAD transmission principle, it can be assumed that the linear prediction of the MAD transmission system is equal to that of the purely digital system, thus, the difference being the transmission of the residual. Figure 5.9 illustrates this idea. This was also the main reason for choosing standard modules in Chapter 3.3 instead of optimizing the LP filter and channel coding for MAD transmission.

5.4.2 Rate Split between LPC and Residual Signal

So far, the residual signal has been in the focus to point out the different behaviour of general digital coding and transmission and the MAD principle, as the digital transmission of the spectral envelope and gains is similar in both systems. To determine how much of the complete bandwidth should be assigned to the description of

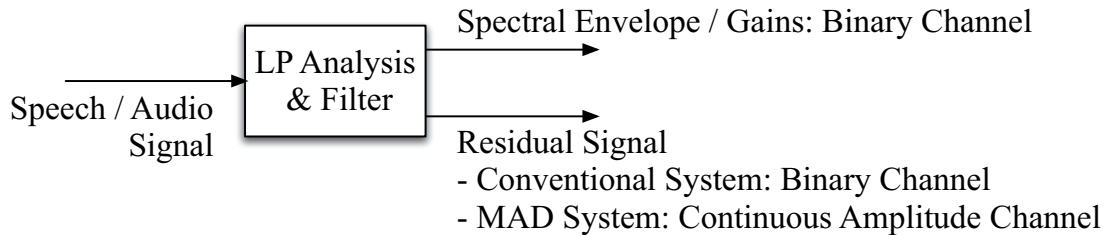


Figure 5.9: Simplified model of time-discrete linear predictive transmission of speech and audio signals.

the auto regressive (AR) model (that is, explicitly, the LP coefficients), one can follow the reasoning of Kleijn and Ozerov [12] who proved that this rate is determined by the model only and is independent from the gross rate available.

Ozerov and Kleijn show that the optimum rate allocation R for an AR model with quantized model parameters $\hat{\Theta}$ and blocklength n is

$$R(\hat{\Theta}) = h(\Theta) + \frac{d}{2} \log\left(\frac{n}{2} C_q\right) \quad (5.57)$$

with $d \leq |\hat{\Theta}|$ being the dimensionality in the log power spectral domain of the manifold which is described by the model parameters, $h(\Theta)$ the differential entropy of the optimum unquantized parameters and C_q a coefficient of quantization, which relates the volume V of the quantization cell to the mean distortion D [12].

$$C_q = \frac{V^{\frac{2}{n}}}{D} \quad (5.58)$$

Using 8 kHz samples speech from the TIMIT database [97] the dimensionality d of the manifold is found following [45] to be $d = 7.9$ and the differential entropy is found to be 8.5 bits, from which an optimum model rate of $R(\hat{\Theta}) = 19$ bits (scalar quantization), or $R(\hat{\Theta}) = 17.2$ bits (vector quantization), respectively, can be derived. This matches very well with the AMR speech codec [95] narrowband LPC quantization scheme (and MAD, see Section 3.3.1), where two sets of LP coefficients are jointly quantized with 38 bits using split matrix quantization.

In application to wideband speech, the same behaviour can again be studied, e.g., in the AMR wideband speech coding standard [96], where the rate assigned to the LP coefficients is constant in most coding modes (see Table 5.1).

For wideband speech transmission, our MAD transmission system uses exactly the same LP analysis (apart from noise shaping) and LPC quantization as does the AMR wideband codec.

Rate (kbit/s)	6.6	8.85	12.65	14.25	15.85	18.25	19.85	23.05
AR Model	36	46	46	46	46	46	46	46
Pitch Model	26	26	30	30	30	30	30	30
Excitation	48	80	144	176	208	256	288	352

Table 5.1: Bit rates of the AMR wideband speech codec (excluding gains and highband energy).

5.4.3 Analysis of the MAD Principle

With the processing of the spectral envelope being equal in conventional digital systems and the MAD transmission system, respectively, (see Section 3.3.1) the special characteristics of each transmission principle must be found in the transmission of the residual signal. As shown, e.g., in [21, 67, 78], the normalized residual of a speech or audio signal will in general have an approximately Gaussian pdf. An exemplary distribution can be found in Figure 5.10.

In whatever way the residual is quantized in the digital system, the binary representation will aim at an equal distribution of ones and zeros to minimize the description length of the code [62].

Thus, the general difference between conventional and MAD transmission can be explained with the different channel capacities when transmitting binary or Gaussian data over the AWGN channel as derived in section 5.1.4.

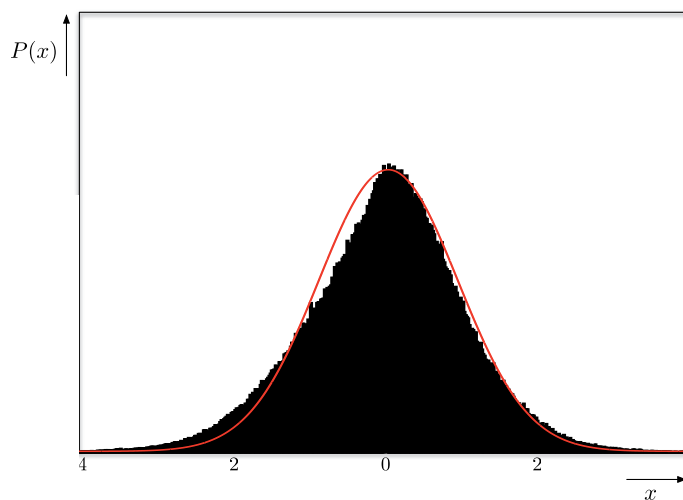


Figure 5.10: Histogram of a normalized residual r_n .

Free Capacity for the Residual Signal

The rate needed for the description of the AR model has been shown to be fixed regardless of the total rate available ([12], Section 5.4.2). Thus, deducting the part of the digital capacity needed for transmission of the model, it is possible to define the optimum working point (split between analogue and digital capacity) for MAD

transmission. Under the assumption of the model taking a total of, e.g.⁴, $0.6 \frac{\text{kbit/s}}{\text{Hz}}$, Figures 5.11 to 5.13 show the capacities left for the description of the residual for MAD (solid yellow line, $C_{ana,res}$), and digital transmission (dashed yellow line, $C_{digi,res}$), respectively⁵. (5.38) now leads to

$$C_{digi,res} < C_{ana,res} \quad (5.59)$$

for all channel conditions. This means that, from an information theoretic point of view, most of the information should be transmitted over the analogue channel. Including the results from Section 5.4.2 only the model (LP coefficients) and gains must be transmitted digitally, while all other information remains pseudo-analogue.

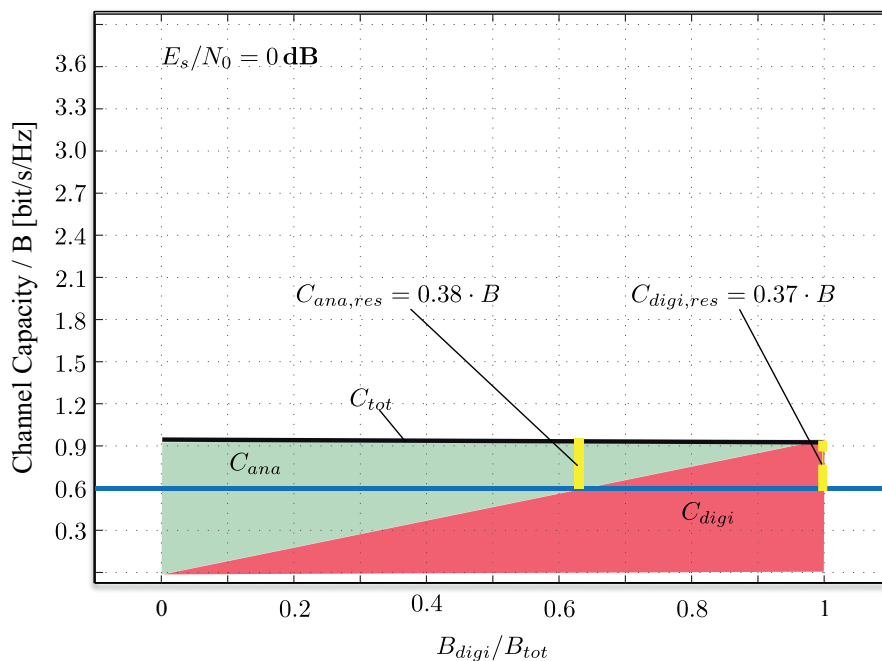


Figure 5.11: Comparison of the capacities $C_{digi,res}$ (dashed yellow line) and $C_{ana,res}$ (solid yellow line) left for transmission of the pseudo analogue residual, if the capacity C_{model} used by model description and gains covers $0.6 \frac{\text{bit/s}}{\text{Hz}}$. $E_s/N_0 = 0 \text{ dB}$. Note: $C_{digi,res} < C_{ana,res}$ for all SNR (5.59).

⁴This value is exemplary to support a clear understanding. In practice it depends on the rate of the digital information and on the bandwidth of the channel.

⁵It may be noted that an analogue transmission of speech or audio signals would not benefit from the complete capacity C_{ana} given on the left hand side of the figures, as neither speech nor audio signals fulfill the prerequisite of having a Gaussian distribution.

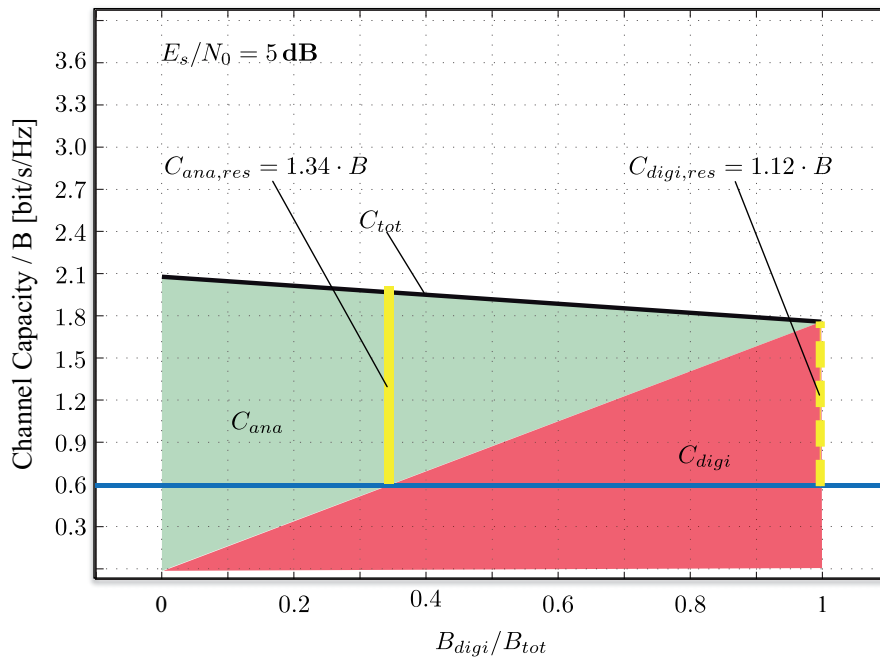


Figure 5.12: Comparison of the capacities $C_{digi,res}$ (dashed yellow line) and $C_{ana,res}$ (solid yellow line) left for transmission of the pseudo analogue residual, if the capacity C_{model} used by model description and gains covers $0.6 \frac{\text{bit/s}}{\text{Hz}}$. $E_s/N_0 = 5 \text{ dB}$.

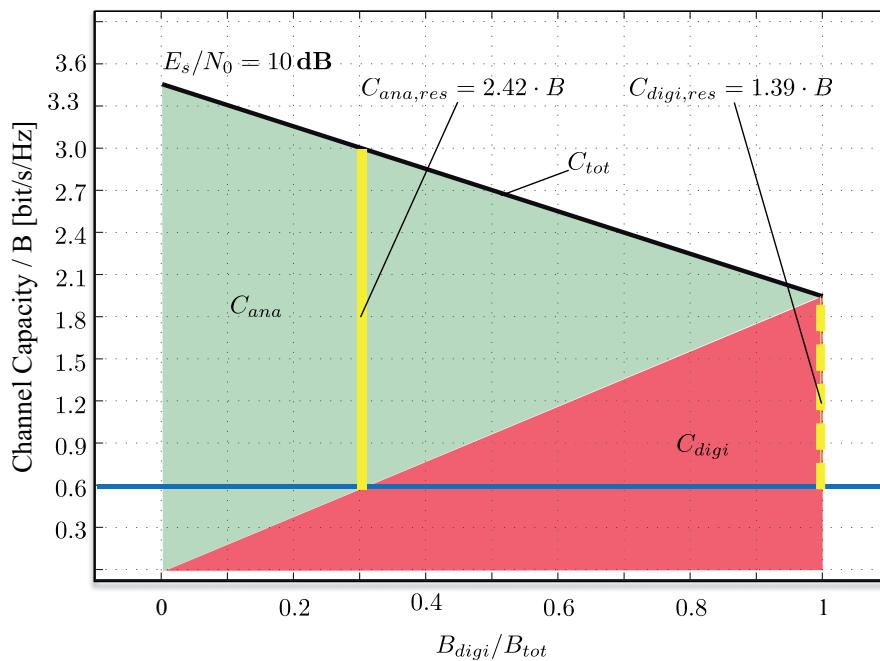


Figure 5.13: Comparison of the capacities $C_{digi,res}$ (dashed yellow line) and $C_{ana,res}$ (solid yellow line) left for transmission of the pseudo analogue residual, if the capacity C_{model} used by model description and gains covers $0.6 \frac{\text{bit/s}}{\text{Hz}}$. $E_s/N_0 = 10 \text{ dB}$.

However, it is not always possible to adjust the relation between B_{digi} and B_{ana} optimally. With multiplexing as suggested in Figure 3.4, e.g., a fixed relation is defined by the number of bits versus the number of samples within a given time interval. With fixed $\frac{B_{digi}}{B_{tot}}$, the capacity C_{model} needed for transmission of the model may exceed the available capacity C_{digi} although the total capacity C_{tot} is still higher than C_{model} , or the capacity available for transmission of the pseudo analogue residual $C_{ana,res}$ is lower than it could be for optimum bandwidth distribution. This effect is visualized in Figures 5.14 to 5.16.

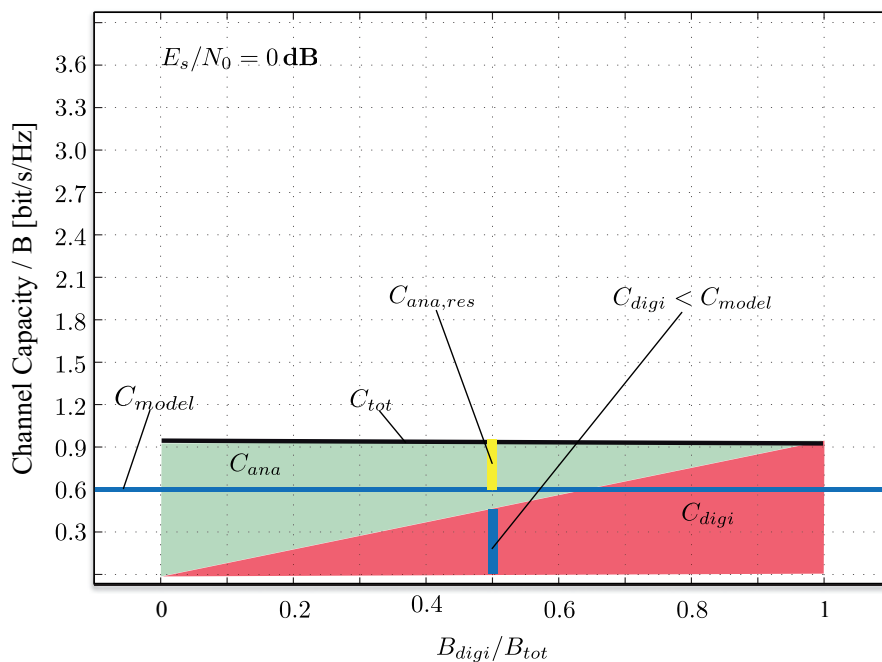


Figure 5.14: Comparison of the capacities C_{model} used by model description (blue line) and $C_{ana,res}$ (solid yellow line) left for transmission of the pseudo analogue residual with $C_{model} = 0.6 \frac{\text{bit/s}}{\text{Hz}}$. $E_s/N_0 = 0 \text{ dB}$.

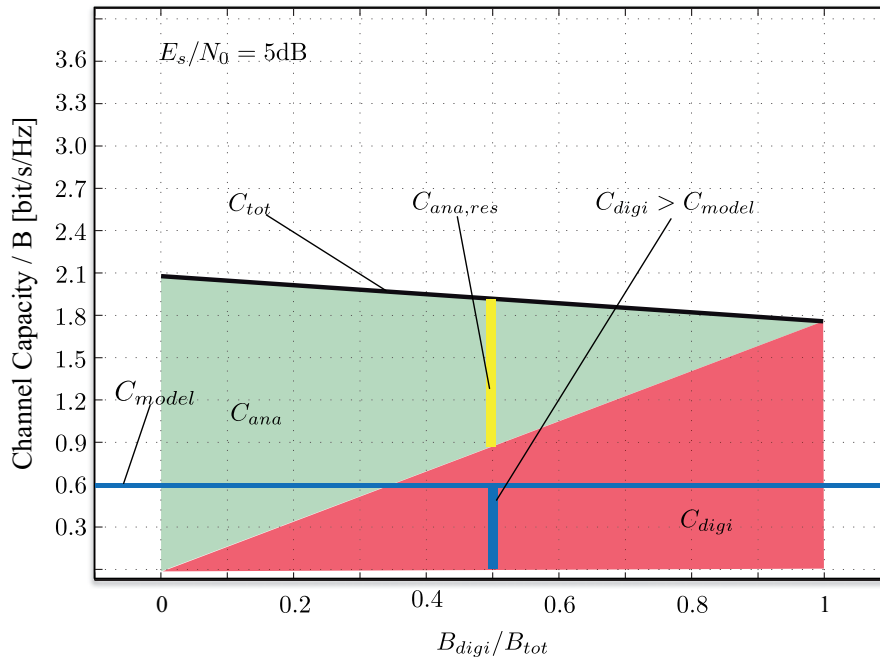


Figure 5.15: Comparison of the capacities C_{model} used by model description (blue line) and $C_{ana,res}$ (solid yellow line) left for transmission of the pseudo analogue residual with $C_{model} = 0.6 \frac{\text{bit/s}}{\text{Hz}}$. $E_s/N_0 = 5 \text{ dB}$.

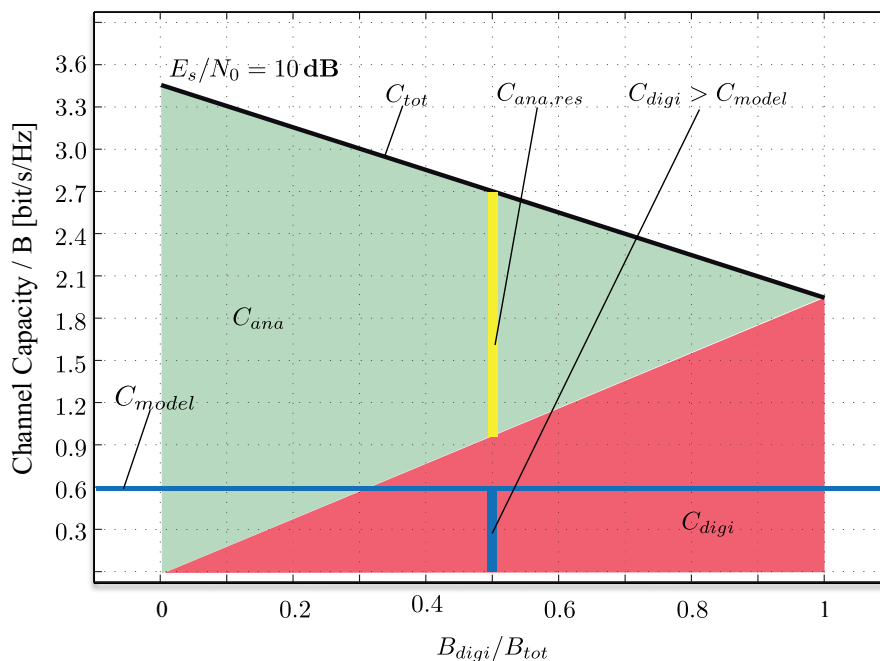


Figure 5.16: Comparison of the capacities C_{model} used by model description (blue line) and $C_{ana,res}$ (solid yellow line) left for transmission of the pseudo analogue residual with $C_{model} = 0.6 \frac{\text{bit/s}}{\text{Hz}}$. $E_s/N_0 = 10 \text{ dB}$.

SDR of MAD Transmission

In Section 5.3 the Optimum Performance Theoretically Attainable (OPTA) for the transmission of the output of a memoryless Gaussian source has been introduced. To evaluate the behaviour of MAD transmission as introduced in Section 3.3 the Signal-to-Distortion rate (SDR) has been determined for two different sources:

- a) Natural speech, 1.3 million samples from male and female speakers, German, English, Russian and Spanish language.
- b) An input signal generated by an auto-regressive (AR) process based on white noise excitation with constant power $\sigma_d^2 = 1$ and an AR filter $H_s(z)$ with fixed filter coefficients that resemble the LP coefficients of a short sequence of voiced speech.

The pseudo analogue transmission path is detailed in Figure 5.17. The assumed pulse shaping filter and the matched filter are both ideal low pass filters with amplitude 1 and cut-off frequency $f_s/2$. For evaluation purposes the following measures are used:

$$CSNR = \frac{E \{x(k)^2\}}{E \{(x(k) - \tilde{x}(k))^2\}} \quad (5.60)$$

$$SDR_r = \frac{E \{r(k)^2\}}{E \{(r(k) - \tilde{r}(k))^2\}} \quad (5.61)$$

$$SDR_s = \frac{E \{s(k)^2\}}{E \{(s(k) - \tilde{s}(k))^2\}} \quad (5.62)$$

Simulations have been carried out for a range of -10 dB to 30 dB E_s/N_0 .

Figure 5.18 shows the behaviour of MAD transmission of natural wideband speech (BPSK and PAM) in relation to OPTA and PAM as introduced in Section 5.3. With the residual being transmitted over a PAM link, it is reasonable that for larger channel SNR the incline equals that of the optimum PAM system. The gain in output SNR that allows the simulated SDR to be higher than OPTA is related to the correlation in the speech signal and to the noise-shaping. This effect will be examined in detail below.

It should be noted that, while the SDR performance increases linearly for larger CSNR, the subjective perception approaches that of the original signal at roughly $20 - 25$ dB CSNR (compare Figure 4.27), so that even with further improvement in SDR, the difference in the received signal is not clearly noticeable.

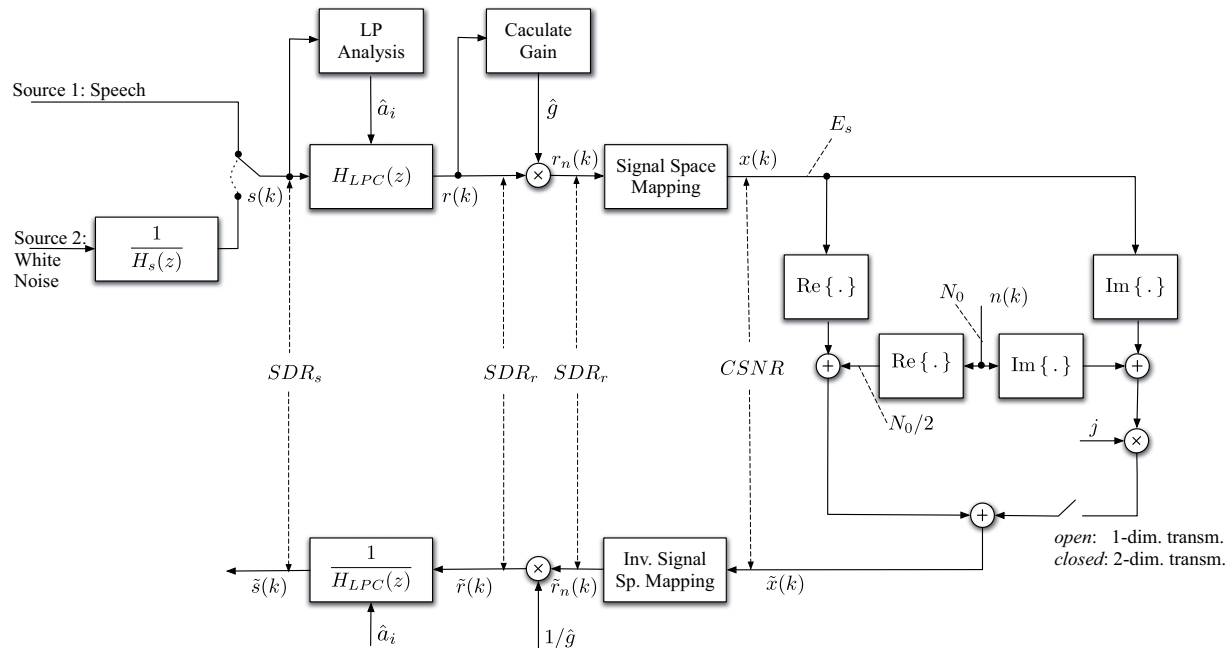


Figure 5.17: Model of the MAD transmission system, pseudo analogue path, showing $CSNR$, SDR_r , and SDR_s .

While any purely digital transmission system will not show any more improvement in terms of SDR if all bits are decoded correctly, MAD transmission will always benefit from improved channel conditions in terms of output SDR. Depending on the application, however, the improvement in SDR in good conditions may not result in a proportional improvement in perceptual quality, as the latter will saturate sooner or later. This effect can well be studied by comparison of Figures 4.26 and 4.27.

Noise Shaping in MAD

As shown in Section 3.3.1, the amount of noise shaping can be varied in MAD. Equation (3.4) gives the traditional approach of noise shaping, in which the modified prediction filter

$$H(z) = \frac{1 - A(z)}{1 - A(z/\gamma)} \quad (5.63)$$

is controlled by a factor of $\gamma = 0$ (full prediction) to $\gamma = 1$ (no prediction). Figure 3.8 illustrates the effect of this variation.

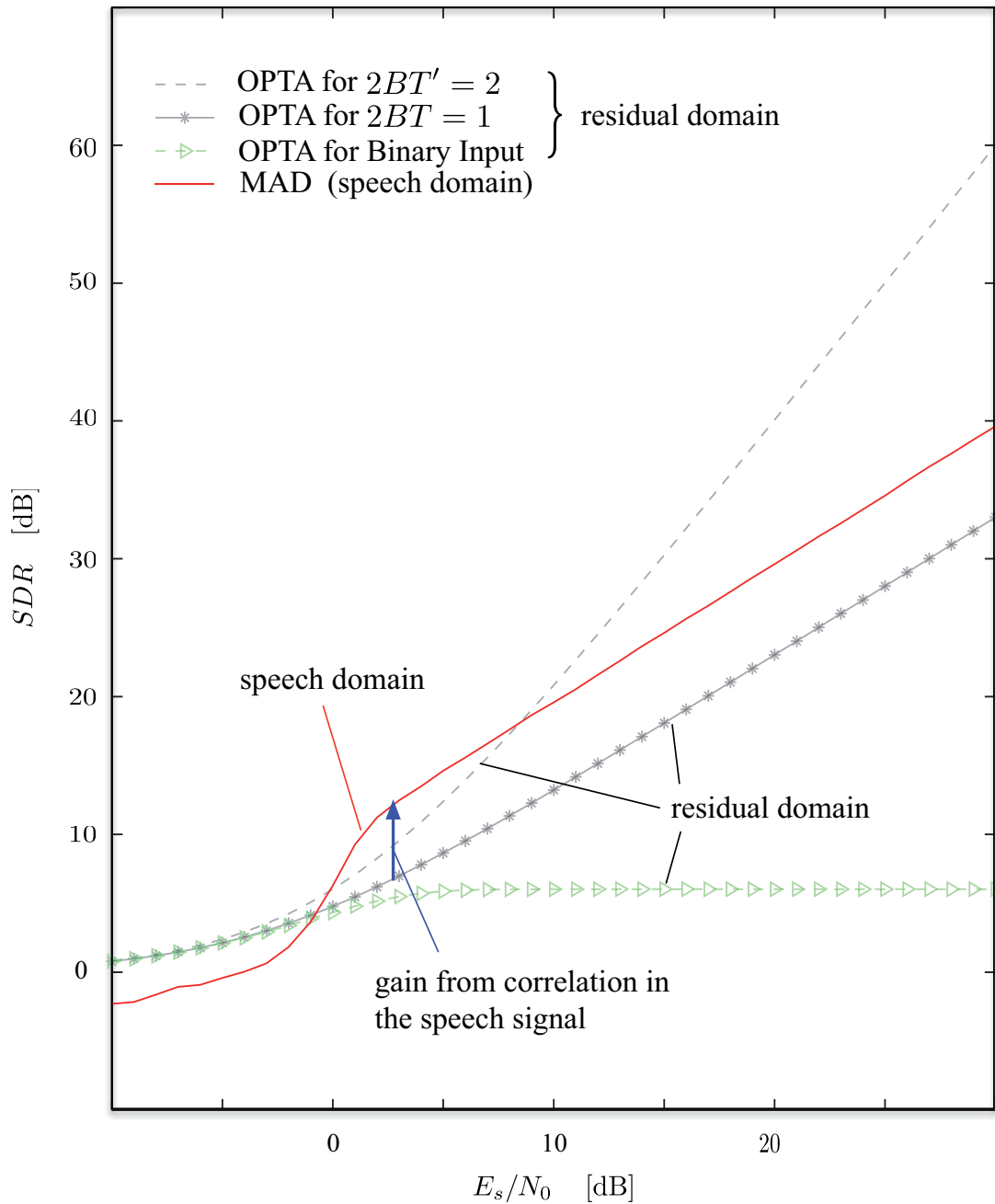


Figure 5.18: Performance of the MAD transmission system. **Note that the SDR in the Speech domain has an offset compared to the SDR in the residual domain.** This offset (which allows the SDR of MAD in the speech domain to appear above the OPTA of the residual domain) is related to the prediction gain and quantified in the following Section.

For a detailed theoretical analysis, however, it is more convenient to examine a related effect, which has been studied in the context of open loop and closed loop quantization by H. Krüger in [13]. He modified the calculation of the LP coefficients in the frequency domain as shown in Figure 5.19 to achieve noise shaping dependent

on a factor α with $0 \leq \alpha \leq 1$. When $\alpha = 0.0$ this yields no prediction (and thus simple PCM coding, equivalent to $\gamma = 1$) while when $\alpha = 1.0$ this yields full prediction. This modification is called *partial decorrelation* [13].

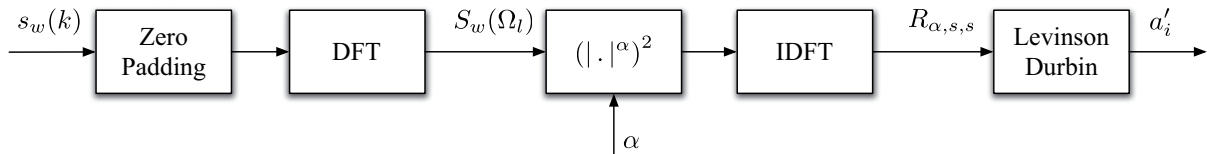


Figure 5.19: Calculation of the LP coefficients as in [13].

An important difference between the resulting LP filter structures is that the coefficients calculated with the partial decorrelation scheme apply to the Finite Impulse Response (FIR) part of the analysis filter $\tilde{H}(z) = 1 - A(z)$ while using γ other than $\gamma = 1$ results in an Infinite Impulse Response (IIR) filter structure of $H(z)$. Perceptually, traditional noise shaping and partial decorrelation yield comparable quality if the respective parameters are chosen adequately. This can be seen in Figure 5.20.

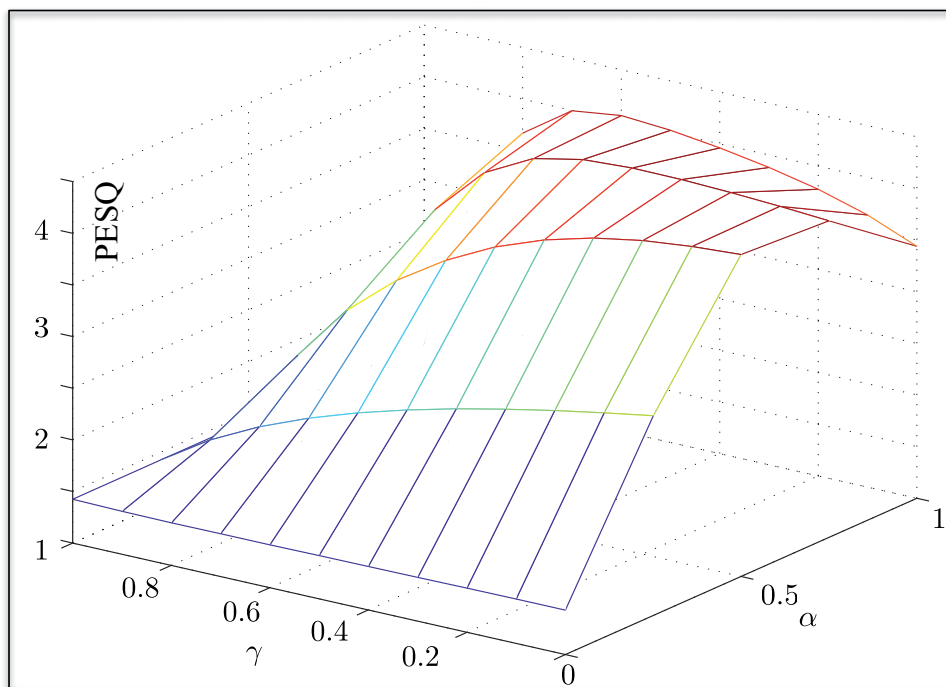


Figure 5.20: Perceptual quality of MAD using a combination of traditional noise shaping (variation of γ) and partial decorrelation (variation of α).

What makes the modifications interesting is that they allow for calculation of the impact of the prediction on the output SNR (SDR). Although partial decorrelation

is based on added quantization noise, it applies accordingly to the AWGN channel noise regarded for MAD transmission, as shown below.

The model regarded by Krüger is given in Figure 5.21. White noise excitation $D_s(z)$ with constant power σ_d^2 passes through an AR filter $H_s(z)$ which may be expressed as a cascade of two separate filters $H_{s,1}(z)$ and $H_{s,2}(z)$

$$\frac{1}{H_s(z)} = \frac{1}{H_{s,1}(z)} \cdot \frac{1}{H_{s,2}(z)} \quad (5.64)$$

to form the input signal

$$S(z) = \frac{1}{H_s(z)} \cdot D_s(z). \quad (5.65)$$

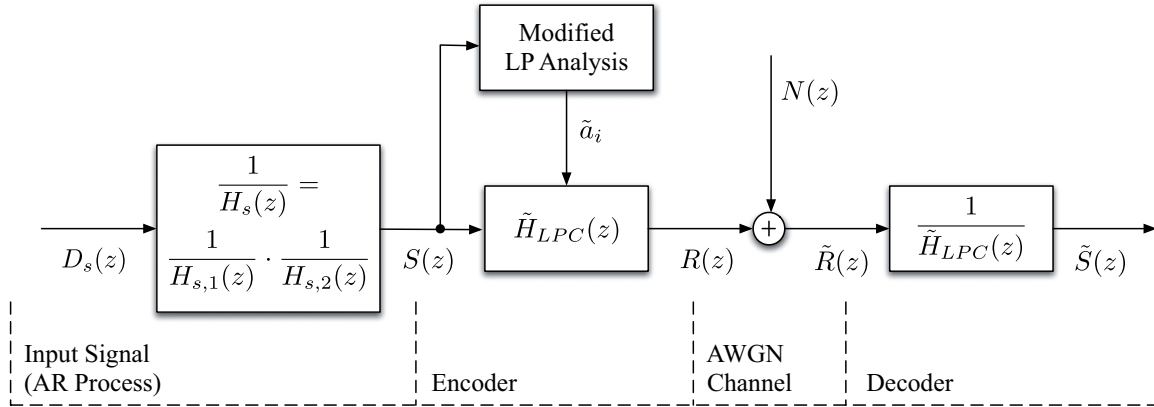


Figure 5.21: Model of the Partial Decorrelation Scheme

From the input signal $S(z)$ the modified LP coefficients \tilde{a}_i are calculated and the modified LP analysis filter

$$\tilde{H}_{LPC}(z) = 1 - \tilde{A}(z) \quad (5.66)$$

is applied. The residual signal $R(z)$ is then transmitted over the AWGN channel modeled by adding white noise $N(z)$ with power σ_n^2 . The channel SNR

$$CSNR = \frac{E\{r^2(k)\}}{E\{(r(k) - \tilde{r}(k))^2(k)\}} = \frac{E\{r^2(k)\}}{E\{n^2(k)\}} \quad (5.67)$$

is assumed to be constant.

On the decoder side the noisy residual $\tilde{R}(z)$ is filtered with the LP synthesis filter

$$\frac{1}{\tilde{H}_{LPC}(z)} \quad (5.68)$$

to get the received signal $\tilde{S}(z)$.

The impact of the parameter α can best be determined in the spectral domain with $z = e^{j\Omega}$. Using Parseval's theorem the signal energy of $S(z)$ can be derived:

$$E\{s^2(k)\} = \frac{\sigma_d^2}{2\pi} \cdot \int_{-\pi}^{\pi} \left| \frac{1}{H_{s,1}(\Omega)} \right|^2 \cdot \left| \frac{1}{H_{s,2}(\Omega)} \right|^2 d\Omega. \quad (5.69)$$

The modified LP analysis filter only partially decorrelates the input signal. Without loss of generality it can be assumed that only the correlation introduced by the second stage of the cascaded filter, $\frac{1}{H_{s,2}(z)}$ is decorrelated by linear prediction, thus

$$\frac{1}{\tilde{H}_{LPC}(z)} = \frac{1}{H_{s,2}(z)}. \quad (5.70)$$

From this the energy of the residual signal $R(z)$ can be determined as

$$E\{r^2(k)\} = \frac{\sigma_d^2}{2\pi} \cdot \int_{-\pi}^{\pi} \left| \frac{1}{H_{s,1}(\Omega)} \right|^2 d\Omega \quad (5.71)$$

and the channel SNR is

$$CSNR = \frac{1}{\sigma_n^2} \cdot \frac{\sigma_d^2}{2\pi} \cdot \int_{-\pi}^{\pi} \left| \frac{1}{H_{s,1}(\Omega)} \right|^2 d\Omega. \quad (5.72)$$

After LP synthesis filtering, the energy of the distortion is

$$E\{(s(k) - \tilde{s}(k))^2\} = \frac{\sigma_n^2}{2\pi} \cdot \int_{-\pi}^{\pi} \left| \frac{1}{H_{s,2}(\Omega)} \right|^2 d\Omega. \quad (5.73)$$

With this, the SDR in the speech domain becomes

$$SDR_s = \frac{\frac{\sigma_d^2}{2\pi} \cdot \int_{-\pi}^{\pi} \left| \frac{1}{H_{s,1}(\Omega)} \right|^2 \cdot \left| \frac{1}{H_{s,2}(\Omega)} \right|^2 d\Omega}{\frac{\sigma_n^2}{2\pi} \cdot \int_{-\pi}^{\pi} \left| \frac{1}{H_{s,2}(\Omega)} \right|^2 d\Omega}. \quad (5.74)$$

The SDR depends on the one hand on the CSNR and on the other hand on the transformation of the channel noise due to the LP synthesis filter. A gain G_{SNR} can be defined that relates the CSNR and the SDR with linear prediction:

$$G_{SNR} = \frac{SDR_s}{CSNR} \quad (5.75)$$

$$= \frac{\int_{-\pi}^{\pi} \left| \frac{1}{H_{s,1}(\Omega)} \right|^2 \cdot \left| \frac{1}{H_{s,2}(\Omega)} \right|^2 \cdot \frac{d\Omega}{2\pi}}{\int_{-\pi}^{\pi} \left| \frac{1}{H_{s,2}(\Omega)} \right|^2 \frac{d\Omega}{2\pi} \cdot \int_{-\pi}^{\pi} \left| \frac{1}{H_{s,1}(\Omega)} \right|^2 \frac{d\Omega}{2\pi}}. \quad (5.76)$$

For an AR process, the transfer function of a conventional linear prediction filter $1/H_{LPC}(\Omega)$ approximates the spectrum $1/H_s(\Omega)$ of the input signal. Spectral flattening of $S(z)$ with parameter α as in Figure 5.19 thus also flattens the approximation of the signal shape, which is the transfer function of the linear prediction filter. With this

$$\frac{1}{H_{s,2}(\Omega)} = \frac{1}{\tilde{H}_{LPC}(\Omega)} = \frac{1}{(H_s(\Omega))^\alpha} \quad (5.77)$$

and

$$\frac{1}{H_{s,1}(\Omega)} = \frac{H_s(\Omega)}{\tilde{H}_{s,2}(\Omega)} = \frac{1}{(H_s(\Omega))^{(1-\alpha)}}. \quad (5.78)$$

Thus, the signal distortion is spectrally shaped according to

$$|S(\Omega) - \tilde{S}(\Omega)| \sim \left| \frac{1}{(H_s(\Omega))^\alpha} \right|. \quad (5.79)$$

Figure 5.22 shows G_{SNR} over α . Due to the symmetric property of G_{SNR} (Equation (5.76)) the maximum gain is achieved with $\alpha = 0.5$. It can be calculated as

$$G_{SNR} = \frac{\int_{-\pi}^{\pi} \left| \frac{1}{H(z)} \right|^2 \frac{d\Omega}{2\pi}}{\int_{-\pi}^{\pi} \left| \frac{1}{H(z)} \right|^{2\alpha} \frac{d\Omega}{2\pi} \int_{-\pi}^{\pi} \left| \frac{1}{H(z)} \right|^{2 \cdot (1-\alpha)} \frac{d\Omega}{2\pi}}. \quad (5.80)$$

Similar studies of what is effectively an optimized adaptive pre-emphasis and de-emphasis filter structure can be found in [29, 31, 36, 72]. Mostly this technique is referred to as *half-whitening*.

Figure 5.23 compares the effect of noise shaping with α or γ . The marginals show the effects seen in Figure 3.8 for traditional noise shaping (variation of γ) and in Figure 5.19 for partial decorrelation (variation of α). Also it is clearly visible that for PCM coding ($\alpha = 0$ or $\gamma = 1$) no gain in terms of SDR can be achieved, which is well known from open loop quantization.

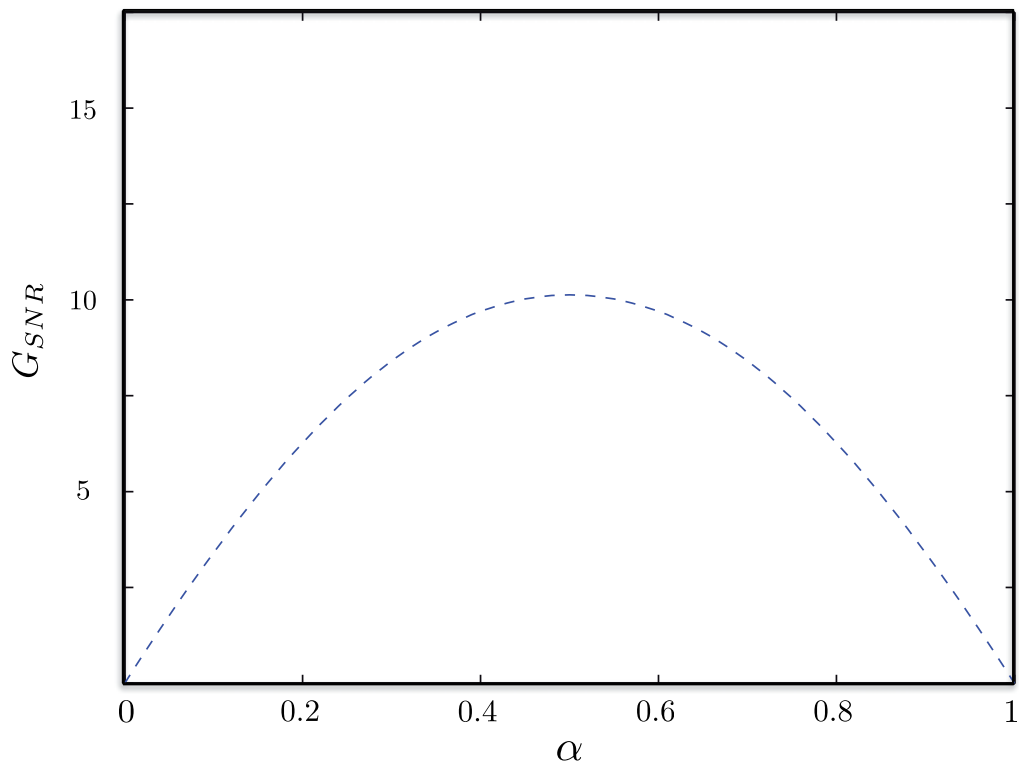


Figure 5.22: G_{SNR} over α for an exemplary AR filter $H_s(z)$ as defined in Appendix C.

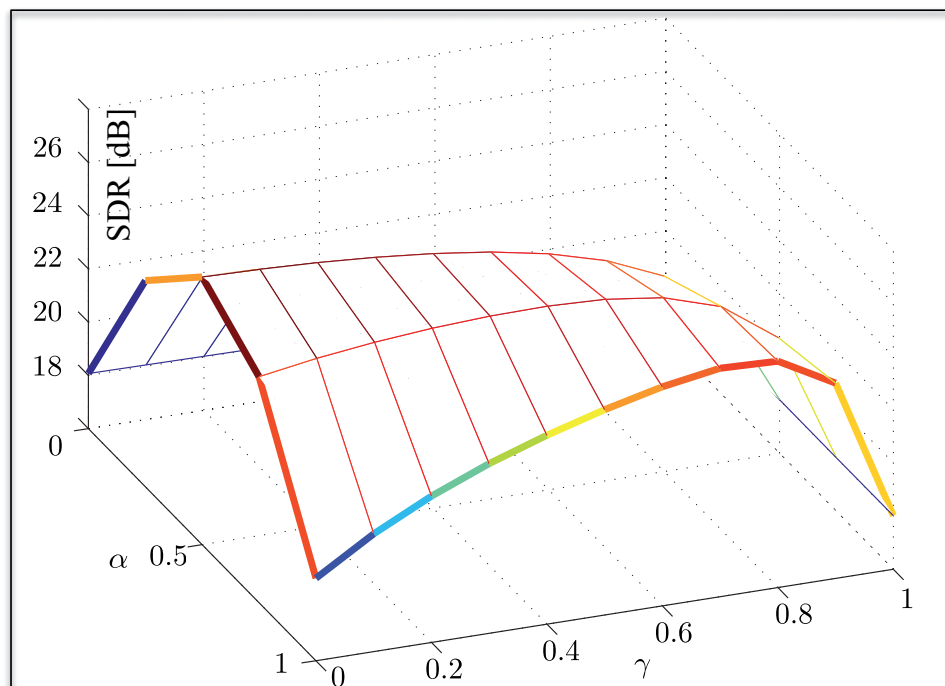


Figure 5.23: Comparison of traditional noise shaping and partial decorrelation with respect to the SDR at 18 dB channel SNR.

Figure 5.24 shows the difference in SDR using partial decorrelation vs. using traditional noise shaping. While perceptually indistinguishable, the partial decorrelation scheme is optimized in terms of SDR and thus tops the traditional scheme. The strong degradation around 0 dB E_s/N_0 is caused by bit errors which occur in the digital path (description of the model, LP coefficients and gains), compare Figure 5.14.

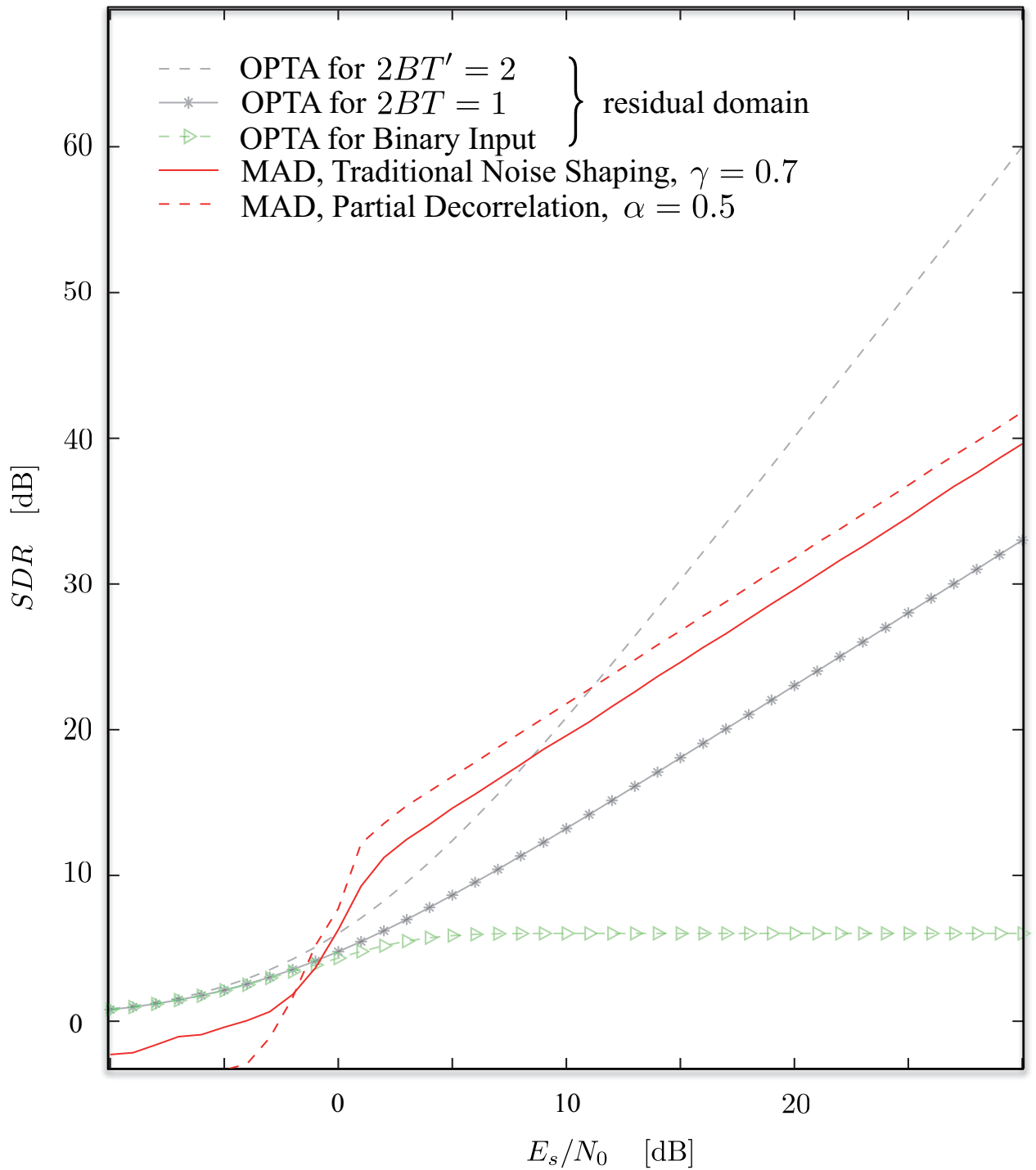


Figure 5.24: SDR performance of the MAD transmission system with traditional noise shaping (variation of γ) vs. MAD using partial decorrelation (variation of α).

Using partial decorrelation with $\alpha = 0.5$ allows to calculate the maximum of G_{SNR} . With the inverse of the spectral flatness SF^{-1} of the magnitude spectrum of the LP synthesis filter $\frac{1}{H(z)}$ being the limit for the prediction gain G_p

$$G_p|_{N \rightarrow \infty} = SF^{-1} \left(\left. \frac{1}{H(z)} \right|_{z=e^{j\Omega}} \right) \quad (5.81)$$

$$= \frac{\int_{-\pi}^{\pi} \left| \frac{1}{H(\Omega)} \right|^2 \frac{d\Omega}{2\pi}}{e^{\left(\int_{-\pi}^{\pi} \ln \left| \frac{1}{H(\Omega)} \right|^2 \frac{d\Omega}{2\pi} \right)}}, \quad (5.82)$$

and $H(z)$ having zero mean property [74]:

$$\int_{-\pi}^{\pi} \ln \left(\left| \frac{1}{H(\Omega)} \right|^2 \right) \frac{d\Omega}{2\pi} = 0.0, \quad (5.83)$$

Krüger has shown in [13] that G_{SNR} can be expressed as

$$\begin{aligned} G_{SNR}(\alpha = 0.5) &= 10 \log_{10} \left(\frac{SF^{-1}(\alpha = 1.0)}{SF^{-1}(\alpha = 0.5) \cdot SF^{-1}(\alpha = 0.5)} \right) \\ &= 10 \log_{10}(SF^{-1}(\alpha = 1.0)) - 2 \cdot 10 \log_{10}(SF^{-1}(\alpha = 0.5)). \end{aligned} \quad (5.84)$$

The spectral flatness is detailed in Appendix A.

Bounds for MAD Transmission

For the AR input signal defined in Section 5.4.3 the spectral flatness is:

$$10 \log_{10}(SF_{AR}^{-1}(\alpha = 1.0)) = 25.44 \text{ dB} \quad (5.85)$$

$$10 \log_{10}(SF_{AR}^{-1}(\alpha = 0.5)) = 7.65 \text{ dB} \quad (5.86)$$

and thus

$$G_{SNR}(\alpha = 0.5) = 25.44 \text{ dB} - 2 \cdot 7.65 \text{ dB} = 10.14 \text{ dB}. \quad (5.87)$$

Knowing G_{SNR} , the Signal-to-Distortion ratios SDR_s and SDR_r , and OPTA can be related as follows:

$$SDR_r < OPTA \quad (5.88)$$

$$SDR_s = SDR_r + G_{SNR} \quad (5.89)$$

and thus we find a bound OPT_{MAD} for MAD transmission

$$SDR_s < OPTA + G_{SNR} \stackrel{!}{=} OPT_{MAD}. \quad (5.90)$$

As pointed out in Sections 5.4.2 and 5.4.3, the digitally transmitted description of the model is vital for all transmission scenarios (meaning it must be there in complete to have any useful output at all), while the way of transmitting the residual signal determines the quality of the received signal. The waterfall effect of incomplete transmission of the model can be studied in Figure 5.24 at about 0 dB E_s/N_0 . To focus on the effects caused by the pseudo analogue transmission of the residual signal, the digital part of the transmission is assumed to be error-free in case of using the AR input signal.

Figure 5.25 shows the performance of MAD transmission with the AR input signal in relation to OPTA and OPT_{MAD} . OPT_{MAD} is not completely reached, as the LP coefficients have been quantized for transmission.

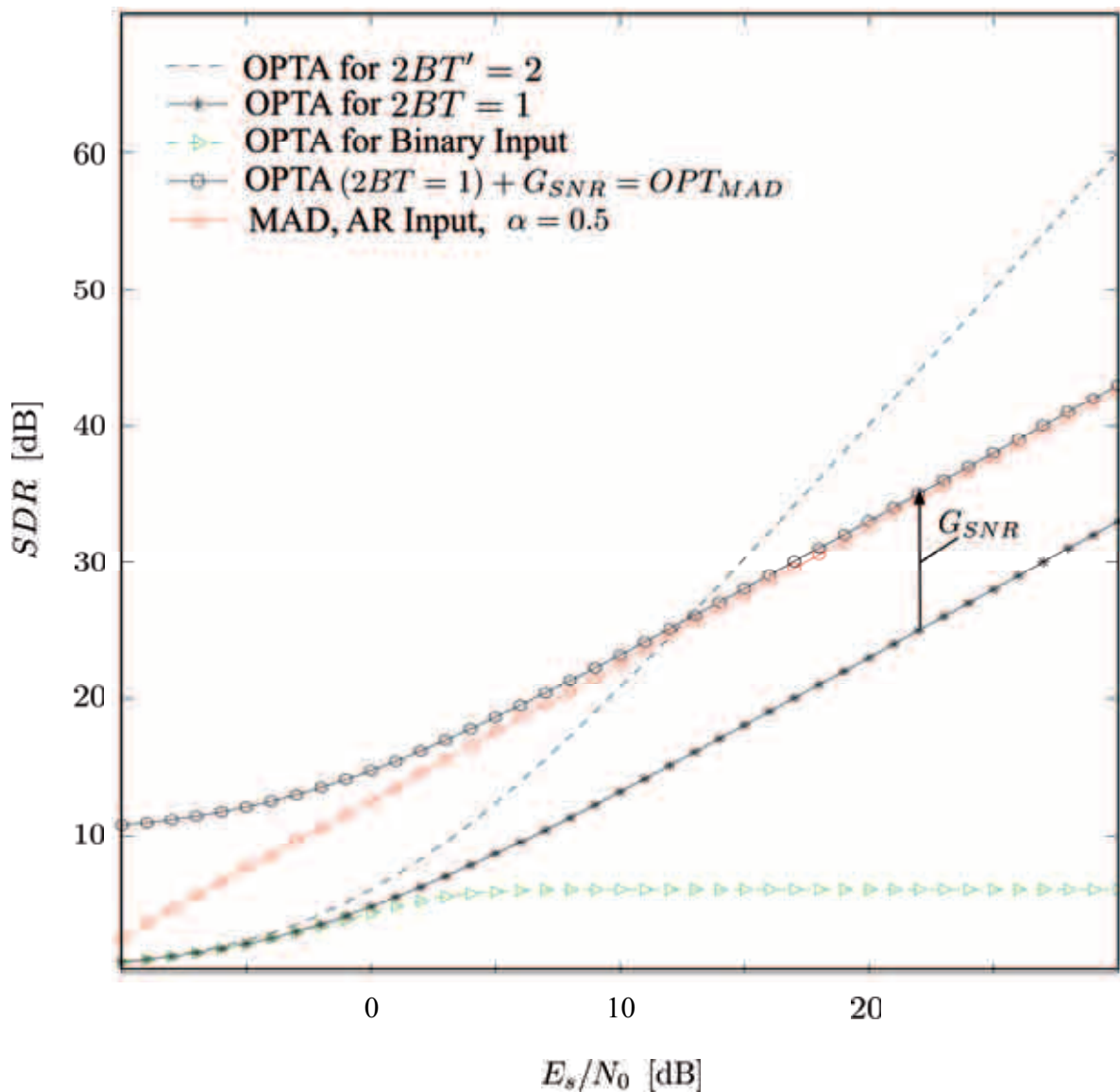


Figure 5.25: A bound for MAD transmission in case of error free transmission of the digital information.

AMR mode (kbit/s)	4.75	5.15	5.9	6.7	7.4
Encoder total	11.63	9.65	11.35	14.03	13.03
LP	1.40	1.40	1.90	1.90	1.90
AMR mode (kbit/s)	7.95	10.2	12.2	TWC	MAD-NB
Encoder total	14.18	13.66	14.05	14.18	3.25
LP	1.90	1.90	2.75	2.75	2.75

Table 5.2: Complexity of AMR and MAD narrowband, wMOPS

AMR-WB mode (kbit/s)	6.6	8.85	12.65	14.25	15.85	18.25
Encoder total	20.46	23.59	26.91	29.24	29.41	30.22
LP	2.10	2.10	2.10	2.10	2.10	2.10
AMR-WB mode (kbit/s)	19.85	23.05	23.85	TWC	MAD-WB	
Encoder total	31.14	30.84	29.07	31.14	3.1	
LP	2.10	2.10	2.10	2.10	2.10	

Table 5.3: Complexity of AMR and MAD wideband, wMOPS

Complexity Aspects

The AMR and AMR-WB speech codecs have been evaluated regarding complexity in [102] and [103], respectively, using the methodology agreed for the standardization of the GSM speech codecs. The actual values (wMOPS, weighted Million Operations per Second) will eventually depend on the final DSP⁶ implementation. The theoretical worst case complexity (TWC) is given in addition to the complexity of each mode of the AMR codecs.

To approximate the complexity of the MAD transmission scheme, the complexity of LP analysis, LP quantization, and LP filtering with the modules inherited from the AMR codecs has been studied. For normalization of the residual, approximately 0.5 wMOPS (narrowband) or 1.0 wMOPS (wideband) are required. In total, the complexity of the MAD system is significantly lower than the one of AMR. Tables 5.2 and 5.3 show the resulting values. In narrowband mode, MAD has about 1/4 of the complexity of AMR. MAD wideband reduces the complexity by a factor of 6.6 to 10 compared to AMR wideband.

Variable Energy Distribution Between Model and Residual

So far, it was implicitly assumed that the digital information is transmitted with unit amplitude/power and the pseudo analogue samples are normalized to unit average power within each 5 ms frame as well. However, within the MAD transmission system, there is no strict demand for this power distribution. The total power available for one speech frame must be one, yet the distribution of power between digital and pseudo analogue is not predetermined. While surely using all power for the digital model omitting the transmission of the residual is just as senseless as omitting

⁶Digital Signal Processor

the digital transmission for the benefit of the pseudo analogue samples, other unequal power distributions can improve the transmission. Allowing more power for the digital part of the transmission will shift the waterfall region to lower SNR while reducing the output SDR for better channels and vice-versa. As a rule-of-thumb, the digital transmission should get just enough power for the channel coder to work effectively, while all other power should be used for the pseudo analogue samples. This can be motivated by the SDR characteristic of digital transmission systems, which flattens out as soon as all bits are decoded correctly. Because raising the CSNR further yields no further improvement in SDR, in case of lowered channel noise, the transmission power used for the digital part can be reduced to remain in the same CSNR working point while at the same time the saved power can boost CSNR for the pseudo analogue transmission.

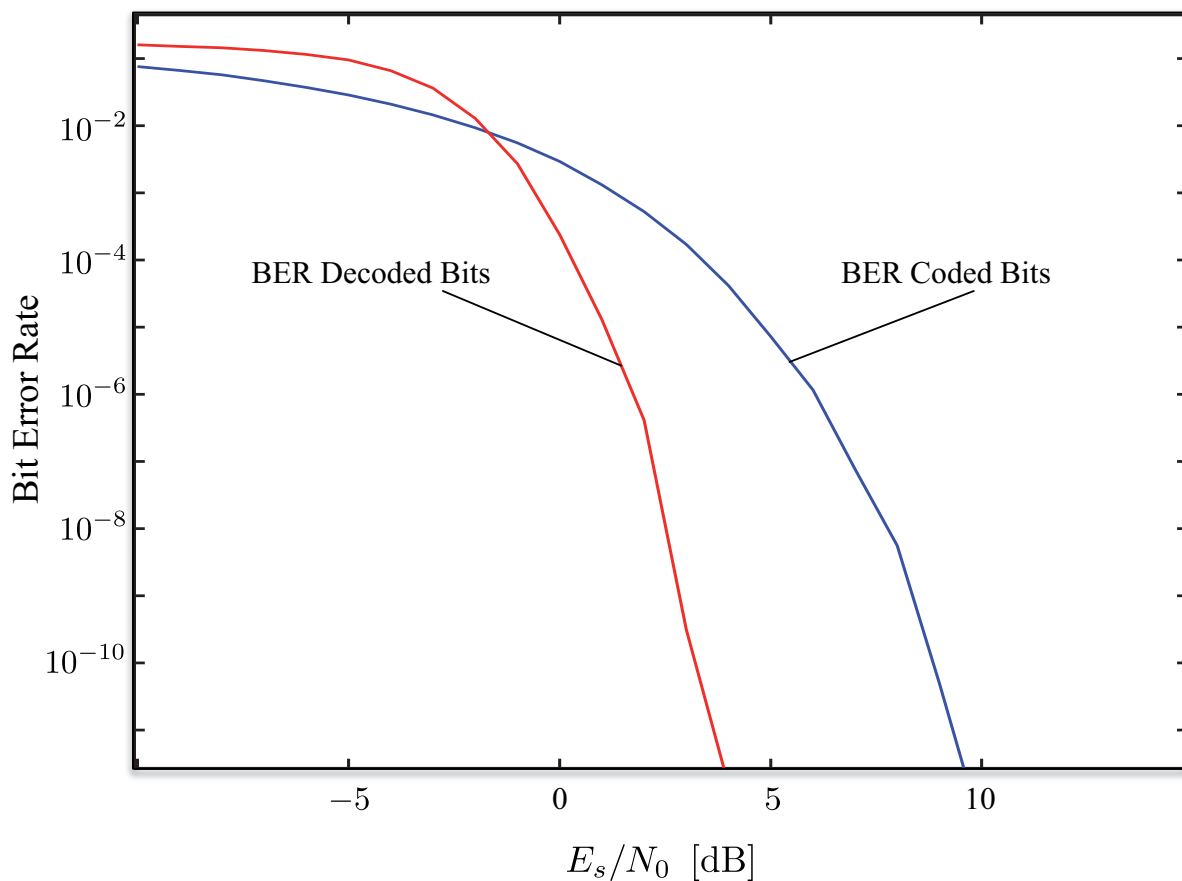


Figure 5.26: Bit Error Rate (BER) in the digital path of the MAD transmission scheme before and after channel decoding.

Figure 5.26 shows the bit error rates that occur in the digital path of MAD transmission depending on the channel E_s/N_0 . From about 3 dB E_s/N_0 the bit error rate (BER) after channel decoding is

$$BER_{dec}(E_s/N_0 \geq 3 \text{ dB}) \leq 10^{-8} \quad (5.91)$$

If the channel conditions are known, the power $E_{s,digi}$ used for transmitting the digital information can be adjusted in such a way that

$$E_{s,digi}/N_0 = 3 \text{ dB} = \text{const.} \quad (5.92)$$

This allows to rise the power $E_{s,ana}$ used for transmitting the pseudo analogue information to keep the average transmit power E_s constant. In the case of wideband MAD, we find

$$E_{s,ana} = E_s + 10 \cdot \log_{10} \left(\frac{B_{tot} - B_{digi} \cdot 10^{\frac{3 \text{ dB} - (E_s/N_0)}{10}}}{B_{ana}} \right) \quad (5.93)$$

$$= E_s + \Delta_{E_s,ana}. \quad (5.94)$$

With $B_{tot} = 23000$ kHz, $B_{digi} = 7000$ kHz and $B_{ana} = 16000$ kHz, $\Delta_{E_s,ana}$ is limited by

$$\Delta_{E_s,ana} \leq 10 \cdot \log_{10} \left(\frac{B_{tot}}{B_{ana}} \right) = 1.5761 \text{ dB}. \quad (5.95)$$

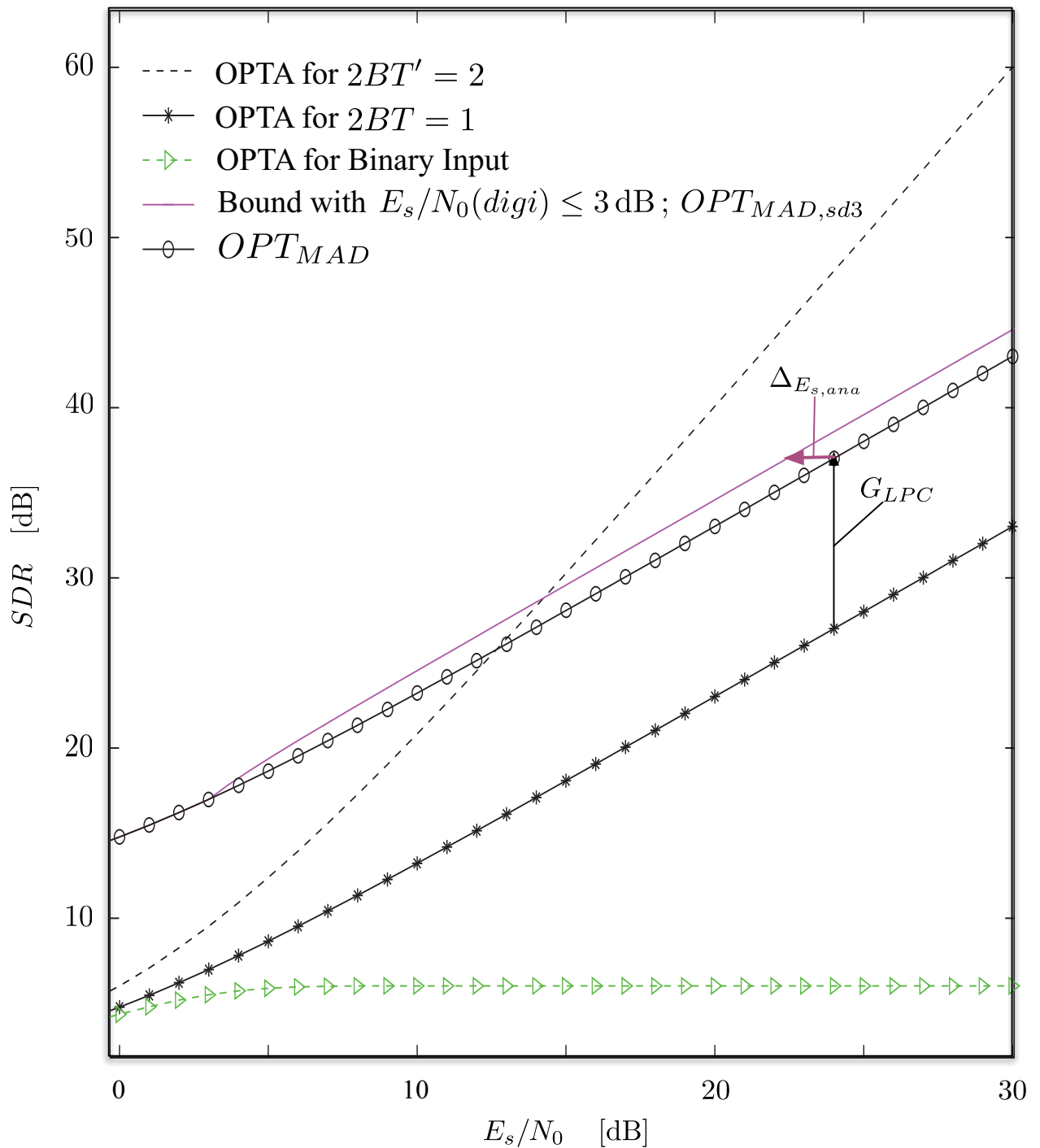


Figure 5.27: Performance bound for MAD transmission with unequal distribution of the total power. $E_{s,digi}/N_0 = 3$ dB for $E_s/N_0 \geq 3$ dB.

The possible performance with $E_{s,digi}/N_0 = 3$ dB is shown in Figure 5.27 in terms of SDR and in Figure 5.29 in terms of speech quality. Figure 5.28 shows the output SDR of MAD transmission using the AR input signal defined in Section 5.4.3.

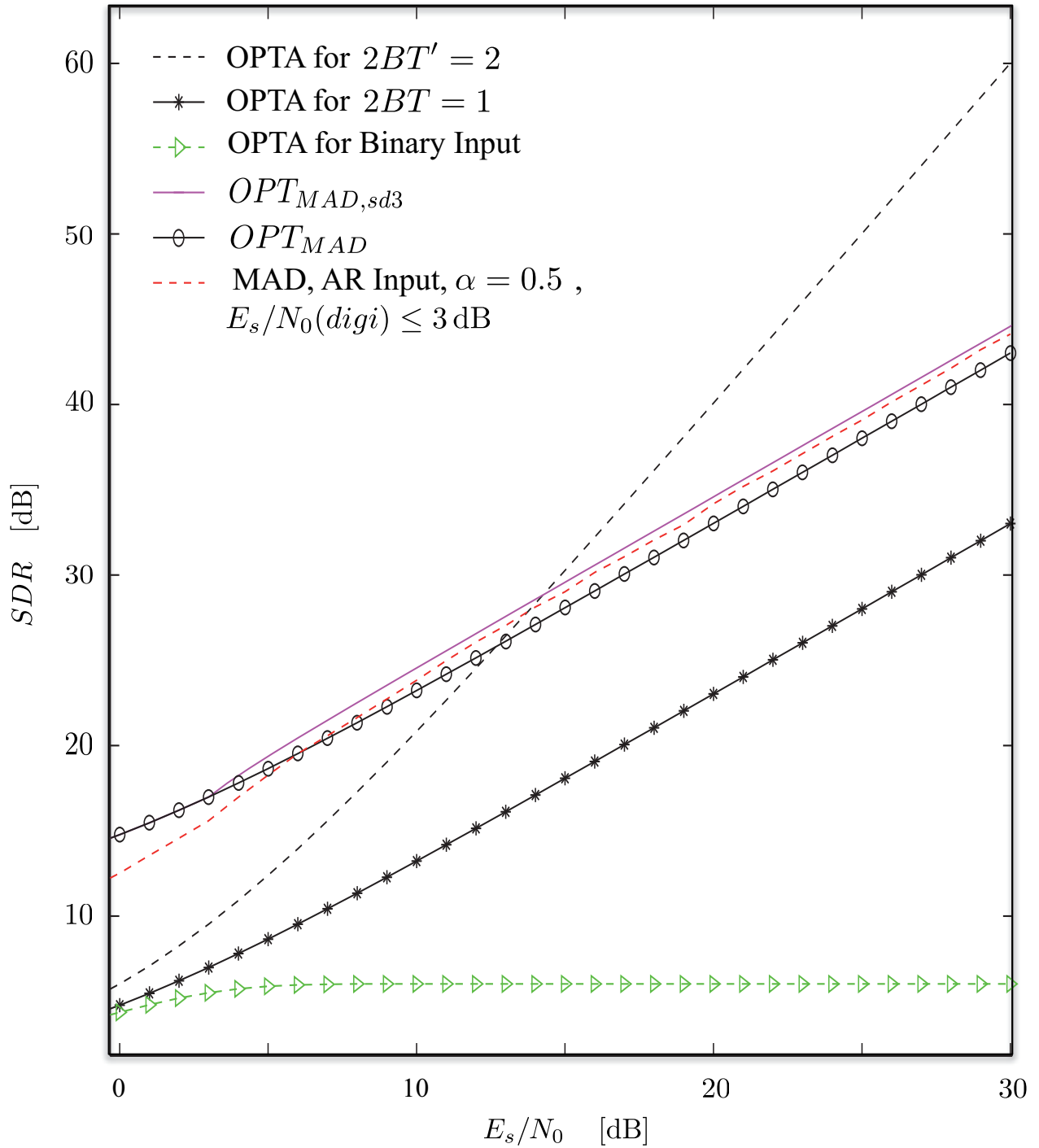


Figure 5.28: Performance of MAD transmission with unequal distribution of the total power. $E_{s,digi}/N_0 = 3$ dB for $E_s/N_0 \geq 3$ dB.

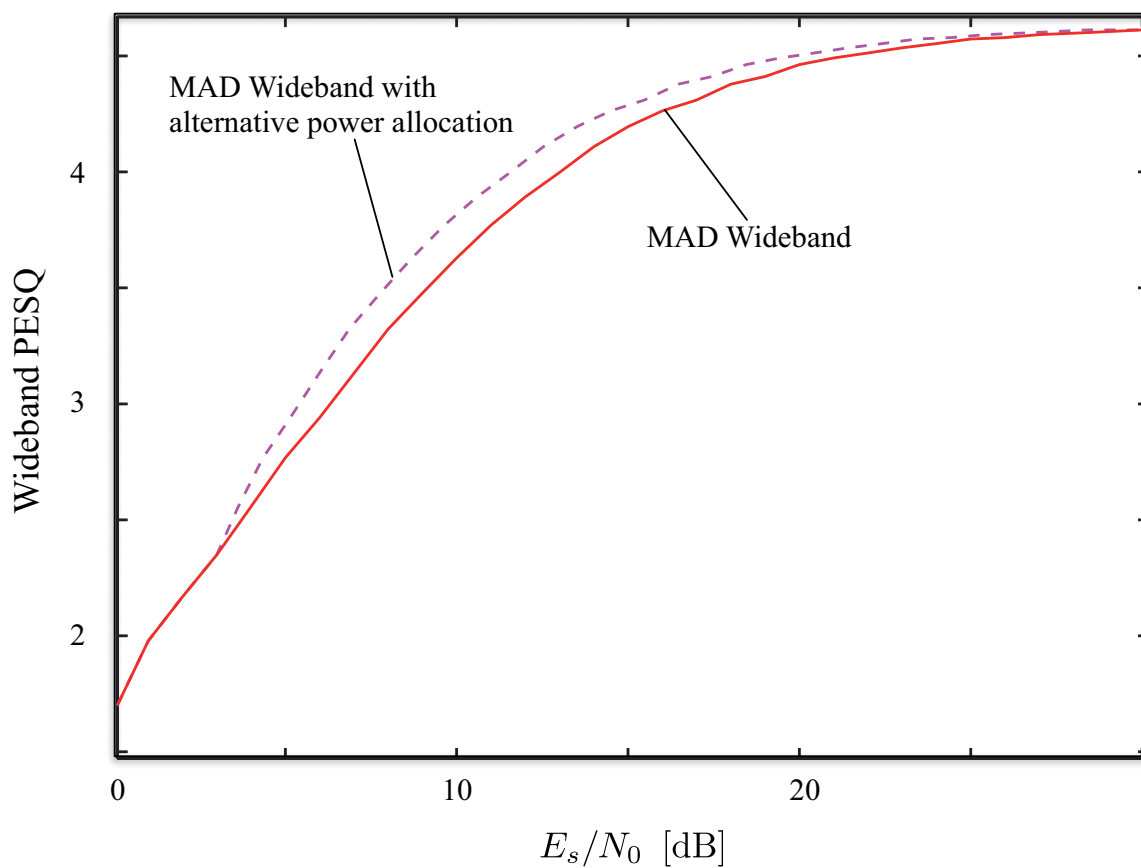


Figure 5.29: Perceptual Performance of MAD transmission with unequal distribution of the total power. $E_{s,digi}/N_0 = 3$ dB for $E_s/N_0 \geq 3$ dB.

5.4.4 Analysis of MAD using ASM Transmission

A practical MAD transmission system using the Archimedes spiral for mapping the normalized residual into the signal space has been introduced in Section 4.2.4. Based on the results of the previous sections this scheme shall be analyzed with respect to the OPTA bounds as defined in Section 5.3. Since ASM uses a 2-dimensional signal space, the performance bound is given in Equation (5.55). The transmission power and the bandwidth used for transmission of the normalized residual are not changed in comparison to PAM transmission.

For an AWGN channel of dimension $N = 1$ and a Gaussian source, PAM is the optimum system [29, 35]. Vaishampayan and Costa have shown in [59] that for $1 : N$, $N > 1$ Shannon-Kotel'nikov mappings, the optimum communication system is necessarily nonlinear. Furthermore Ziv has shown that no single modulation scheme can achieve the ideal rate distortion bound on the Mean Square Error for all SNR [61], which means a family of modulation schemes is needed to obtain the optimum MSE for each channel condition. This effect can well be studied in Figure 4.26, where, depending on the channel SNR, spirals with different tightness are optimum.

Mappings with $N > 1$ can be analyzed with the relevant theory from [79, 83, 90]. Expanding the dimension by a factor of N can be viewed as representing a scalar, continuous amplitude, and time discrete memoryless source $r \in \mathbb{R}$ by N components in an N -dimensional channel space ($\mathbf{y}(r) \in \mathbb{R}^N$). Each component can be a linear or non-linear function of the amplitude of r . They form a curve in the channel space referred to as *locus* (see Figures 5.31 to 5.33). With Gaussian noise $\mathbf{n} \in \mathbb{R}^N$ the received signal is

$$\hat{\mathbf{y}}(\mathbf{r}) = \mathbf{y}(r) + \mathbf{n} \quad (5.96)$$

with the likelihood function

$$f_{\hat{\mathbf{y}}(\mathbf{r})} = \left(\frac{1}{2\pi\sigma_n^2} \right)^{N/2} e^{-\frac{\|\hat{\mathbf{y}} - \mathbf{y}(r)\|^2}{2\sigma_n^2}} \quad (5.97)$$

which is maximized for minimum norm $\|\hat{\mathbf{y}}(\mathbf{r}) - \mathbf{y}(r)\|$. With large CSNR the maximum likelihood (ML) estimate, which corresponds to the point on the locus $\mathbf{y}(r)$ with the smallest geometrical distance to the received point, approaches that of the MSE optimum estimate (see [83]).

As shown in 4.2.4, the noise that applies to ASM transmission over an AWGN channel consists of two components, weak noise (the noise component parallel to the locus) and strong noise (the noise component perpendicular to the locus which may cause decoding of a wrong branch).

The Weak Noise Case

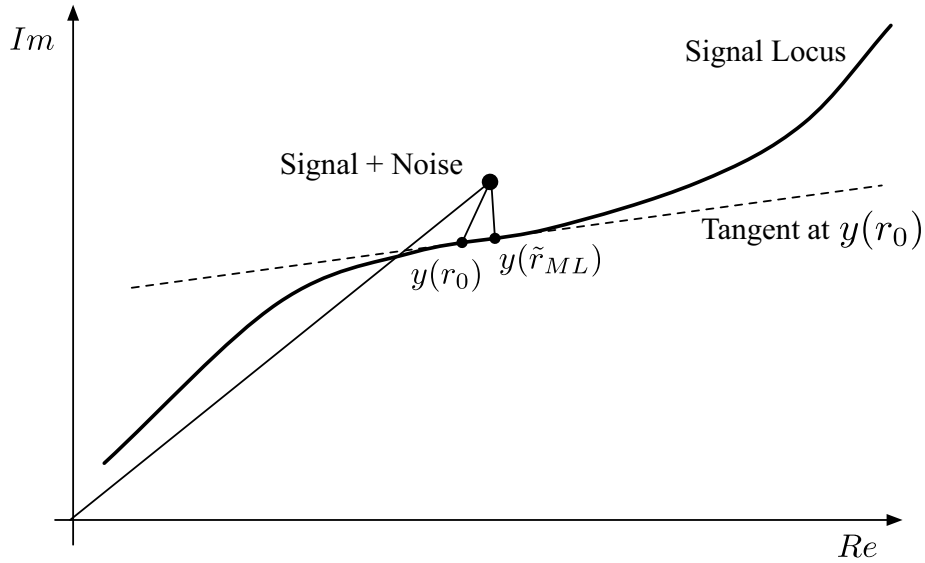


Figure 5.30: ML estimate in the weak noise case.

Assuming the locus is continuously differentiable, in the weak noise case, the locus can be approximated by

$$\mathbf{y}(r) \approx \mathbf{y}(r_0) + \mathbf{y}_0' \cdot (r - r_0) \quad (5.98)$$

with $\mathbf{y}_0' = \frac{d\mathbf{y}(r)}{dr}|_{r=r_0}$. As shown in Figure 5.30, the reconstructed sample $\mathbf{y}(\tilde{r}_{ML})$ corresponds to the projection of the received value onto the tangent through $\mathbf{y}(r_0)$ on the locus

$$\mathbf{y}(\tilde{r}_{ML}) = \mathbf{y}(r_0) + \frac{\langle \mathbf{n}, \mathbf{y}_0' \rangle}{\|\mathbf{y}_0'\|^2} \mathbf{y}_0' \quad (5.99)$$

with the inner product $\langle a, b \rangle = |a| \cdot |b| \cdot \cos(\text{angle}(a, b))$.

Given that r_0 was transmitted, the minimum mean square error (MMSE) is

$$\varepsilon_{LowNoise}^2 = E\{(\tilde{r}_{ML} - r)^2 | r = r_0\} \quad (5.100)$$

From (5.98) and (5.99) Kotel'nikov derived [79]

$$\varepsilon_{LowNoise}^2 = \frac{E\{\langle \mathbf{n}, \mathbf{y}_0' \rangle^2\}}{\|\mathbf{y}_0'\|^4} \quad (5.101)$$

which in case of Gaussian noise is [3]

$$\varepsilon_{LowNoise}^2 = \frac{\sigma_n^2}{\|\mathbf{y}'(r_0)\|^2} \quad (5.102)$$

With the Euclidean norm of the tangent in the denominator, the impact of the noise is only reduced, if the length of the locus is increased. Simple bending without stretching induces no change. Limiting the power, thus, makes it necessary to use a nonlinear mapping (the locus is no straight line, see Figures 5.31 to 5.33) to increase the length of the locus in order to get closer to OPTA.

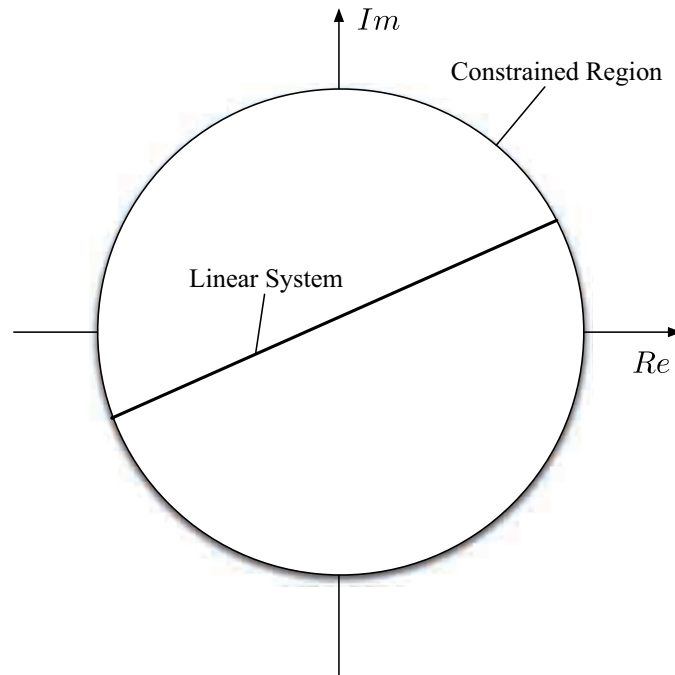


Figure 5.31: 1 : 2 Shannon-Kotel'nikov mapping; the straight line illustrates the locus of a linear mapping $r \rightarrow y(r) = mr + q$ with $m, q \in \mathbb{R}$. The circle represents a region of constrained power.

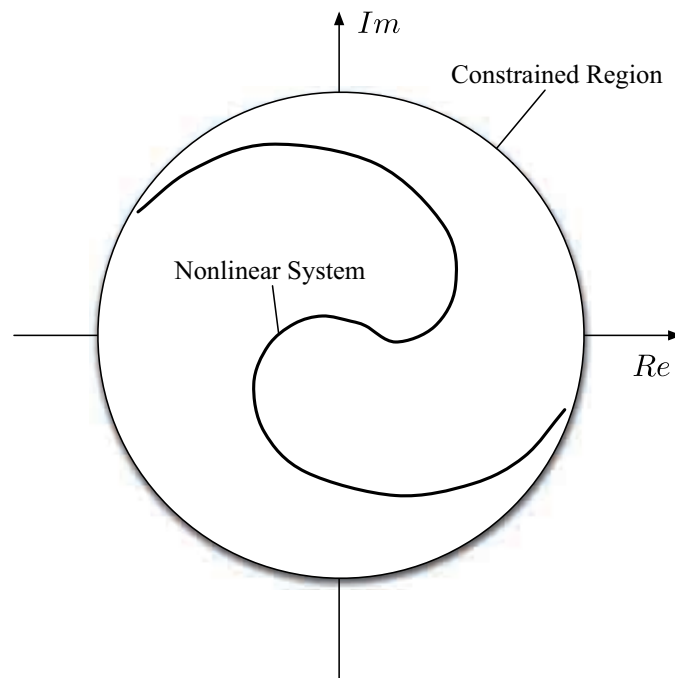


Figure 5.32: 1 : 2 Shannon-Kotel'nikov mapping; the curved line represents the locus of a nonlinear mapping.

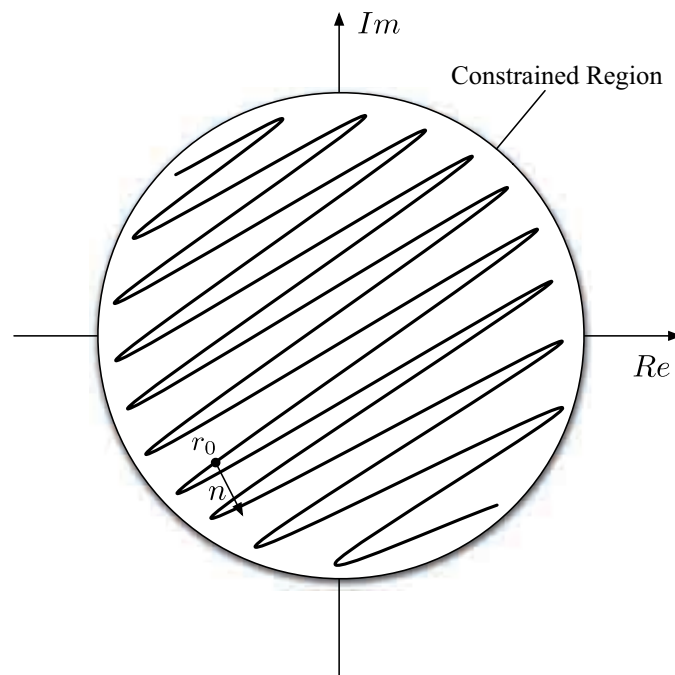


Figure 5.33: 1 : 2 Shannon-Kotel'nikov mapping; nonlinear mapping that has been stretched by a significant amount.

In [3] and [4], Ramstad and Floor calculate the SDR of an Archimedes spiral with slightly different mapping of r_n to their spiral. Adopting these calculations to the spiral used for ASM, from the definition in (4.11) follows

$$y(r_n) = r_n \cos\left(\frac{|r_n|}{c}\right) + jr_n \sin\left(\frac{|r_n|}{c}\right). \quad (5.103)$$

The derivative $y'(r_n)$ is found for real and imaginary part separately:

$$y'_{real}(r_n) = \cos\left(\frac{|r_n|}{c}\right) - \frac{|r_n|}{c} \sin\left(\frac{|r_n|}{c}\right) \quad (5.104)$$

$$y'_{imaginary}(r_n) = j \left[\sin\left(\frac{|r_n|}{c}\right) + \frac{|r_n|}{c} \cos\left(\frac{|r_n|}{c}\right) \right] \quad (5.105)$$

and thus

$$y'(r_n) = \left[\cos\left(\frac{|r_n|}{c}\right) - \frac{|r_n|}{c} \sin\left(\frac{|r_n|}{c}\right) \right] + j \left[\sin\left(\frac{|r_n|}{c}\right) + \frac{|r_n|}{c} \cos\left(\frac{|r_n|}{c}\right) \right]. \quad (5.106)$$

The Euclidean norm $\|y'(r_n)\|^2$ is

$$\begin{aligned} \|y'(r_n)\|^2 &= \left[\cos\left(\frac{|r_n|}{c}\right) - \frac{|r_n|}{c} \sin\left(\frac{|r_n|}{c}\right) \right]^2 + \left[\sin\left(\frac{|r_n|}{c}\right) + \frac{|r_n|}{c} \cos\left(\frac{|r_n|}{c}\right) \right]^2 \\ &= 1 + \frac{|r_n|^2}{c^2} = 1 + \frac{r_n^2}{c^2}. \end{aligned} \quad (5.107)$$

Averaging the low noise for all r_n leads to

$$\bar{\epsilon}_{LowNoise}^2 = \sigma_n^2 \int_{-\infty}^{\infty} \frac{1}{\|y'(r_n)\|^2} p_{r_n}(r_n) dr_n, \quad (5.108)$$

where $p_{r_n}(r_n)$ is the probability density function of r_n . Assuming Gaussian distribution of r_n yields

$$\bar{\epsilon}_{LowNoise}^2 = \sigma_n^2 \int_{-\infty}^{\infty} \frac{1}{1 + \frac{r_n^2}{c^2}} \frac{1}{\sigma_{r_n} \sqrt{2\pi}} e^{-\frac{r_n^2}{2\sigma_{r_n}^2}} dr_n. \quad (5.109)$$

The Strong Noise Case

Increasing the length of the locus constrained to limited power, results in different parts (branches) of the locus coming closer to each other which increases the probability P_{error} of decoding to the wrong branch⁷ (σ_n being significant compared

⁷While P_e in Equation (4.25) is the combined probability for any threshold error, P_{error} is the probability for either error, decoding the outer or inner branch. From the symmetry of $p_{\delta_r}(\delta_r)$ follows $P_{error} = \frac{1}{2}P_e$.

to the distance Δ between different parts of the locus, see Figure 5.34). Unfortunately, this threshold effect cannot be avoided when using nonlinear mappings with constrained power (see Figures 5.31 to 5.33). The objective to make the locus as long as possible while keeping the probability of the threshold effect sufficiently low equals the situation in channel coding, where the maximum number of codewords shall be put in a constrained region while keeping their distance large enough.

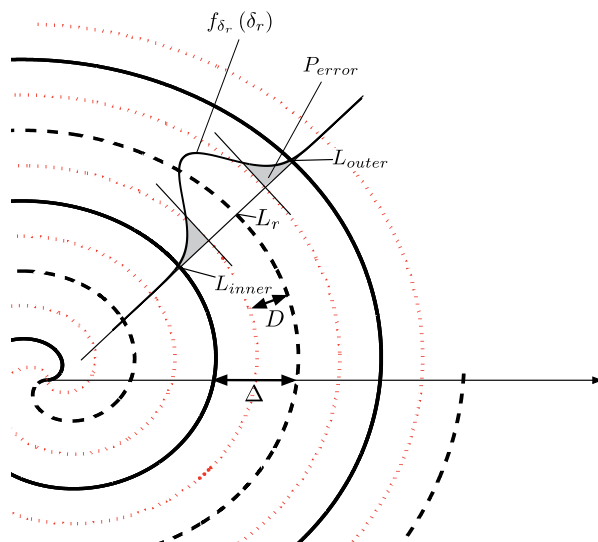


Figure 5.34: Approximate calculation of the threshold noise.

The minimum distance between branches is $2D$ (see Figure 5.34). Following [3], D is approximated as

$$D = \frac{\Delta}{2} \quad (5.110)$$

as there is no closed representation for D . Assuming complex additive white Gaussian noise with independent real and imaginary part, the distribution of this noise is symmetrical towards rotation and thus any section will have a Gaussian distribution with variance σ_n^2 . The probability $2 \cdot P_{error}$ of decoding to either wrong branch equals the probability of $\delta_r = |r_n - \tilde{r}_n| > D$. For δ_r having a probability density function $p_{\delta_r}(\delta_r)$,

$$\begin{aligned} P_{error} &= \frac{1}{2} P(\delta_r > D) = \frac{1}{2} (1 - P(\delta_r < D)) \\ &= \frac{1}{2} \left(1 - 2 \int_0^D p_{\delta_r}(\delta_r) d\delta_r \right). \end{aligned} \quad (5.111)$$

Assuming Gaussian distribution of δ_r yields

$$\begin{aligned} P_{error} &= \frac{1}{2} \left(1 - 2 \int_0^D \frac{1}{\sqrt{2\pi}} e^{-\frac{\delta_r^2}{2}} d\delta_r \right) \\ &= \frac{1}{2} \left(1 - \operatorname{erf} \left(\frac{D}{\sqrt{2}\sigma_n} \right) \right) = \frac{1}{2} \operatorname{erfc} \left(\frac{D}{\sqrt{2}\sigma_n} \right) \end{aligned} \quad (5.112)$$

and with (5.110)

$$P_{error} = \frac{1}{2} \operatorname{erfc} \left(\frac{\pi \cdot c}{2\sqrt{2}\sigma_n} \right). \quad (5.113)$$

To estimate the average threshold noise, given that L_r (radius r_n) was transmitted, the average power of error resulting from decoding L_{inner} on the inner neighbour branch (radius $r_{n,inner}$, inverse sign compared to r_n)

$$\bar{\varepsilon}_{L_{inner}}^2 = \int_{-\infty}^{\infty} P_{error} \cdot (r_n - r_{n,inner})^2 p_{r_n}(r_n) dr_n \quad (5.114)$$

and that of decoding L_{outer} on the outer branch (radius $r_{n,outer}$, again inverse sign compared to r_n),

$$\bar{\varepsilon}_{L_{outer}}^2 = \int_{-\infty}^{\infty} P_{error} \cdot (r_n - r_{n,outer})^2 p_{r_n}(r_n) dr_n \quad (5.115)$$

have to be summed up⁸.

$$\begin{aligned} \bar{\varepsilon}_{StrongNoise}^2 &= \int_{-\infty}^{\infty} P_{error} \cdot (r_n - r_{n,inner})^2 p_{r_n}(r_n) dr_n \\ &\quad + \int_{-\infty}^{\infty} P_{error} \cdot (r_n - r_{n,outer})^2 p_{r_n}(r_n) dr_n. \end{aligned} \quad (5.116)$$

and thus

$$\bar{\varepsilon}_{StrongNoise}^2 = P_{error} \int_{-\infty}^{\infty} [(r_n - r_{n,inner})^2 + (r_n - r_{n,outer})^2] p_{r_n}(r_n) dr_n. \quad (5.117)$$

With the distance between branches $\Delta = \pi \cdot c$ (see Equation (4.15)) $|r_{n,outer}|$ and $|r_{n,inner}|$ can be calculated as

$$\begin{aligned} |r_{n,inner}| &= |r_n| - \pi \cdot c \\ |r_{n,outer}| &= |r_n| + \pi \cdot c. \end{aligned}$$

With the alternating sign of neighbouring branches respected, $r_{n,outer}$ and $r_{n,inner}$ follow to be

$$\begin{aligned} r_{n,inner} &= -(r_n - \pi \cdot c) \\ r_{n,outer} &= -(r_n + \pi \cdot c). \end{aligned}$$

⁸The noise from decoding L_r to the second neighbouring branch outside of L_{outer} or inside of L_{inner} can be neglected, as the probability for this kind of error is low and the decoded sign equals that of r_n , which makes the power of the error small. For branches further away the probability of error is neglectably small.

Thus,

$$\begin{aligned}
& \bar{\varepsilon}_{StrongNoise}^2 \\
&= \frac{1}{2} \left(1 - \operatorname{erf} \left(\frac{\pi \cdot c}{2\sqrt{2}\sigma_n} \right) \right) \cdot \int_{-\infty}^{\infty} \left((2r_n - \pi c)^2 + (2r_n + \pi c)^2 \right) \frac{1}{\sigma_{r_n} \sqrt{2\pi}} e^{-\frac{r_n^2}{2\sigma_{r_n}^2}} dr_n \\
&= \frac{1}{2} \left(1 - \operatorname{erf} \left(\frac{\pi \cdot c}{2\sqrt{2}\sigma_n} \right) \right) \cdot \int_{-\infty}^{\infty} (8r_n^2 + 2\pi^2 \cdot c^2) \frac{1}{\sigma_{r_n} \sqrt{2\pi}} e^{-\frac{r_n^2}{2\sigma_{r_n}^2}} dr_n \quad (5.118)
\end{aligned}$$

Combination of Weak Noise and Strong Noise

Including weak and strong noise, the total SDR of ASM can be calculated as

$$SDR = \frac{\sigma_{r_n}^2}{\bar{\varepsilon}_{LowNoise}^2 + \bar{\varepsilon}_{StrongNoise}^2}. \quad (5.119)$$

This is depicted in Figure 5.35 for different tightness factors c .

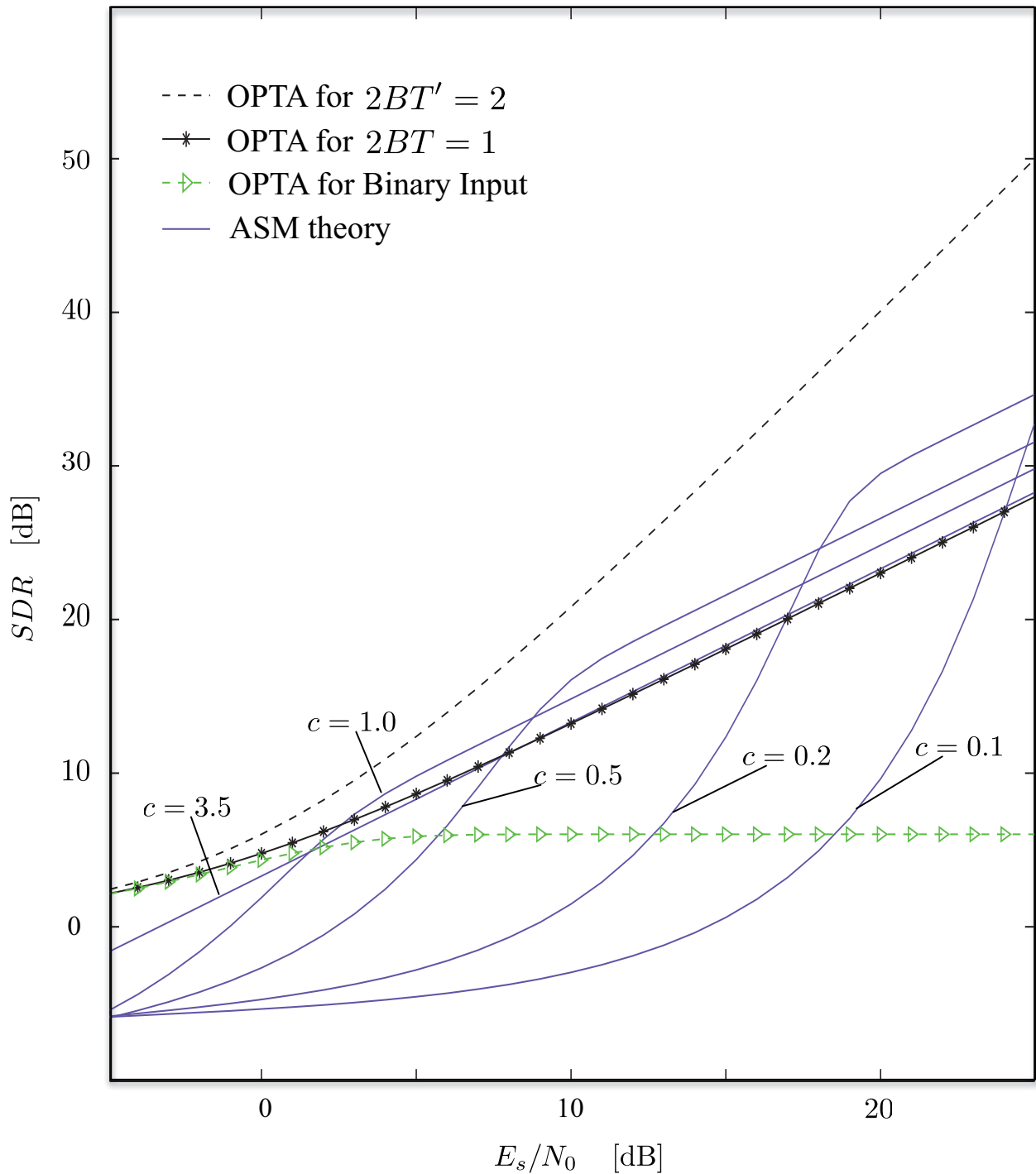


Figure 5.35: Theoretical performance of ASM with different tightness factors c .

Simulations of MAD using ASM transmission and the AR input signal defined in Section 5.4.3 show a good match between theory and simulation, as can be seen in Figure 5.36.

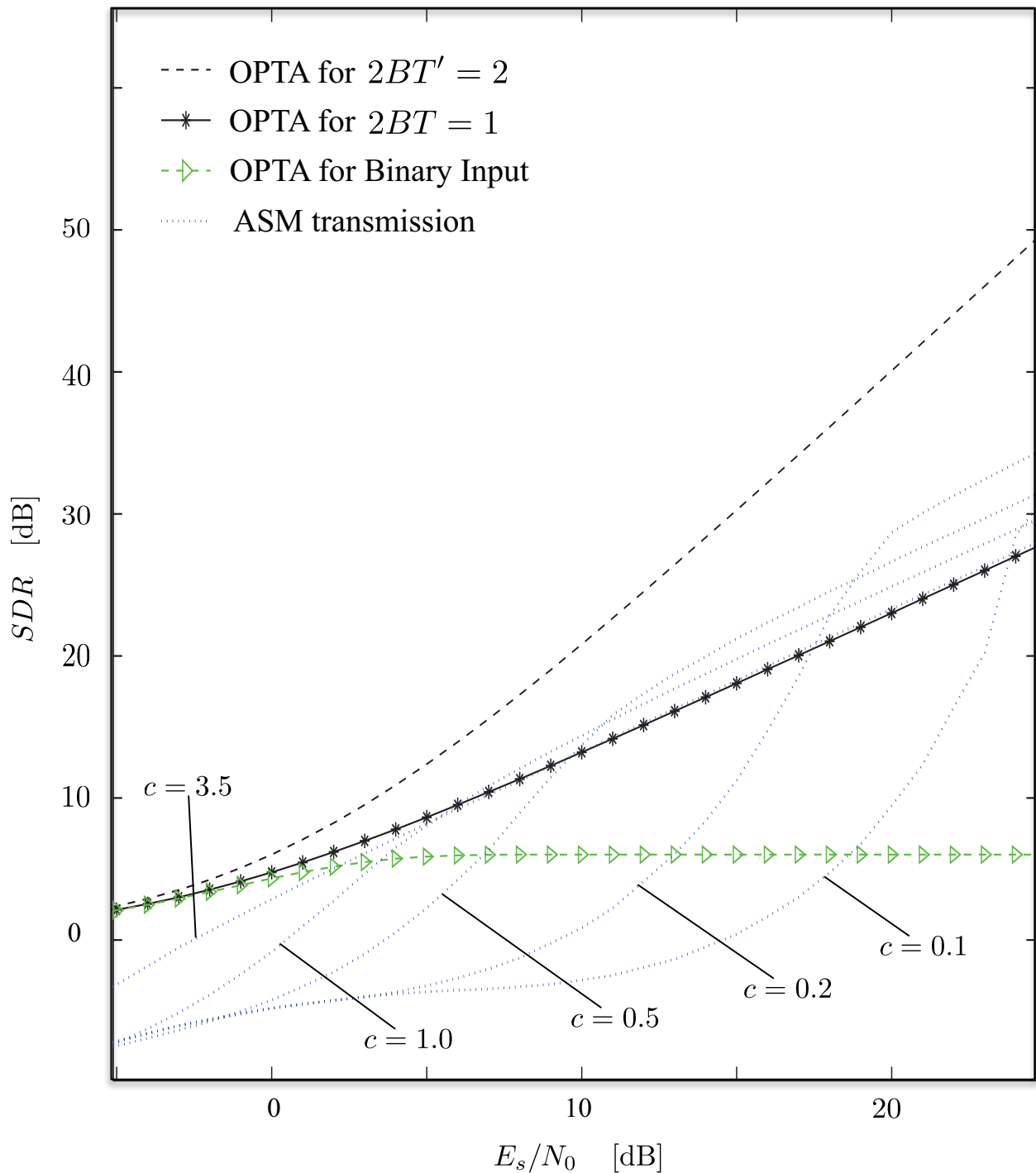


Figure 5.36: Measured performance of MAD using ASM with different tightness factors c for the AR input signal defined in Section 5.4.3.

Optimum Tightness

Looking at Figure 5.35, it is obvious that, for optimum performance, the tightness factor c should be chosen differently for different channel conditions. With carefully

chosen c there is a considerable gain compared to the optimum linear system (PAM). As there is no closed solution, Equation (5.119) needs to be maximized to obtain the optimum tightness factor c depending on σ_n . Figure 5.37 shows the result of the numerical maximization.

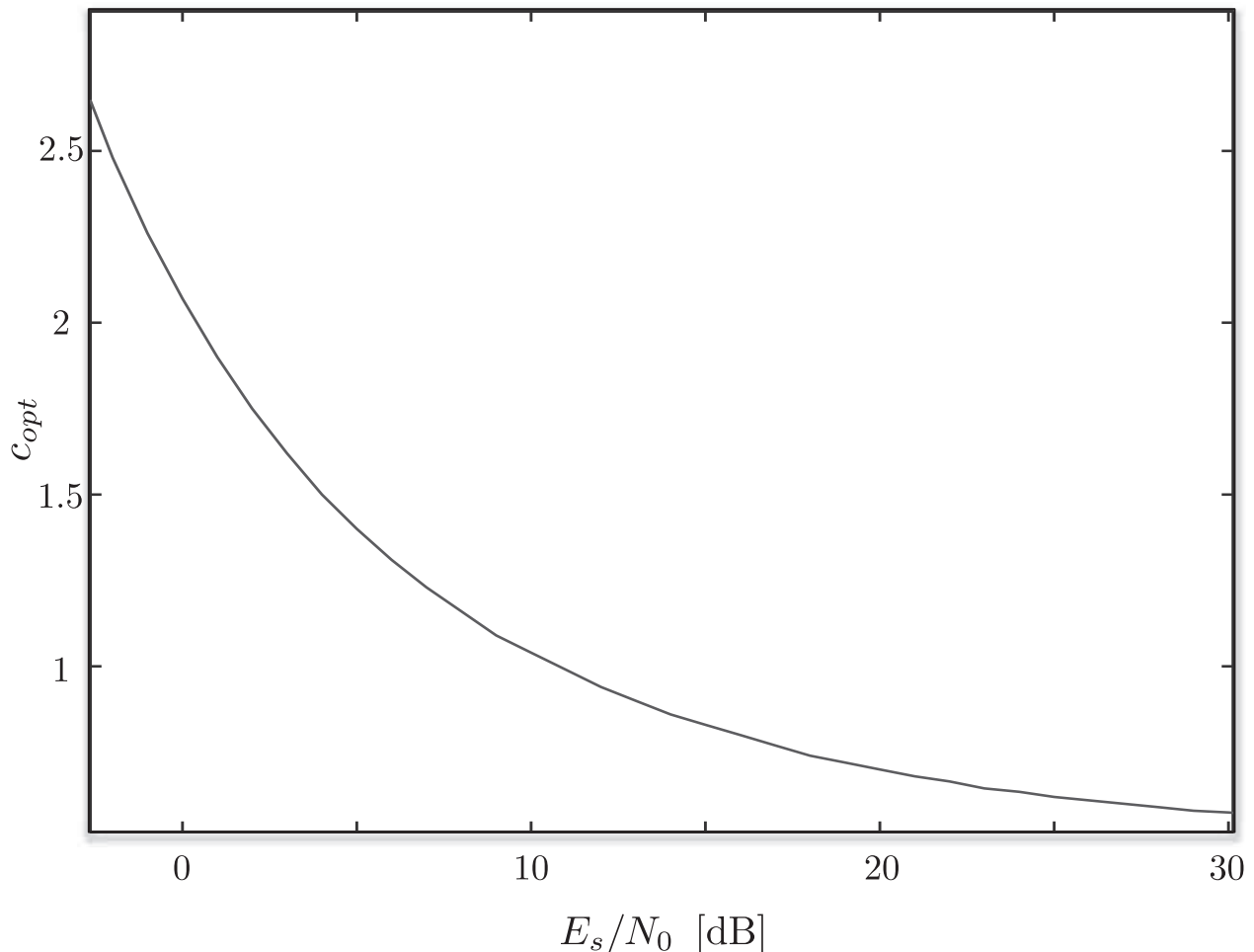


Figure 5.37: The optimum tightness factor c results from maximizing Equation (5.119). It depends on the channel SNR.

It should be emphasized that ASM does not introduce any delay (important for delay constrained applications as, e.g., in Section 6.2.1), does not use more bandwidth than double-sided PAM, and that its encoding/decoding process is very simple.

Optimum Signal Distribution

While the theory derived above matches ASM transmission as introduced in Section 4.2.4, it does not describe the optimum mapping to an Archimedes spiral (oASM), as the amplitude r_n by definition is the same as that of PAM.

Generalizing the amplitude constraint by allowing for modification of the distribution of r_n on the locus of the spiral, Floor and Ramstad have shown in [3] that for optimum SDR the mapping must follow

$$y(r_n) = \frac{\Delta}{\pi} (g_{osd} \phi(r_n)) \exp(j(g_{osd} \phi(r_n))) \quad (5.120)$$

with a gain factor g_{osd} to define the average power and $\phi(r_n)$ the inverse of the length of the spiral L_{spiral} , which is approximately proportional to $\sqrt{r_n}$. The average weak noise in this case is

$$\bar{\epsilon}_{LowNoise,oASM}^2 = \frac{\sigma_n^2}{g_{osd}^4} \quad (5.121)$$

while the strong noise is calculated as

$$\begin{aligned} \bar{\epsilon}_{StrongNoise,oASM}^2 = \frac{1}{2} \left(1 - \operatorname{erf} \left(\frac{\Delta}{2\sqrt{2}\sigma_n} \right) \right) \\ \int_{-\infty}^{\infty} [(r_n - r_n^+)^2 + (r_n - r_n^-)^2] p_{r_n}(r_n) dr_n \end{aligned} \quad (5.122)$$

with

$$r_n^{\pm} = -0.16\Delta \left(\sqrt{\frac{r_n}{0.16\Delta}} \mp \frac{\pi}{g_{osd}} \right)^2 \quad (5.123)$$

(see [3]). The theoretical behaviour of the oASM scheme is shown in Figure 5.38.

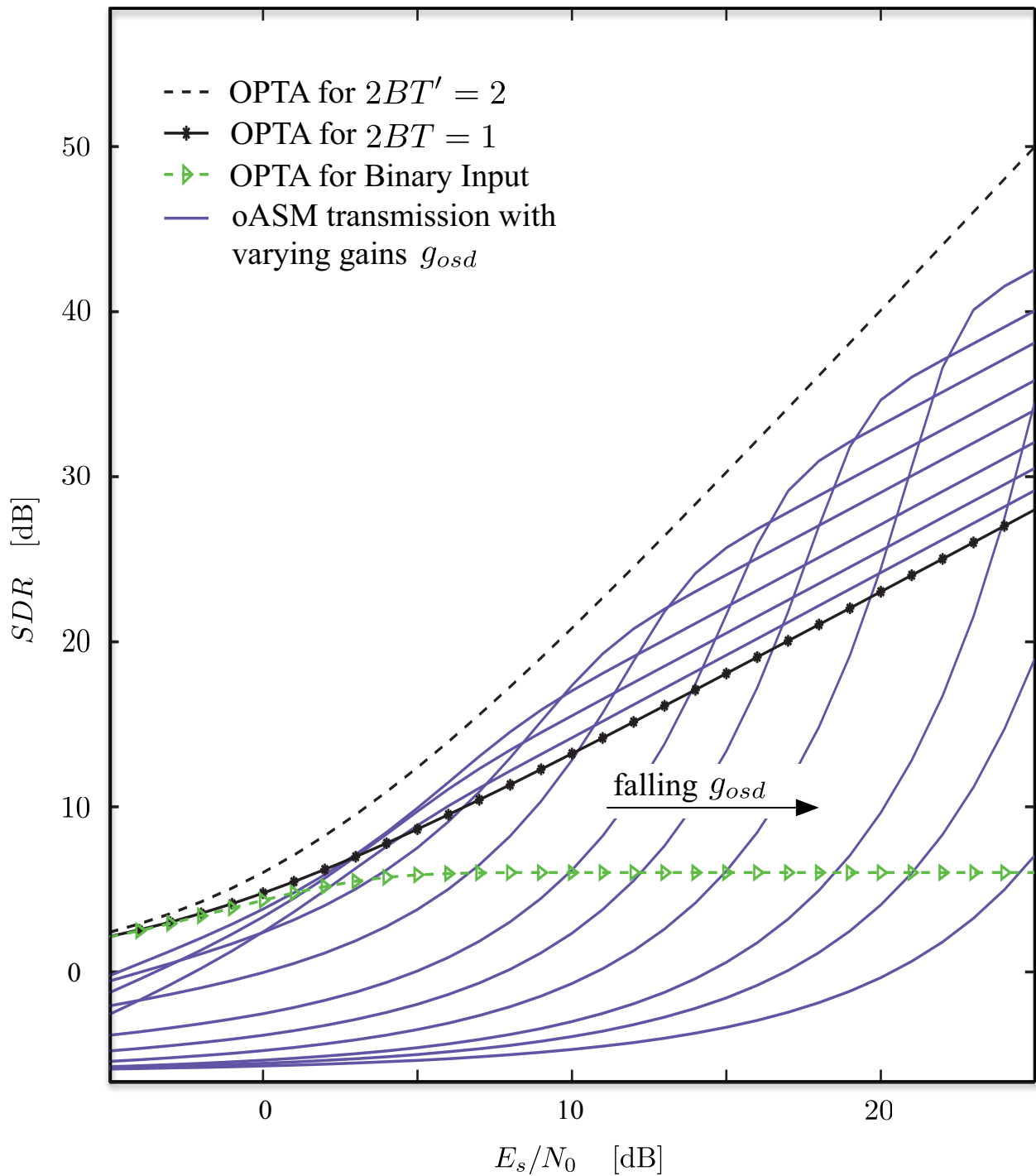


Figure 5.38: Optimum performance of ASM with different gains g_{osc} .

Optimum ASM in the context of MAD

In MAD transmission the pseudo analogue residual must be transmitted with an average power of $\sigma_{rn} = 1$. Thus, the optimum system described in the previous section cannot be applied to MAD exactly as defined in [3]. Instead, the gain factor

g_{osd} is combined with the gain factor g available in MAD and updated blockwise. The resulting block diagram is given in Figure 5.39.

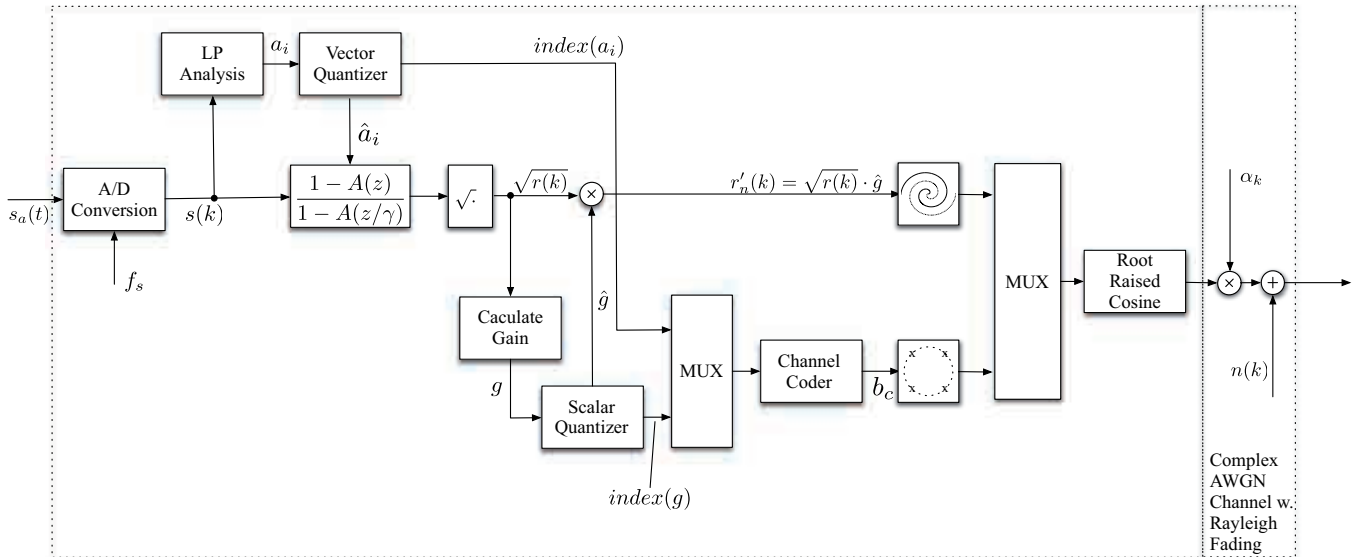


Figure 5.39: Optimum ASM for MAD transmission.

The resulting performance in terms of SDR of optimum ASM transmission using the AR input signal from Section 5.4.3 is shown in Figure 5.40.

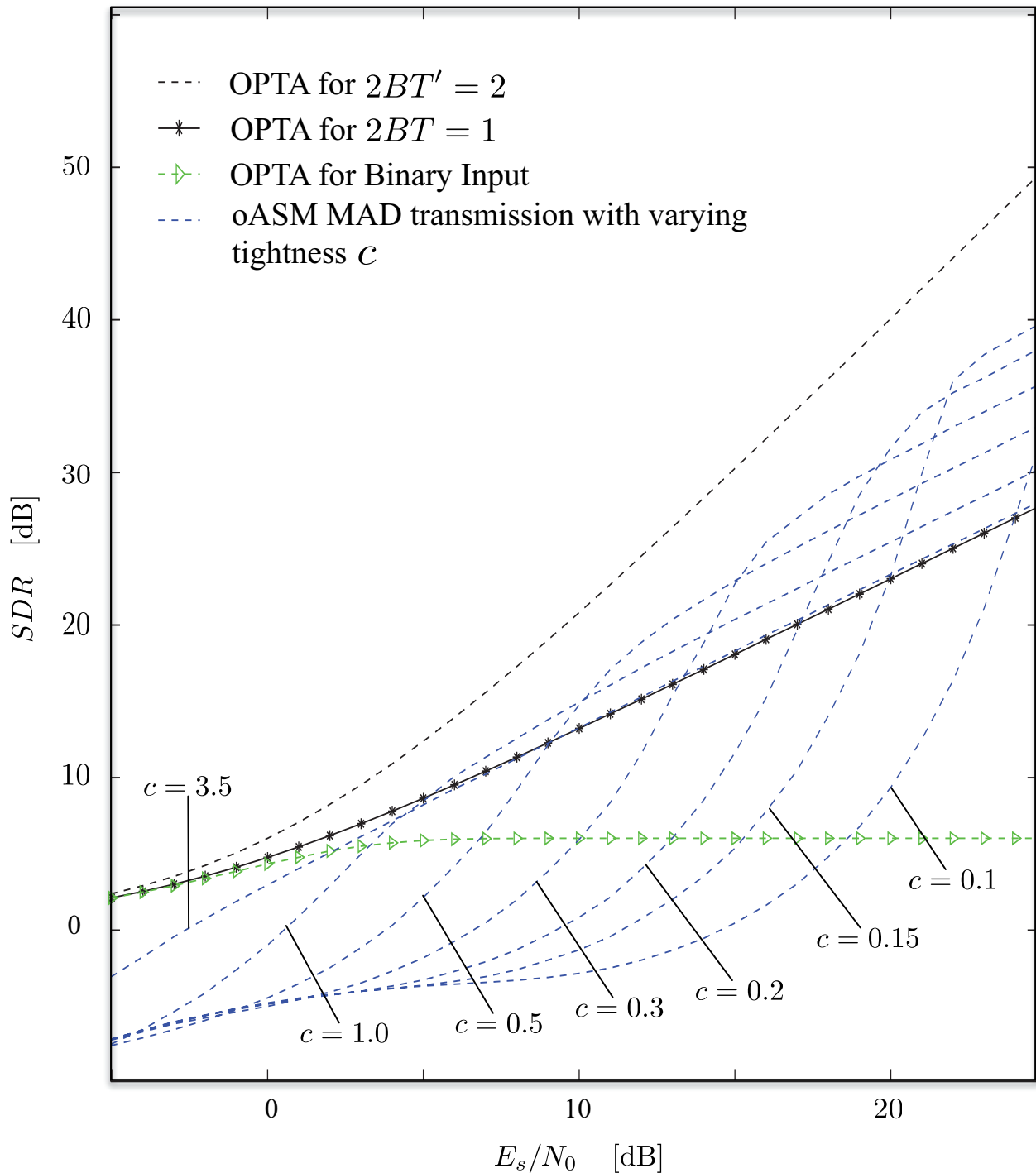


Figure 5.40: Measured SDR of MAD with optimum ASM (AR Input).

Varying the distribution of transmission power between model and residual corresponding to Section 5.4.3 allows to minimize the degradation. The resulting performance using optimum tightness factors c (see Section 5.4.4) is shown in Figure 5.41 in the residual domain as well as in the speech domain.

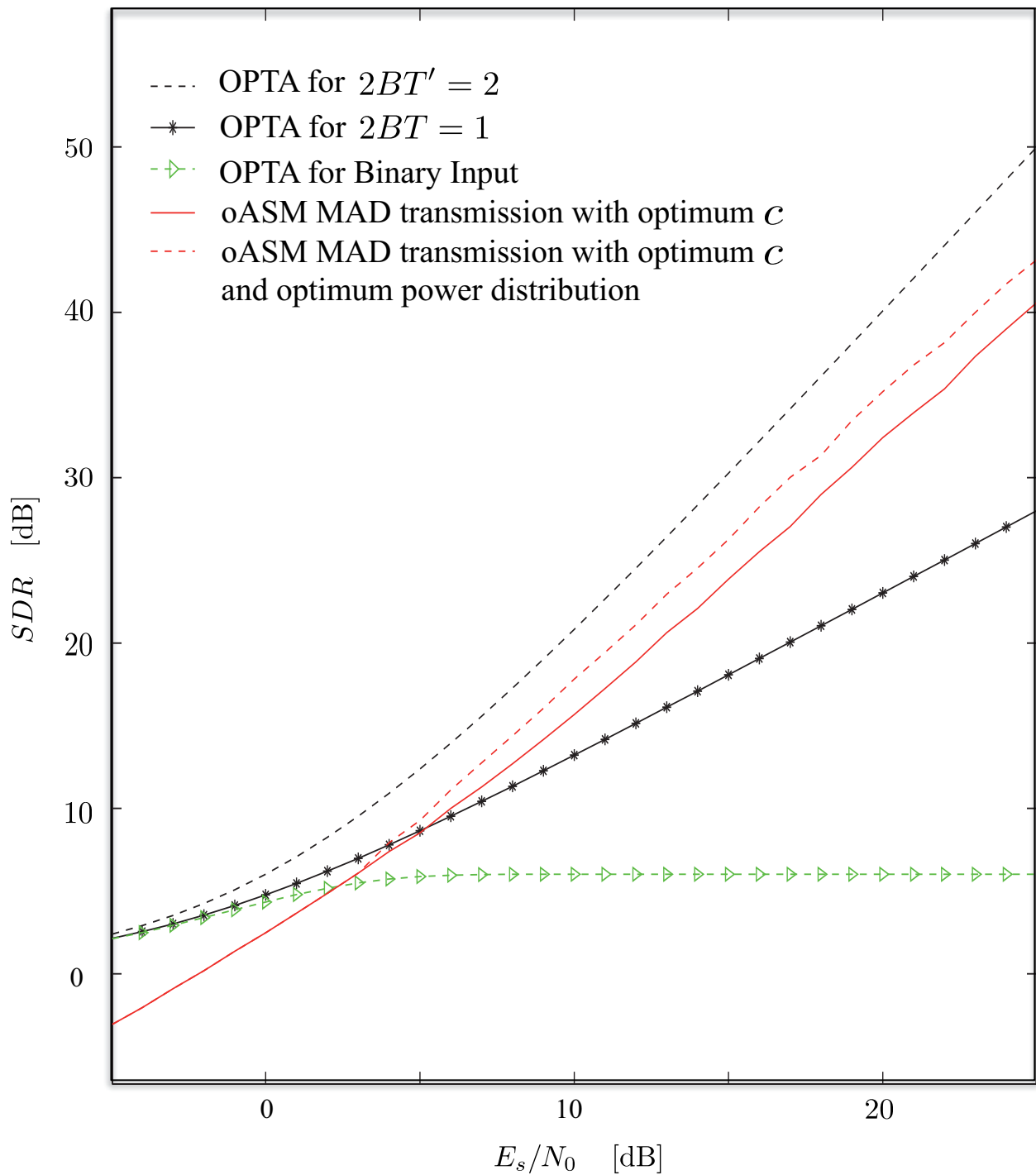


Figure 5.41: Performance of MAD transmission using optimum ASM and optimum tightness factors c in the residual domain.

Using the 1.3 million samples of natural speech defined as input signal before, the subjective performance of MAD incorporating all the optimizations derived in this chapter is shown in Figure 5.42. Although an improvement for high E_s/N_0

is hard to perceive, the improvement over the initial MAD transmission concept is substantial between 10 dB and 20 dB E_s/N_0 .

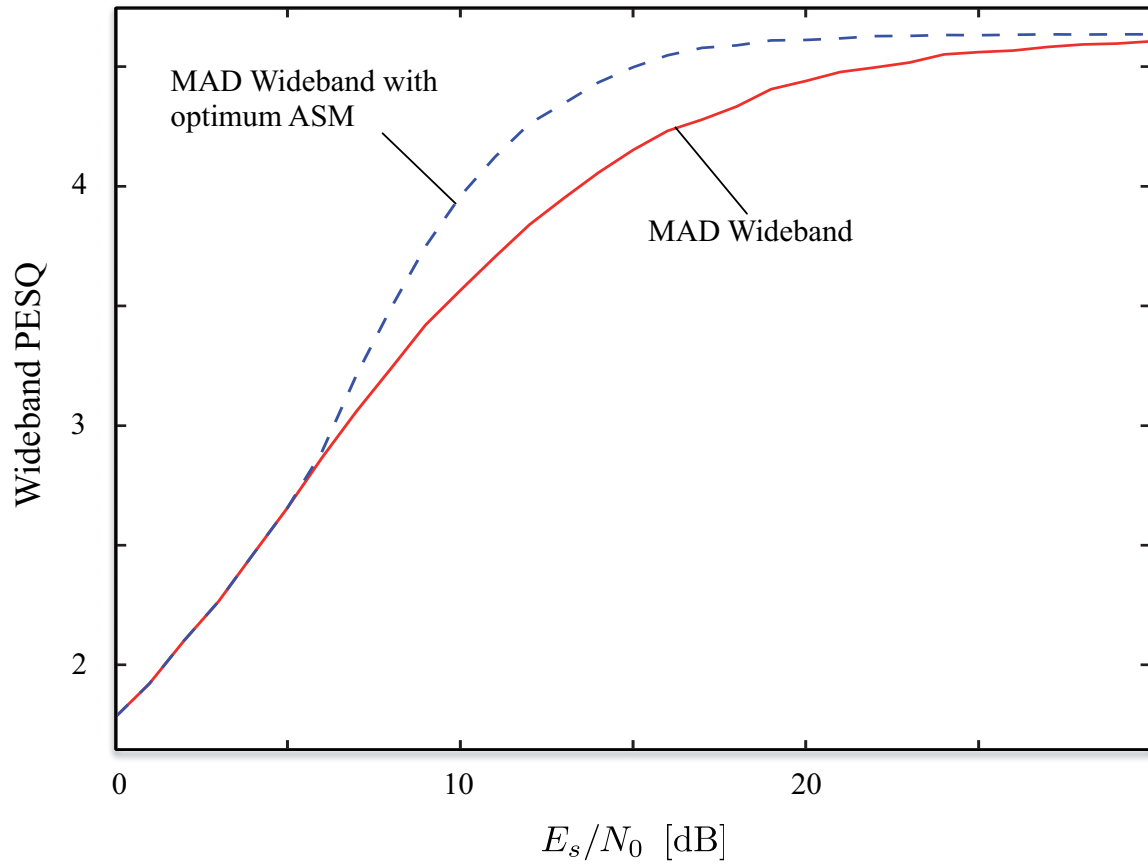


Figure 5.42: Perceptual quality of MAD using optimized ASM.

6

Application of MAD Transmission

6.1 MAD Speech Transmission

The MAD speech transmission concept introduced in Section 3.3 and evaluated in Chapter 5 is based on the AMR speech codec. There are numerous aspects which could benefit from an optimization for MAD transmission, e.g., LP filtering [49], frame structure, channel coding, etc.

An optimized MAD speech transmission system could be applied in any scenario, where narrow bandwidth, high quality, low delay and low complexity are the main demands.

6.2 MAD Audio Transmission

Today, the most widespread techniques for digital coding of audio signals are MP3¹, or AAC², respectively. These use lossy compression based on a hearing model, (M)DCT³, and spectral quantization to create a compact binary representation of the audio signal targeting primarily at *storage* of audio signals. Good quality is achieved with 64 kbit/s for AAC, or 128 kbit/s for MP3, respectively.

MAD audio, on the other hand, was designed for *transmission* of audio signals. Storage of the normalized residual would take as much space on a digital storage device, as would storage of the PCM samples of the original signal. The digital information even adds up to that. Thus, using MAD audio for storage is not an option.

Considering audio transmission using the model described in Section 3.3.2 and 1/2-rate channel coding as introduced in Section 3.3.1, the band pass bandwidth

¹Motion Picture Experts Group MPEG-1 Audio Layer 3

²Advanced Audio Coding

³(Modified) Discrete Cosine Transform

(assuming a root raised cosine filter with $\alpha = 1.5$) needed for digital transmission of AAC and MP3 can be calculated to be

$$B'_{Audio,AAC} = 64 \text{ kbit/s} \cdot 2 \cdot 1.5 \frac{Hz}{bit} = 192 \text{ kHz}, \quad (6.1)$$

or

$$B'_{Audio,MP3} = 128 \text{ kbit/s} \cdot 2 \cdot 1.5 \frac{Hz}{bit} = 384 \text{ kHz}, \quad (6.2)$$

while MAD audio providing CD quality for a mono signal (sampling rate $f_s = 44.1 \text{ kHz}$) only needs a bandwidth of

$$B'_{MAD_{Audio44.1, mono}} = 1.5 \cdot (f_{s_{Audio}} + R_{digi}) \quad (6.3)$$

$$= (44.1 + 7) \cdot 1.5 \text{ kHz} \quad (6.4)$$

$$= 76.65 \text{ kHz}. \quad (6.5)$$

The stereo signal which is more comparable to AAC or MP3 thus needs a bandwidth of

$$B'_{MAD_{Audio44.1, stereo}} \leq 2 \cdot B'_{MAD_{Audio44.1, mono}} = 153.3 \text{ kHz}. \quad (6.6)$$

This example shows the benefit of using MAD audio for transmission purposes.

An especially interesting application of a low-delay version of MAD audio is suggested in the following section.

6.2.1 MAD Wireless Microphones, Headsets or Hearing Aids with Wireless Audio Input

In Section 6.2 MAD transmission has been used to transmit general audio signals. Applying the transmission of audio signals to equipment which is used for real-time transmission, as, e.g., wireless microphones in a live show or concert, or wireless headsets which shall deliver an audio signal, which is lip-synchronous with a video stream (e.g. TV headsets), the total delay which is allowed for the transmission and processing will strictly be constrained. The general advantages of MAD transmission can still be exploited in such a scenario, however, the standard LP predictor introduced in Section 3.3.1 same as the channel coder (primarily the interleaver) would have to be modified or replaced with low delay versions. Possible solutions for the predictor include backward prediction or more frequent filter updates as used, e.g., in [14]. Considering the interleaver, the size must be minimized constrained to the probability of burst errors of the channel in use. Additionally, error concealment techniques must be introduced.

Targeting high definition audio, also the sampling rate must be adjusted to the desired frequency range, mostly 48 kHz in current applications.

6.3 Further Studies & Prospects

The scope of this contribution was to present the ideas of MAD transmission and to analyze the theoretic background thereof. Extending the basic principles, further research might follow from these first results. Some ideas are briefly sketched in the following.

6.3.1 Pseudo Analogue Channel Coding

For the digital information a convolutional channel coder is used in MAD to allow for Forward Error Correction (FEC). The theory of separating source and channel coding has been introduced by Shannon [86] and is common practice.

Concerning analogue signals, in Section 4.2.4 it was pointed out that Shannon mappings can be regarded as error control coding [2]. These perform very well and help MAD-ASM to operate substantially closer to the Shannon limit than MAD using PAM.

Other sources have begun to look into the application of traditionally digital channel coding techniques like Turbo coding [10, 18, 40], or Low density parity check (LDPC) codes [50] to analogue signals. Alternative schemes based on chaotic systems include [23, 30, 59].

Gastpar, Rimoldi, and Vetterli point out in [33], that if a source and a channel are probabilistically matched, the communication system is optimum without any channel coding. The Gaussian distribution of the normalized residual in combination with a Gaussian channel is the main reason for MAD outperforming other systems. Thus, not only analogue channel coding in the traditional way of thinking is worth further elaboration, but also a filter structure which combines noise shaping with probabilistic source-channel-matching.

6.3.2 Channel Adaptive MAD

Matching the source to the channel by a revised LP filter is not the only area, where a more sophisticated MAD transmission channel should incorporate feedback about the actual channel situation. ASM needs to adapt the tightness factor c , the power distribution between analogue and digital parts should base on the channel condition, to allow for performance close to the Shannon limit, which is never possible with a static system (compare [61]).

7

Summary & Conclusions

In this contribution a novel speech and audio transmission scheme has been introduced which combines both, analogue and digital techniques, for the sake of reduced transmission bandwidth, reduced computational complexity, and increased signal quality compared to purely analogue or digital transmission. The novel scheme is called Mixed Pseudo Analogue-Digital (MAD) transmission.

After a short introduction to analogue transmission, digital transmission, and mixed transmission, the fundamental concepts proposed in this thesis have been established in Chapter 2. Starting with the definition of narrowband signals (300 Hz - 3400 Hz audio bandwidth) and wideband signals (50 Hz - 7000 Hz audio bandwidth), the basic speech coding principle of Linear Prediction (LP), has been described. This was followed by a brief presentation of Residual Excited Linear Prediction (RELP) and Code Excited Linear Prediction (CELP) together with the most common channel coding strategies (block codes and convolutional codes). Finally, two instrumental means of quality evaluation have been described: Perceptual Evaluation of Speech Quality (PESQ) and Perceptual Evaluation of Audio Quality (PEAQ).

In Chapter 3 the term *pseudo analogue signal* has been defined as time discrete, quasi-continuous amplitude signal. A survey of state-of-the-art pseudo analogue transmission schemes has revealed two basic principles. T. Miki et. al. use ADPCM (Adaptive Differential Pulse Code Modulation) in combination with a multi-level digital modulation scheme such as M -ary Differential Phase Shift Keying (M-DPSK) or M -ary Frequency Shift Keying (M-FSK) to transmit the residual which is quantized with large M [44]. N. Phamdo and U. Mittal rather transmit the quantization error of a digital encoder over the pseudo analogue channel with either linear, or nonlinear analogue coding, see Figure 3.3. Their system is called *Hybrid Digital-Analog Coding (HDA)* [19, 48].

7.1 MAD Transmission

The novel MAD transmission scheme is a generalization of Miki's concept. It is based on Linear Prediction with digital transmission of the LP parameters. The unquantized residual of the linear predictor is normalized to unit average power and transmitted as pseudo analogue samples. Normalized samples $r_n(k)$ and coded bits b_c (digital parameters after channel coding) are transmitted sequentially over the baseband channel. The general concept of MAD transmission is shown in Figure 3.1.

In the signal space, Binary Phase Shift Keying (BPSK) is considered for transmission of the digital information. The unquantized (which means substantial savings in terms of complexity), quasi-continuous residual r_n is transmitted with Pulse Amplitude Modulation (PAM). To prevent inter-symbol interference, the multiplexed analogue and digital pulses are shaped with the same RRC filter (roll-off factor $\alpha = 0.5$). At the receiver, the decoded signal is affected by channel noise, but not by quantization noise. For the combined signal, the required two-sided low pass bandwidth, which is relevant for band pass transmission, equals

$$\begin{aligned} B' &= B'_{ana} + B'_{digi_{BPSK}} \\ &= (1 + \alpha) \cdot (R_{ana} + R_{digi}) \\ &= 1.5(R_{ana} + R_{digi}) \end{aligned} \tag{7.1}$$

with the analogue sample rate R_{ana} and the digital bit rate R_{digi} . Figure 3.11 shows the gain in bandwidth and speech quality using MAD speech transmission compared to transmission over the same Additive White Gaussian Noise (AWGN) channel using the Adaptive Multi Rate (AMR) speech codec mode 12.2 kbit/s. The computational complexity of wideband MAD transmission differs only with respect to the sampling frequency from that of narrowband MAD. Both are substantially below the one of a narrowband CELP codec. MAD transmission of audio signals yields comparable results. Competitive results are also achieved in transmission scenarios which include Rayleigh fading (flat fading).

7.2 Modulation for MAD Transmission

In Chapter 4 different modulation schemes are suggested for transmission of the digital bits on one hand, and for transmission of the pseudo analogue samples on the other hand. While the simple concept used to evaluate the general behaviour of MAD transmission, BPSK for the bits and PAM for the samples, already proved the potential of the novel scheme, more sophisticated modulation can further improve the output quality and reduce the channel bandwidth required for MAD transmission. Assuming 2-dimensional modulation, Quadrature Phase Shift keying (QPSK) is evaluated for the digital path, reducing the required bandwidth by a factor of two. Further possible schemes like QAM are well covered in literature and, thus,

not detailed further in this contribution. In the pseudo analogue path Archimedes Spiral Mapping (ASM, see Figure 4.15) is introduced. While ASM does not offer a further reduction of use of channel bandwidth, it improves the Signal-to-Noise Ratio (SNR) at the receiver side - comparable to digital channel coding - without adding considerable complexity. The idea of ASM is visualized in Figures 4.23 to 4.25 while its effect is analyzed in Figures 4.26 and 4.27.

7.3 Information Theoretic Evaluation

Chapter 5 deals with the theoretic evaluation of the MAD concept and the derivation of bounds which define the minimum distortion that can be achieved at the receiver side. First the elementary concepts of information theory are reviewed in terms of capacities

$$C_{Gauss} = B \cdot \text{ld}\left(1 + \frac{S}{N}\right) \quad (7.2)$$

of an AWGN channel with Gaussian input (Shannon Capacity) and

$$C_{bin} = 2B \cdot \left[1 - \frac{1}{\sqrt{2\pi}\sigma_n} \cdot \int_{-\infty}^{\infty} e^{-\frac{(y+1)^2}{2\pi\sigma_n^2}} \cdot \text{ld}\left(1 + e^{\frac{2y}{\sigma_n^2}}\right) dy \right] \quad (7.3)$$

of an AWGN channel with binary input. B denotes the channel bandwidth, $\frac{S}{N}$ the signal-to-noise ratio. The property that $C_{Gauss} \geq C_{bin}$ for all channel conditions is shown to be the main reason for the effectiveness of MAD transmission. If a channel is split into two subchannels, digital and pseudo analogue, the capacities of each of the subchannels add up:

$$C_{tot,p} = C_{digi} + C_{ana} = C_{bin} + C_{Gauss} \quad (7.4)$$

From this we find:

- a) For exclusive transmission of pseudo analogue or digital information only C_{ana} will be greater than C_{digi} for all channel conditions.
- b) With rising E_s/N_0 , C_{ana} will rise faster than C_{digi} .

Next, Rate Distortion Theory is revisited to obtain the transmission rate required to remain below a defined distortion. For a zero-mean, memoryless Gaussian source with variance σ_x^2 and MSE distortion the Rate-Distortion Function (RDF) is considered. Equating the channel capacity C and rate distortion $R(D)$ the Optimum Performance Theoretically Attainable (OPTA) can be determined:

$$\frac{1}{2} \cdot \text{ld}\frac{S}{D} = B \cdot T \cdot \text{ld}\left(1 + \frac{S}{N}\right) \quad (7.5)$$

Based on a simplified transmission model (see Figure 5.9) and taking the work of Kleijn and Ozerov [12] into account, who have shown that the rate needed for an LP filter is determined by the autoregressive model only and that it is independent from the gross rate available, the part of the total capacity which is available for the residual signal in both cases, purely digital LP transmission or MAD transmission, is derived and illustrated in Section 5.4.3. Using a half-whitening LP filter [13] allows for explicit calculation of the prediction gain G_{SNR} which raises the output signal-to-noise ratio after quantization of the residual signal (digital transmission) or after adding noise on the channel (pseudo analogue transmission). Combining G_{SNR} and OPTA on the subchannel used for the residual signal yields the bounds limiting the quality of MAD transmission.

7.4 Implementation

Considering different modulation schemes for transmitting the residual signal, PAM is the optimum choice for 1-dimensional transmission scenarios. Looking at 2-dimensional transmission, ASM is analysed and the possible distortions are split into two kinds: weak noise which corresponds to a displacement on the true branch of the spiral, and strong noise which represents the probability of decoding a received sample to a lower or higher branch (see Figure 5.34). From the combination of these, the theoretical performance of ASM is derived in dependence of the tightness of the spiral. The optimum tightness which depends on the channel SNR is calculated and the optimum signal distribution within the locus of the spiral is found. The performance of MAD transmission with optimized ASM is compared to the performance of wideband MAD in Figures 5.41 and 5.42. Optimizing MAD allows to improve the wideband PESQ speech quality measure by up to 0.5, or to lower E_s/N_0 by up to 8 dB without compromising the speech quality.

Regarding complexity, MAD particularly benefits from the fact that a quantization of the residual signal is not necessary. Compared to the Adaptive Multi Rate (AMR) standard, the complexity of MAD is only about 1/4 (narrowband) or even 1/10 (wideband) of that of the respective AMR standard while the output quality of MAD is even or (in most channel conditions) better in terms of perceptual quality.

7.5 Application

The final chapter discusses application concepts and further prospects for MAD transmission. Any scenario, where lightweight, bandwidth efficient transmission is required to deliver a high quality signal to or from a (possibly mobile, battery powered) device is a target for MAD transmission. Examples are wireless microphones, headsets, or hearing aids with wireless audio input. Being a pure *transmission* concept, MAD does not compete mp3 or AAC audio codecs, which primarily aim at the *storage* of audio signals.

The intention of this contribution is to introduce the MAD concept and to establish a theoretical framework for its evaluation in comparison to traditional digital transmission schemes. This exploration gives rise to further aspects. These include, e.g., pseudo analogue channel coding, extended channel adaptation or secured MAD transmission.

The basic MAD transmission scheme already has proven to be a very competitive, low complexity, low bandwidth audio transmission scheme allowing for high quality transmission with seamless degradation towards worse channel conditions and seamless refinement towards improving channels.

A

The Spectral Flatness Measure

The Spectral Flatness Measure (SFM) was first defined by Makhoul and Wolf [75] as

$$SF(s) = \frac{e^{\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln \phi_{ss}(e^{j\Omega}) d\Omega\right)}}{\frac{1}{2\pi} \int_{-\pi}^{\pi} \phi_{ss}(e^{j\Omega}) d\Omega} \quad (\text{A.1})$$

with $\phi_{ss}(e^{j\Omega})$ the PSD¹ of a zero-mean stationary random signal $s(k)$. With this definition, the SFM resembles the ratio of the geometric to the arithmetic mean of the PSD. It can take values of

$$0 \leq SF \leq 1 \quad (\text{A.2})$$

where an SFM of $SF = 1$ describes a white noise signal, see [74].

The capability of the SFM to describe how white a noise signal is (or how flat the spectrum of the signal is), makes it especially interesting for the analysis of the whitening process of an LP analysis filter (compare also Section 5.4.3). Markel and Gray have shown in [37, 74] that the relation of the SFM of the residual signal $SF^2(r)$ to that of the input signal $SF(s)$ follows

$$SF(r) = SF(s) \cdot G_p(N) \quad (\text{A.3})$$

with $G_p(N)$ the prediction gain of an LP filter of order N . If maximum whitening is achieved with $N \rightarrow \infty$, the residual becomes a white process with $SF(r) = 1$ and thus the largest possible prediction gain equals the inverse of the SFM:

$$G_p^{max} = \lim_{N \rightarrow \infty} G_p(N) = SF^{-1}(s) \quad (\text{A.4})$$

¹Power Spectral Density

B

The Complementary Error Function

The so-called error function $\text{erf}(x)$ is twice the integral of a Gaussian distribution with zero mean and variance $\frac{1}{2}$. It is defined as:

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (\text{B.1})$$

Figure B.1 shows $\text{erf}(x)$.

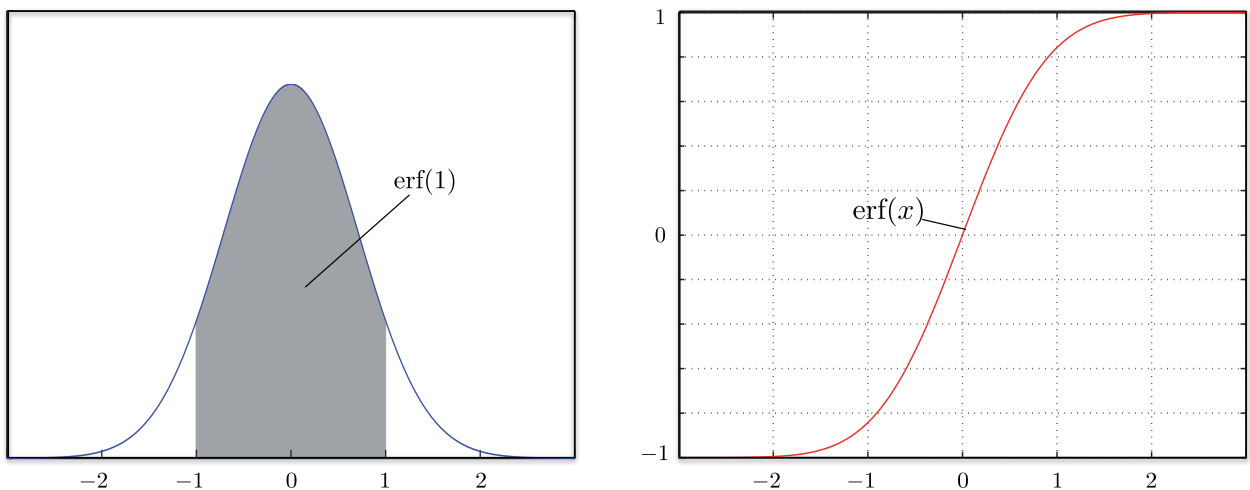


Figure B.1: The error function. The grey area (left) is measured by $\text{erf}(x)$ (right).

$\text{erfc}(x)$ is the complementary error function defined as

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt = 1 - \text{erf}(x). \quad (\text{B.2})$$

The complementary error function is visualized in Figure B.2. Regarding a Gaussian probability density function, $\text{erfc}(x)$ resembles the probability of a probe being greater than the threshold x .

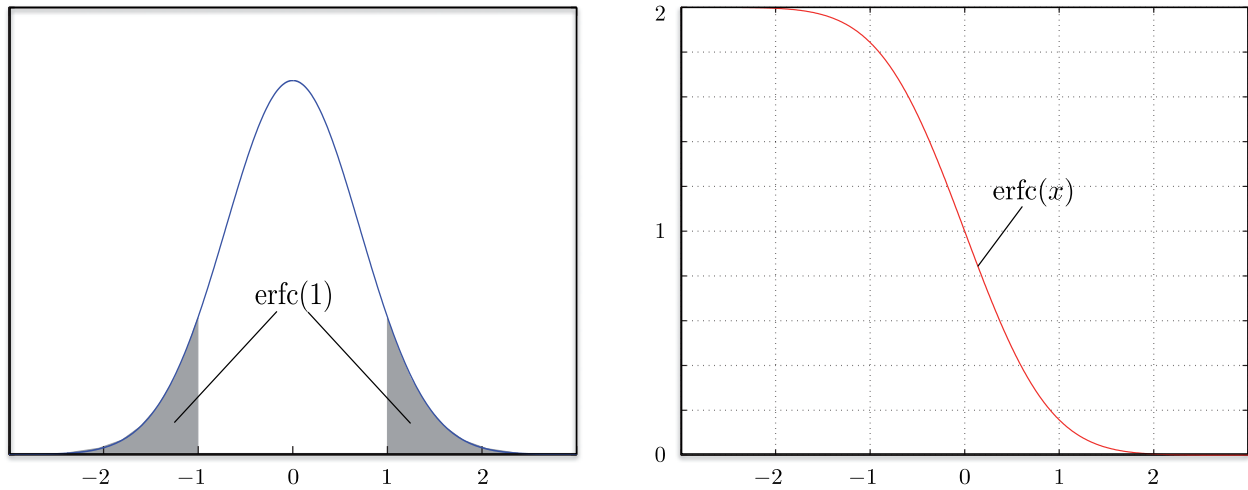


Figure B.2: The complementary error function. The grey area (left) is measured by $\text{erfc}(x)$ (right).

C

The Filter $H_s(z)$

The AR filter $H_s(z)$ has fixed filter coefficients that resemble the LP coefficients of a short sequence of voiced speech:

$$H_s(z) = 1 - A(z) = 1 - \sum_{i=1}^N a_i \cdot z^{-i} = \sum_{i=0}^N a_i \cdot z^{-i}. \quad (\text{C.1})$$

with the coefficients

$$\begin{aligned} a_0 &= 1 \\ a_1 &= -1.76099336051425 \\ a_2 &= 0.750604031626137 \\ a_3 &= 0.180256097286767 \\ a_4 &= -0.323318345333759 \\ a_5 &= 0.0911077080256123 \\ a_6 &= 0.339634337443129 \\ a_7 &= -0.307851250928285 \\ a_8 &= 0.0397709367549065 \\ a_9 &= -0.0954339147246177 \\ a_{10} &= 0.114005517647705 \\ a_{11} &= -0.116059063996878 \\ a_{12} &= 0.0422230850909369 \\ a_{13} &= 0.0763341734216013 \\ a_{14} &= 0.1728099315279 \\ a_{15} &= -0.334353026285037 \\ a_{16} &= 0.150568055017579. \end{aligned} \quad (\text{C.2})$$

D

Deutschsprachige Zusammenfassung

Heutige Systeme zur Codierung und Übertragung von Sprach-, Audio- und Videodaten sind entweder digital oder analog, wobei es in den letzten Jahrzehnten einen starken Trend von analogen zu digitalen Systemen gegeben hat. Digitale Systeme können bei verrauschten Kanälen mittels Kanalcodierung Übertragungsfehler ausgleichen, sind aber in der Regel komplexer als analoge Systeme. Auch wird bei digitalen Systemen die maximal erreichbare Qualität durch das Codierverfahren bestimmt - werden alle Bits fehlerfrei decodiert ist keine Verbesserung mehr möglich ohne das Codierverfahren im Sender zu ändern.

In der hier vorliegenden Arbeit wird ein neuartiges Sprach- und Audioübertragungsschema vorgestellt, welches analoge und digitale Techniken kombiniert um sowohl die Übertragungsbandbreite als auch die Komplexität zu reduzieren und darüber hinaus die Audioqualität des empfangenen Signals zu verbessern. Das neue Verfahren wird als gemischt pseudoanalog-digitale Übertragung (MAD: *Mixed Pseudo Analogue-Digital transmission*) bezeichnet.

D.1 Grundlagen

Nach einer kurzen Einführung in analoge, digitale und gemischte Übertragungsverfahren werden die fundamentalen Konzepte vorgestellt, auf denen in dieser Arbeit aufgebaut wird. Ausgehend von der Definition eines schmalbandigen (Telefonqualität mit einer Audiobandbreite von 300 Hz - 3400 Hz) und eines breitbandigen Signals (Audiobandbreite 50 Hz - 7000 Hz) wird zunächst das Prinzip der linearen Prädiktion (LP) beschrieben, bei der der aktuelle Abtastwert aus einer Linearkombination der letzten N Abtastwerte geschätzt wird. Es folgt eine kurze Vorstellung der gängigen Sprachcodierungsprinzipien RELP (*Residual Excited Linear Prediction*, z.B. im GSM Vollratencodex [99] verwendet) und CELP (*Code Excited Linear Prediction*, heute als verbesserter Vollratencodex im GSM oder auch im UMTS

genutzt [95, 96]). Im Anschluss werden mit Blockcodes und Faltungscodes die verbreitetsten Strategien der Kanalcodierung beschrieben. Schließlich werden instrumentelle Maße zur Bestimmung der Sprachqualität (PESQ: *Perceptual Evaluation of Speech Quality*) und Audioqualität (PEAQ: *Perceptual Evaluation of Audio Quality*) vorgestellt, mit denen in dieser Arbeit die unterschiedlichen Übertragungsverfahren verglichen werden.

D.2 Gemischt pseudoanalog-digitale Übertragung

Ein *pseudoanaloges Signal* wird als Signal mit zeitdiskreten, aber wertekontinuierlichen Abtastwerten definiert. Die Untersuchung aktueller pseudoanaloger Übertragungsschemata zeigt zwei generelle Prinzipien. Miki et. al. nutzen die sogenannte *Adaptive Differentielle Puls Code Modulation* (ADPCM) in Kombination mit einem Multi-Level Modulationsschema - etwa der *M-fachen differentiellen Phasenumtastung* (M-DPSK) oder *M-fachen Frequenzumtastung* (M-FSK) - um das relativ fein quantisierte Restsignal zu übertragen [44]. Dagegen übertragen Phamdo und Mittal in ihrem Konzept den Quantisierungsfehler eines digitalen Encoders über einen pseudoanalogen Kanal, s. Abbildung D.1. Ihr System nennt sich *Hybrid Digital-Analoge Codierung* (HDA) [19, 48].

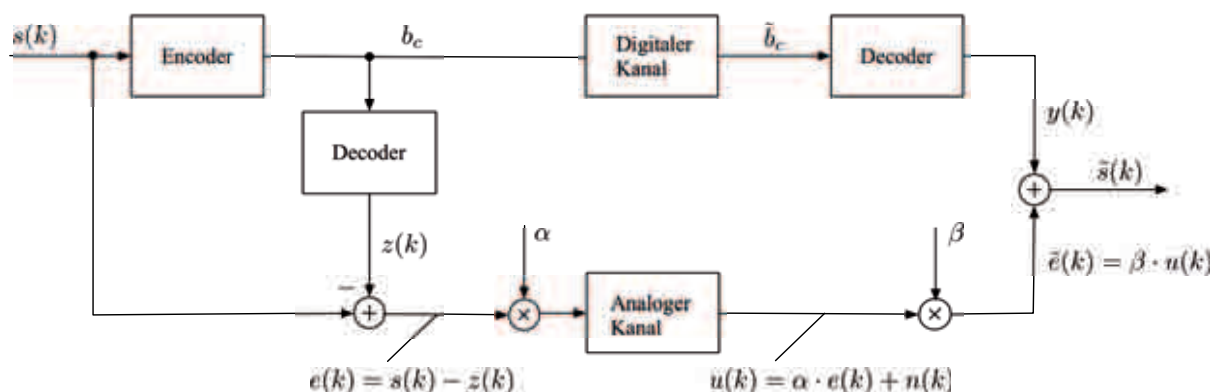


Figure D.1: *Hybrid Digital-Analoge Codierung* nach [19, 48] mit analoger Übertragung des digitalen Quantisierungsfehlers.

Das neue gemischt pseudoanalog-digitale Übertragungsschema MAD ist eine Generalisierung von Mikis Konzept [44]. Es basiert auf *linearer Prädiktion* (LP) und digitaler Übertragung der LP Parameter. Das unquantisierte Restsignal $r(k)$ nach LP Filterung wird auf eine mittlere Leistung von 1 normiert und in Form von pseudoanalogen Abtastwerten übertragen. Normierte Abtastwerte $r_n(k)$ und codierte Bits b_c (digitale Parameter nach Kanalcodierung) werden sequentiell über den Basisbandkanal übertragen. Das generelle Konzept der gemischt pseudoanalog-digitalen Übertragung MAD ist in Abbildung D.2 gezeigt.

Im Signalraum wird zunächst die *binäre Phasenumtastung* (BPSK: *Binary Phase Shift Keying*) zur Übertragung der digitalen Information betrachtet. Das unquantisierte quasi-kontinuierliche Restsignal $r_n(k)$ wird mit *Puls Amplituden Modulation*

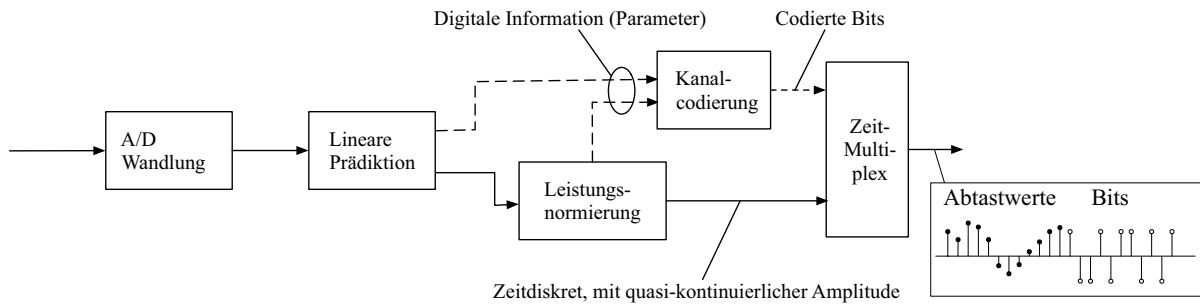


Figure D.2: Prinzip der gemischt pseudoanalog-digitalen Übertragung MAD.

(PAM) übertragen. Durch das Weglassen der Quantisierung kann - etwa im Hinblick auf die aufwändigen Codebuchsuchen in CELP Codecs - deutlich an Komplexität gespart werden. Um Inter-Symbol Interferenzen zu vermeiden, wird das Multiplexsignal aus pseudoanalogen Samples und digitalen Pulsen mit einem gemeinsamen *Root-Raised-Cosine Filter* geformt.

Am Empfänger ist das decodierte Restsignal durch Kanalrauschen verzerrt, dafür tritt kein Quantisierungsrauschen auf. Die zur Übertragung benötigte zweiseitige Tiefpassbandbreite B' , die für die Bandpassübertragung relevant ist, kann gemäß

$$\begin{aligned}
 B' &= B'_{ana} + B'_{digi_{BPSK}} \\
 &= (1 + \alpha) \cdot (R_{ana} + R_{digi}) \\
 &= 1.5(R_{ana} + R_{digi})
 \end{aligned}
 \tag{D.1}$$

angegeben werden. Hierbei sind R_{ana} die analoge Abtastrate, R_{digi} die digitale Bitrate und $\alpha = 0.5$ der *Roll-off-Faktor* des *Root-Raised-Cosine Filters*. Abbildung D.3 zeigt den Gewinn - bezogen auf Übertragungsbandbreite und Sprachqualität - für gemischt pseudoanalog-digitale Schmalband- und Breitband-Übertragung im Vergleich zum im GSM verwendeten *Adaptiven Multiratencodec* (AMR) im 12.2 kbit/s Modus, der incl. Kanalcodierung eine Rate von 22.8 kbit/s hat, also nach Gleichung D.1 auf eine Übertragungsbandbreite von 34.2 kHz kommt. In allen Fällen wird der Übertragungskanal durch einen AWGN-Kanal (*Additive White Gaussian Noise*) modelliert. Der Rechenaufwand zur MAD-Übertragung breitbandiger Audiosignale unterscheidet sich nur durch die höhere Abtastrate von der schmalbandigen MAD-Übertragung. In beiden Fällen liegt die Komplexität mit 3.25 bzw. 3.1 wMOPS (gewichtete MOPS, *Millionen Operationen pro Sekunde*) deutlich unter der von CELP Codecs wie dem AMR (9.65 bis 14.18 wMOPS je nach Modus) oder Breitband-AMR (20.46 bis 31.14 wMOPS), während die Qualität vergleichbar oder besser ist. Ähnlich interessante Ergebnisse erreicht MAD-Übertragung auch in Szenarios mit flachem Fading (*Rayleigh Fading*).

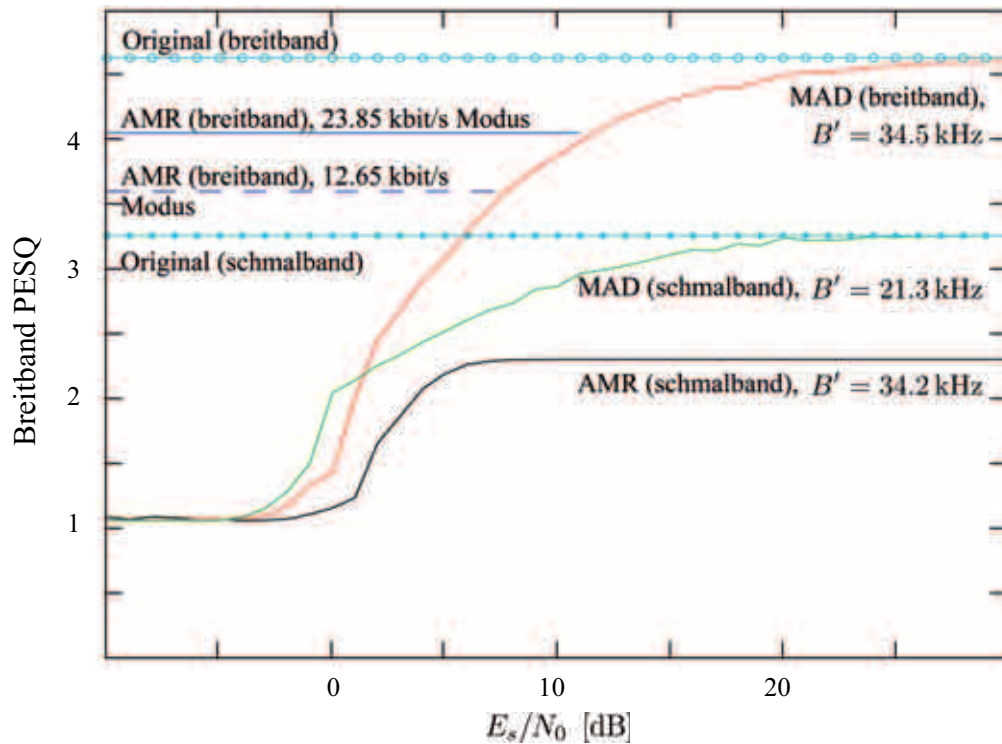


Figure D.3: Vergleich von AMR und MAD-Codierung. E_s : Energie pro codiertem Bit oder mittlere Energie pro Sample. N_0 : Rauschleistung. PESQ bestimmt die Sprachqualität zwischen 1 (sehr schlecht) und 5 (exzellent).

D.3 Modulation bei MAD-Übertragung

Zur Übertragung der digitalen Bits auf der einen und der pseudoanalogen Samples auf der anderen Seite, werden verschiedene Modulationsschemata vorgestellt. Die zur Analyse des prinzipiellen Verhaltens vorgeschlagenen BPSK für die Bits und PAM für die Samples zeigen bereits das Potential des neuen Übertragungsschemas bei reeller Übertragung. Zweidimensionale Verfahren in der komplexen Ebene können die Qualität der Übertragung noch weiter steigern und die Kanalbandbreite noch stärker begrenzen. Im digitalen Pfad wird in dieser Arbeit als zweidimensionale Modulation die *Quadratur-Phasenumtastung* (QPSK) genutzt, die die benötigte Kanalbandbreite bereits um einen Faktor 2 reduziert. Höherwertige Schemata, wie etwa die *Quadratur-Amplituden-Modulation* (QAM), sind in der Literatur hinlänglich vertreten und werden hier nicht weiter untersucht.

Erstmals für den pseudoanalogen Pfad wird das Mapping auf eine Archimedes spirale (ASM: *Archimedes Spiral Mapping*, s. Abbildung D.4) untersucht, bei dem die Sendepunkte x_{ana} nach

$$x_{ana}(k) = r_n(k) \cdot \exp(j \frac{|r_n(k)|}{c}). \quad (\text{D.2})$$

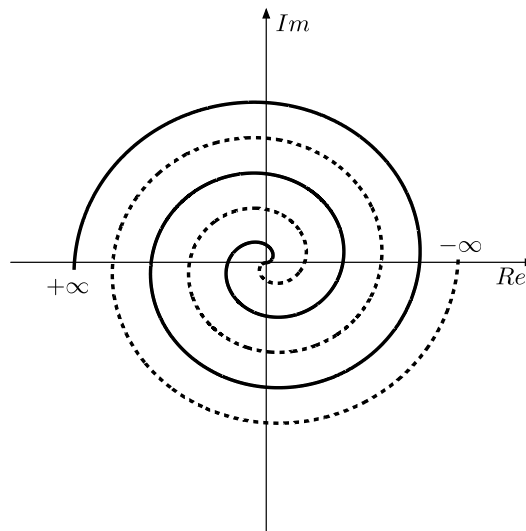


Figure D.4: Archimedes Spiral Mapping.

aus den normierten Abtastwerten r_n und einem Dichte-Faktor c berechnet werden. Die Information über den Abtastwert wird also nicht nur wie bei der PAM in der Amplitude, sondern zusätzlich in der Phase übertragen, wodurch sich ähnlich der Kanalcodierung im digitalen Fall ein Gewinn erzielen lässt. Für $c \rightarrow \infty$ geht ASM in die PAM über. ASM bietet zwar keine weitere Reduktion der nötigen Übertragungsbandbreite, das Signal-Rauschleistungsverhältnis (SNR: *Signal-to-Noise Ratio*) am Empfänger kann aber verbessert werden ohne viel zusätzliche Komplexität zu benötigen. Die Idee des ASM ist folgende: Wird der Graph des Signalraums zur Übertragung gestreckt, reduziert sich bei der Rücküberführung am Empfänger der effektive Anteil des Rauschens relativ zum Signal. Ist die Spirale aber zu dicht, kann es durch das Rauschen zur Decodierung eines falschen Zweiges kommen, was einem Vorzeichenfehler entspricht. Der Effekt des Rauschens wird in Abbildung D.5 verdeutlicht.

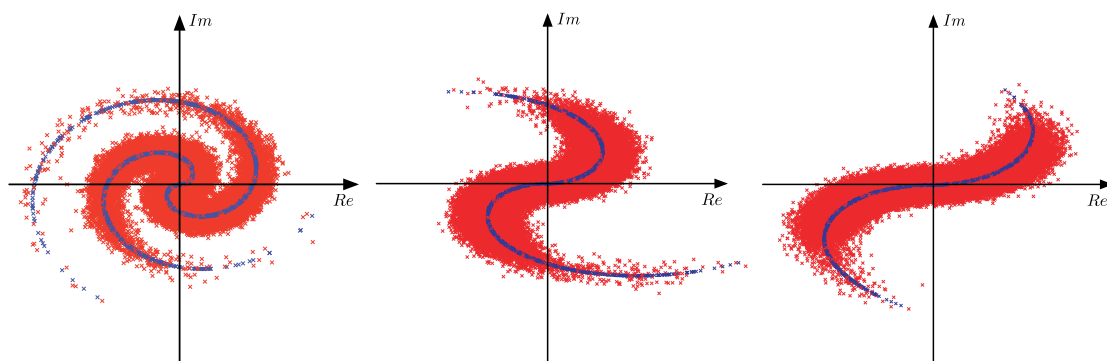


Figure D.5: ASM bei einem SNR von 10 dB und von links nach rechts $c = 0.5$, $c = 1.5$, bzw. $c = 2.5$.

Abbildung D.6 zeigt für verschiedene Dichtefaktoren c das SNR des Ausgangssignals. Es ist deutlich zu sehen, wie eine immer dichtere Spirale bei besseren Kanälen deutliche Gewinne gegenüber der PAM bewirkt, wie eine zu dichte Spirale aber bei geringeren E_s/N_0 die Ausgangsqualität stark verringert.

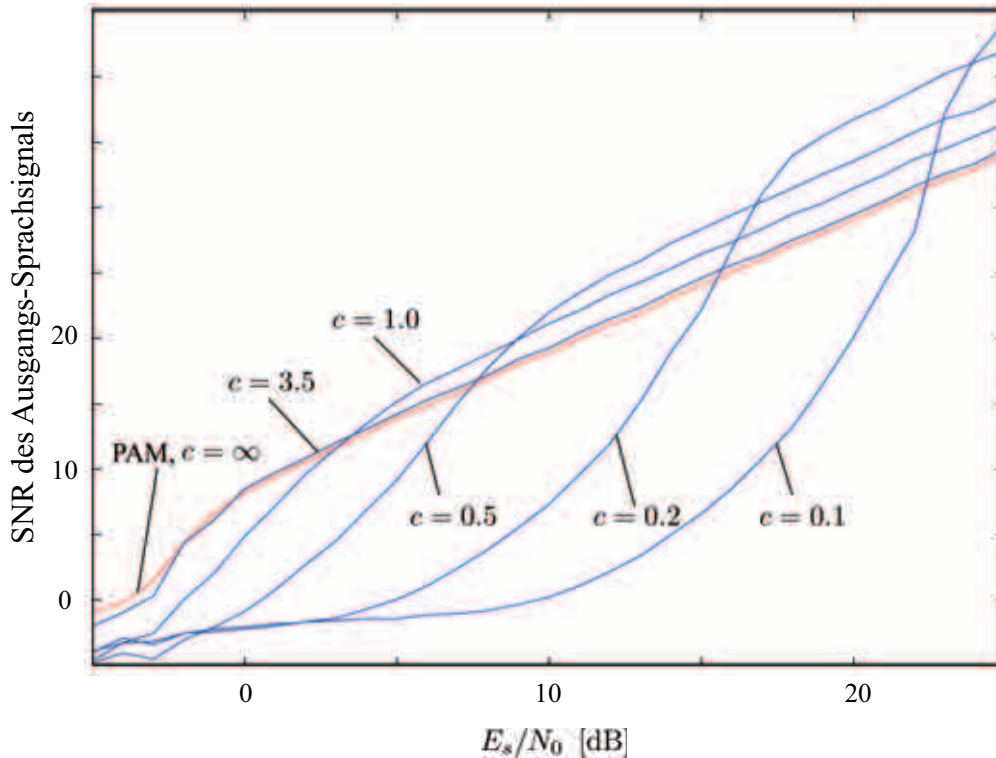


Figure D.6: Effekt des Mappings auf die Archimedes Spirale im SNR Bereich.

D.4 Informationstheoretische Betrachtung

Im Rahmen dieser Arbeit wird das MAD-Konzept informationstheoretisch behandelt und es werden Grenzen abgeleitet, die die minimale Verzerrung definieren, welche am Empfänger erreicht werden kann. Hierzu werden zunächst die elementaren Konzepte der Informationstheorie vorgestellt, insbesondere die Kanalkapazitäten

$$C_{Gauss} = B \cdot \text{ld}\left(1 + \frac{S}{N}\right) \quad (\text{D.3})$$

eines AWGN Kanals mit einem Gauss'schen Eingangssignal (Shannon Kapazität) und

$$C_{bin} = 2B \cdot \left[1 - \frac{1}{\sqrt{2\pi}\sigma_n} \cdot \int_{-\infty}^{\infty} e^{-\frac{(y+1)^2}{2\pi\sigma_n^2}} \cdot \text{ld}\left(1 + e^{\frac{2y}{\sigma_n^2}}\right) dy \right] \quad (\text{D.4})$$

eines AWGN Kanals mit binärem Eingangssignal. B gibt die einseitige Kanalbandbreite an, S/N das Rauschleistungsverhältnis (SNR). Es wird gezeigt, dass das Verhältnis

$$C_{Gauss} \geq C_{bin}, \quad (D.5)$$

welches für alle Kanalzustände gilt, der entscheidende Grund für die Effektivität der gemischt pseudoanalog-digitalen Übertragung ist, da das normierte pseudoanaloge Restsignal annähernd gaussverteilt und damit ideal an den AWGN-Kanal angepasst ist. Informationstheoretisch betrachtet wird der Übertragungskanal in zwei Teilkanäle aufgeteilt, digital und pseudoanalog, deren Kapazitäten sich addieren:

$$C_{tot,p} = C_{digi} + C_{ana} = C_{bin} + C_{Gauss}. \quad (D.6)$$

Hieraus folgt, dass bei exklusiver Übertragung von pseudoanaloger oder digitaler Information C_{ana} stets größer oder gleich C_{digi} ist. Mit der Ableitung von Gleichung (D.5) wird zusätzlich gezeigt, dass mit steigendem E_s/N_0 die Kapazität C_{ana} des pseudoanalogen Pfades stets schneller steigt als C_{digi} .

Nach einer kurzen Einführung wird mit den Mitteln der *Rate-Distortion-Theorie* die minimal notwendige Rate bestimmt, die es erlaubt bei einer Übertragung unter einer definierten Verzerrung zu bleiben. Betrachtet wird die *Rate-Distortion-Funktion* (RDF) einer mittelwertfreien, gedächtnislosen, gaussverteilten Quelle mit Varianz σ_x^2 bezüglich der mittleren quadratischen Verzerrung (MSE: *Mean Square Error Distortion*). Werden die Kanalkapazität C und die *Rate Distortion* $R(D)$ gleichgesetzt, kann die theoretisch optimal erzielbare Güte (OPTA: *Optimum Performance Theoretically Attainable*) bestimmt werden:

$$\frac{1}{2} \cdot \text{ld} \frac{S}{D} = B \cdot T \cdot \text{ld} \left(1 + \frac{S}{N} \right). \quad (D.7)$$

Die weitere Betrachtung basiert auf einem vereinfachten Modell der Übertragung gemäß Abbildung D.7 und berücksichtigt die Arbeit von Kleijn und Ozerov [12]. Hier wurde gezeigt, dass die Rate, welche zur Übertragung der Koeffizienten des LP Filters (und damit der spektralen Einhüllenden) benötigt wird, nur vom autoregressiven Modell abhängt und unabhängig von der insgesamt zur Verfügung stehenden Übertragungsrate ist.

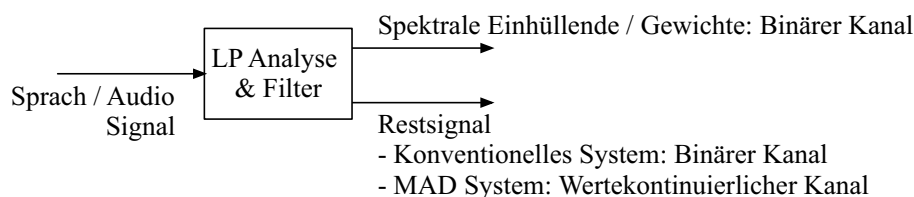


Figure D.7: Vereinfachtes Modell der zeitdiskreten, linear prädiktiven Übertragung von Sprach- und Audiosignalen.

Hieraus wird der Teil der Gesamtrate bestimmt, der zur Übertragung des Restsignals verwendet werden kann - unabhängig davon, ob diese Übertragung digital oder pseudoanalog erfolgt.

Wird das von Krüger in [13] vorgestellte LP Filter genutzt, kann der Prädiktionsgewinn G_{SNR} , um den das Ausgangs-SNR nach Quantisierung des Restsignals (digitale Übertragung) bzw. nach Hinzufügen von Kanalrauschen (pseudoanaloge Übertragung) durch die Prädiktionsfilterung angehoben wird, explizit berechnet werden. Die Kombination von G_{SNR} und OPTA auf dem Teilkanal, der für die Übertragung des Restsignals genutzt wird, ergibt die hier vorgestellten Grenzen der Qualität der gemischt pseudoanalog-digitalen Übertragung. Dies wird in Abbildung D.8 verdeutlicht: Zunächst sind für den Übertragungskanal des Restsignals jeweils OPTA für ein binäres und ein reellwertiges, gaussverteiltes Eingangssignal bei komplexer Übertragung und OPTA für ein reellwertiges, gaussverteiltes Eingangssignal bei reeller Übertragung dargestellt. Die letztgenannte Kurve ist dann um G_{SNR} angehoben, was einer Grenze für die Qualität der MAD Übertragung im Sprachbereich entspricht (OPTA_MAD). Schließlich in rot dargestellt ist die simulierte MAD Übertragung - wobei die digitalen Parameter als fehlerfrei angenommen und nur die Übertragung des Restsignals betrachtet wurde.

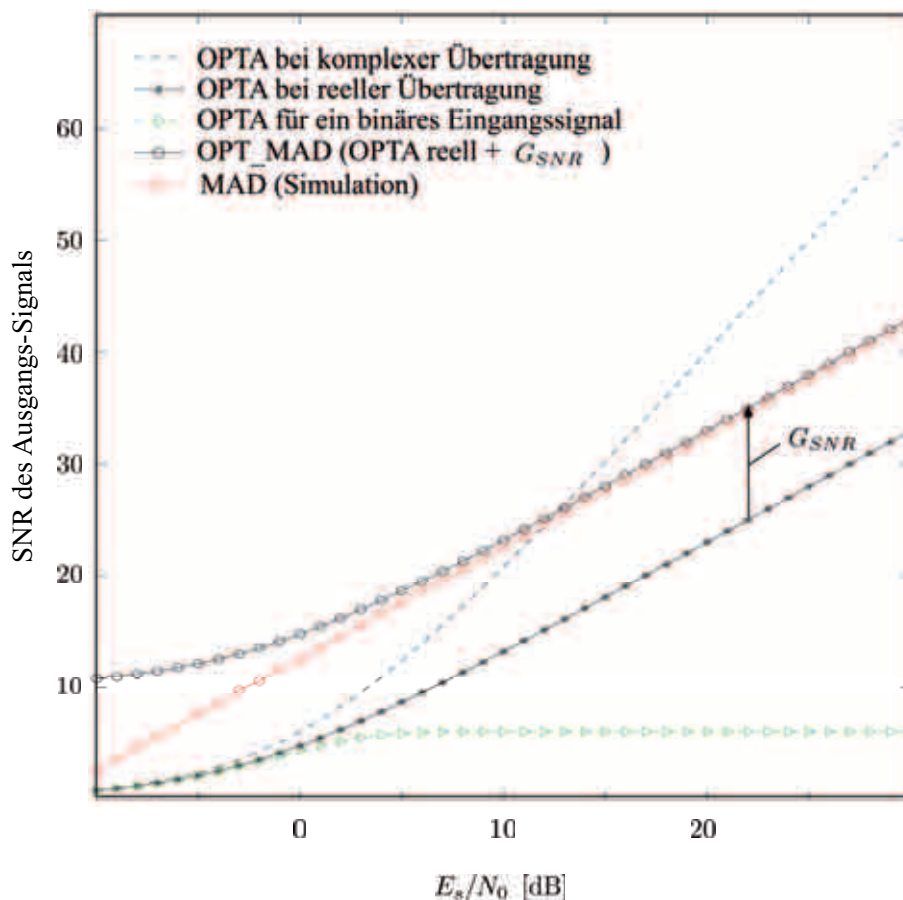


Figure D.8: Grenze für die Güte der reellen MAD-Übertragung.

Bei Betrachtung verschiedener Modulationsschemata zur Übertragung des pseudoanalogen Restsignals, ist *Puls Amplituden Modulation* (PAM) die optimale Wahl für eine reelle Übertragung. Für eine komplexe Übertragung wird das Mapping auf die Archimedesspirale analysiert. Hier können zwei unterschiedliche Arten von Verzerrungen auftreten. Schwaches Rauschen $\bar{\varepsilon}_{LowNoise}^2$, welches einer Verschiebung des decodierten Signalpunktes gegenüber dem gesendeten Signalpunkt auf dem selben Zweig der Spirale entspricht und starkes Rauschen $\bar{\varepsilon}_{StrongNoise}^2$, welches auftritt, wenn der decodierte Signalpunkt auf einem benachbarten Zweig der Spirale liegt, wodurch ein falsches Vorzeichen decodiert wird. Die Wahrscheinlichkeit für diese Art Rauschen ist in Abbildung D.9 dargestellt. Das Rauschen ist jeweils abhängig von der Stärke des Kanalrauschens und vom Dichte-Faktor c der Spirale (also dem Abstand der einzelnen Zweige).

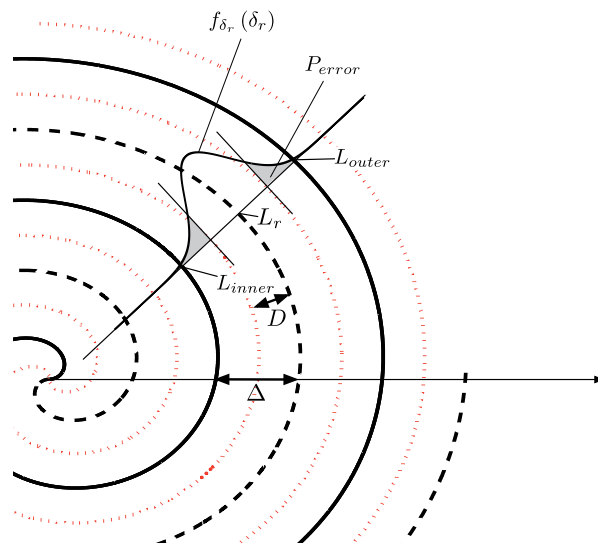


Figure D.9: Wahrscheinlichkeit für das Auftreten von starkem Rauschen, also der Decodierung eines Signalpunktes auf der durchgezogenen Linie, wenn ein Punkt auf der gestrichelten Linie (komplementäres Vorzeichen) gesendet wurde.

In der vorliegenden Arbeit wurden erstmals Ausdrücke für die beiden Rauscharten bei ASM nach Gleichung D.2 hergeleitet. Diese sind für das schwache Rauschen

$$\bar{\varepsilon}_{LowNoise}^2 = \sigma_n^2 \int_{-\infty}^{\infty} \frac{1}{1 + \frac{r_n^2}{c^2}} \frac{1}{\sigma_{r_n} \sqrt{2\pi}} e^{-\frac{r_n^2}{2\sigma_{r_n}^2}} dr_n \quad (D.8)$$

und für das starke Rauschen

$$\bar{\varepsilon}_{StrongNoise}^2 = \frac{1}{2} \left(1 - \operatorname{erf} \left(\frac{\pi \cdot c}{2\sqrt{2}\sigma_n} \right) \right) \cdot \int_{-\infty}^{\infty} (8r_n^2 + 2\pi^2 \cdot c^2) \frac{1}{\sigma_{r_n} \sqrt{2\pi}} e^{-\frac{r_n^2}{2\sigma_{r_n}^2}} dr_n. \quad (D.9)$$

Das Verhältnis von Signal- zu Verzerrungsleistung (SDR: *Signal-Distortion-Ratio*) kann nach

$$SDR = \frac{\sigma_{r_n}^2}{\bar{\varepsilon}_{LowNoise}^2 + \bar{\varepsilon}_{StrongNoise}^2} \quad (D.10)$$

berechnet werden. Hieraus ergeben sich die optimalen Dichte-Faktoren c in Abhängigkeit vom Kanal-SNR. Diese können im Falle von rückgekoppelter Kanalinformation für eine optimierte Übertragung genutzt werden.

Abbildung D.10 zeigt die Qualität von breitbandiger MAD-Übertragung wenn das so optimierte ASM zur Übertragung des Restsignals verwendet wird. Die rote Vergleichskurve stammt aus Abbildung D.3, sie stellt MAD Breitband mit BPSK und PAM dar.

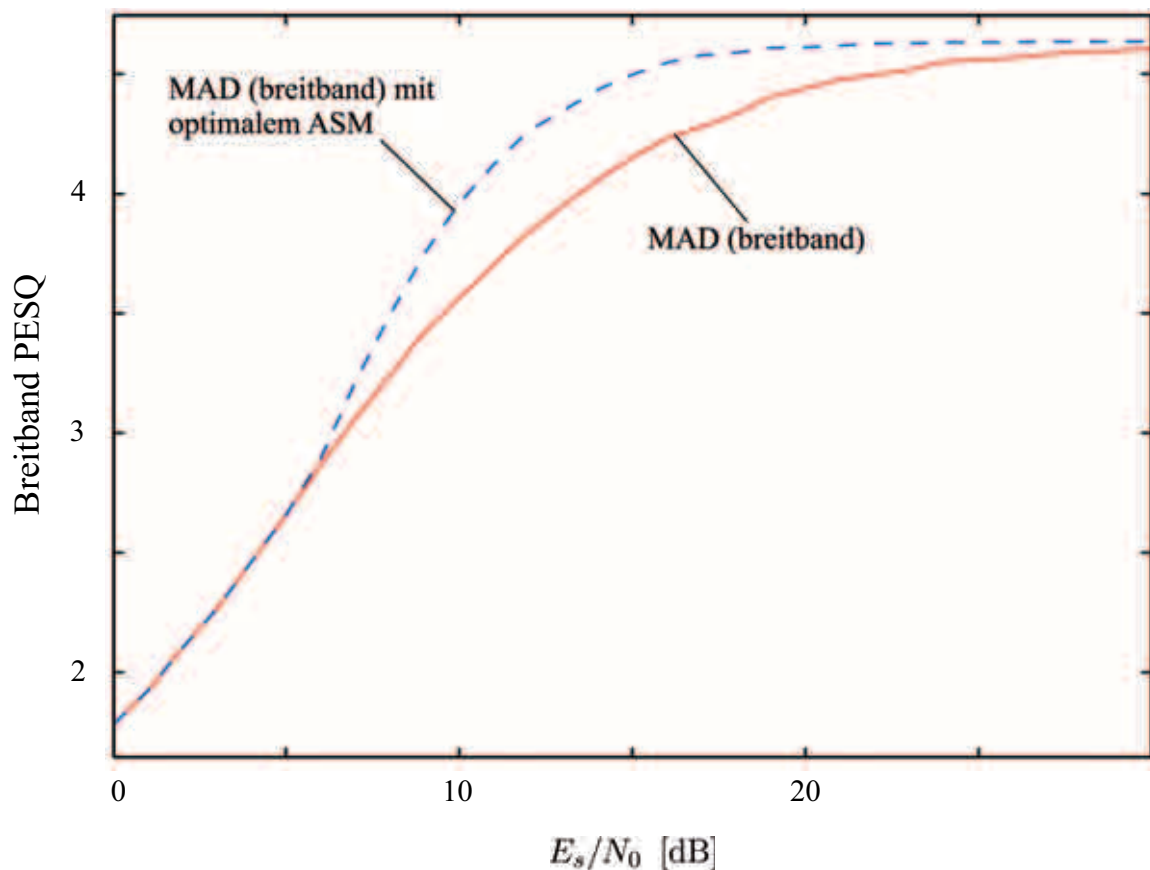


Figure D.10: Perzeptuelle Qualität von MAD mit optimiertem ASM.

In Bezug auf die Komplexität profitiert das MAD-Übertragungsschema hauptsächlich von der entfallenen Quantisierung des Restsignals. Der Rechenaufwand der MAD-Übertragung beträgt nur etwa 1/4 (schmalband) und im breitbandigen Fall sogar nur 1/10 desjenigen, des entsprechenden AMR Standards, während die Ausgabequalität von MAD für die meisten Kanalzustände besser oder zumindest gleich ist.

D.5 Anwendung

Zum Abschluss der Arbeit werden Anwendungskonzepte für das pseudoanalog-digitale Übertragungsverfahren vorgestellt und weitere Ausblicke gegeben. Immer dann, wenn wenig rechenintensive, bandbreiteneffiziente Übertragungsverfahren benötigt werden, um mit einem möglicherweise mobilen und batteriebetriebenen Endgerät eine hohe Qualität darstellen zu können, kann MAD eingesetzt werden. Beispielfür sind drahtlose Mikrofone, Kopfhörer oder Hörgeräte mit drahtloser Audioanbindung.

Ziel dieser Arbeit war, das Konzept der gemischt pseudoanalog-digitalen Übertragung einzuführen und theoretisch im Vergleich zu rein digitalen Übertragungsschemata zu analysieren. Die Ausarbeitung bietet Anhaltspunkte für weitere Vertiefungen, darunter etwa pseudoanaloge Kanalcodierung, erweiterte Anpassung an den jeweiligen Kanal oder MAD-Übertragung in Verbindung mit kryptographischen Verfahren.

Mit dem grundlegenden MAD-Übertragungsverfahren konnte ein konkurrenzfähiges, komplexitätsarmes, bandbreitensparendes Audio-Übertragungssystem entwickelt werden. Neben einer hochqualitativen Übertragung erlaubt das Konzept darüberhinaus eine nahtlose Skalierung mit der Kanalqualität.

Abbreviations & Principal Symbols

List of Abbreviations

2-WM	2-Way Mapping
AAC	Advanced Audio Coding
ACB	Adaptive Codebook
A/D	Analog-to-Digital
ADPCM	Adaptive Differential Pulse Code Modulation
AES	Advanced Encryption Standard
AM	Amplitude Modulation
AMR	Adaptive Multi-Rate Speech Codec
AMR-NB	Narrowband Adaptive Multi-Rate Speech Codec
AMR-WB	Wideband Adaptive Multi-Rate Speech Codec
AR	AutoRegressive
ASK	Amplitude Shift Keying
ASM	Archimedes Spiral Mapping
AWGN	Additive White Gaussian Noise
BPSK	Binary Phase Shift Keying
CD	Compact Disc
CELP	Code Excited Linear Prediction
CPU	Central Processing Unit
CRC	Cyclic Redundancy Check
CSNR	Channel Signal-to-Noise Ratio
D/A	Digital-to-Analog
DAB	Digital Audio Broadcasting
DCT	Discrete Cosine Transform
DPSK	Differential Phase Shift Keying
DRF	Distortion-Rate Function
DSP	Digital Signal { Processor Processing }

DVB	Digital Video Broadcasting
DVD	Digital Video Disc
DVD	(later) Digital Versatile Disc
EFR	Enhanced Full Rate { Speech Codec }
FCB	Fixed Codebook
FEC	Forward Error Correction
FSK	Frequency Shift Keying
FM	Frequency Modulation
GSM	Groupe Spéciale Mobil
GSM	(later) Global System for Mobile Communications
HDA	Hybrid Digital-Analog
IEEE	Institute of Electrical and Electronics Engineers
IIR	Infinite Impulse Response
IND	Institut für Nachrichtentechnik und Datenverarbeitung
ITU	International Telecommunication Union
LDPC	Low-Density Parity Check
LP	Linear Prediction
LPC	Linear Predictive { Coding Coefficients }
LSB	Least Significant Bit
LTP	Long Term Predictor
MA	Moving Average
MAD	Mixed Pseudo Analogue-Digital {Transmission }
MAD-WB	MAD for wideband signals
M-DPSK	M-ary Differential Phase Shift Keying
MDCT	Modified Discrete Cosine Transformation
M-FSK	M-ary Frequency Shift Keying
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MOS	Mean Opinion Score
MP3	MPEG-1 Audio Layer 3
MPEG	Moving Picture Experts Group
M-PSK	M-ary Phase Shift Keying
MSE	Mean Squared Error
MSVQ	Multi-Stage Vector Quantization
MUX	{ Multiplex Multiplexer }
ODG	Objective Difference Grade
OPTA	Optimum Performance Theoretically Attainable
PAM	Pulse Amplitude Modulation
PCM	Pulse Code Modulation
PDF	Probability Density Function
PEAQ	Perceptual Evaluation of Audio Quality
PESQ	Perceptual Evaluation of Speech Quality
PM	Phase Modulation
PMR	Personal Mobile Radio

PSD	Power Spectral Density
PSK	Phase Shift Keying
QPSK	Quadrature Phase Shift Keying
QPAM	Quadrature Pulse Amplitude Modulation
QAM	Quadrature Amplitude Modulation
R/D	Rate Distortion
RDF	Rate-Distortion Function
REL P	Residual-Excited Linear Prediction
RRC	Root Raised Cosine
RWTH	Rheinisch Westfälische Technische Hochschule
SDG	Subjective Difference Grade
SDR	Signal-to-Distortion Ratio
SFM	Spectral Flatness Measure
SMQ	Split Matrix Quantization
SNR	Signal-to-Noise Ratio
SPSK	Skewed Phase Shift Keying
SVQ	Split Vector Quantization
TETRA	TErrestrial Trunked RAdio
TV	Television
TWC	Theoretical Worst Case
UEP	Unequal Error Protection
VQ	Vector { Quantization Quantizer }
wMOPS	weighted Million Operations Per Second

List of Principal Symbols

The following conventions are used in this thesis: matrices and vectors are written in bold letters, e.g., $\mathbf{s}(k)$, scalars are not bold, e.g., $s(i)$, quantities in the frequency-domain are written in uppercase letters, e.g., $\mathbf{S}(k)$, quantized quantities are labeled with a hat, e.g., $\hat{s}(k)$, and received, noisy, or estimated quantities are labeled with a tilde, e.g., $\tilde{s}(k)$.

a_i	LP filter coefficients
A	Amplitude
$A(z)$	Z-transform of a Linear Predictor
b_c	Coded bits
B	Bandwidth of a channel
B'	Band pass bandwidth of a channel
B_{tot}	Total bandwidth
c	Tightness constant of a spiral
C	Capacity (in <i>bits/second</i>)
C'	Capacity (in <i>bits/sample</i>)

D	{ Distortion Distance }
$D(R)$	Distortion-Rate Function
E_s	Energy per coded bit, average energy per sample
$E\{\cdot\}$	Expectation
f	Frequency
f_c	Carrier frequency
f_{max}	Highest frequency within a signal
f_s	Sampling frequency
g	Gain factor
G_p	Prediction gain
G_{SNR}	Gain yielding an SNR improvement
$H(X)$	Entropy
$H(X, Y)$	Joint entropy
$H(X Y)$	Conditional entropy
$H(z)$	Z-transform of a prediction filter
i	Integer number
$I(X; Y)$	Mutual information
k	Integer number, used as discrete time index
L	Length, Length of a locus
m_k	Message bit
M	Integer number
n	{ Noise integer number }
$n(k)$	Discrete noise
N	{ Noise power integer number }
N_0	Normalized noise power
N_s	Number of samples in a subframe
$p(x)$	Probability mass function
$p_{r_n}(r_n)$	Probability density function of r_n
$P(x)$	Probability
$P(x, y)$	Joint probability
$P(x y)$	Conditional probability
P	Power
$P_{LTP}(z)$	Z-transform of a LTP filter
P_{error}	Probability of error
q	Number of bits used for quantization
Q	Quantizer
$r(k)$	Residual of a discrete-time signal
$r_n(k)$	Normalized residual of a discrete-time signal
R	Rate
$R(D)$	Rate Distortion Function
$R_G(D)$	Rate Distortion Function for a memoryless Gaussian source and MSE distortion
R_{ana}	Analog sample rate
R_{digi}	Digital bit rate

$s(k)$	Discrete-time signal
$\hat{s}_{LP}(k)$	LP estimate of a discrete-time signal
$s_a(t)$	Continuous-time signal
S	Signal power
SF^{-1}	Spectral flatness measure
T	Duration
$u(t)$	Modulated signal
V	Volume
$W(z)$	Transfer function of a weighting filter
$x(t)$	Signal transmitter side
$x_b(t)$	Binary signal
X	Random variable
\mathcal{X}	Alphabet of X
\tilde{x}_{ana}	Complex signal points
\tilde{x}_{digi}	Complex signal points
$y(t)$	Signal receiver side
α	Roll-off factor
α	Noise shaping factor
α_k	Rayleigh fading factor
β	Gain measure
δ	Closest distance from \tilde{x}_a to Archimedes spiral
Δ	Radial distance in the Archimedes Spiral
ε	Error
γ	Noise shaping factor
θ	Model
Θ	Model parameters
φ	Skewing angle
φ_0	Initial carrier phase
$\varphi_{Ar}(r_n)$	Angle in Archimedes Spiral
σ_n^2	Noise power
σ_x^2	Signal power

BIBLIOGRAPHY

- [1] S. ten Brink, "Exploiting the Chain Rule of Mutual Information for the Design of Iterative Decoding Schemes", *Allerton Conference on Communication, Control, and Computing*, pp293-300, Monticello, IL, USA, 10/2001.
- [2] X. Cai, J.W. Modestino, "Bandwidth Expansion Shannon Mapping for Analog Error-Control Coding", *40th Annual Conference on Information Science and Systems*, Princeton University, 3/2006.
- [3] P.A. Floor, T.A. Ramstad, "Noise Analysis for Dimension Expanding Mappings in Source-Channel Coding", *IEEE Workshop on Signal Processing Advances in Wireless Communications*, Cannes, 7/2006.
- [4] P.A. Floor, T.A. Ramstad, "Optimality of Dimension Expanding Shannon-Kotel'nikov Mappings", *ITW*, Lake Tahoe, CA, 6.2007.
- [5] A. Fuldseth, T.A. Ramstad, "Bandwidth Compression for Continuous Amplitude Channels Based on Vector Approximation to a Continuous Subset of the Source Signal Space", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Munich, 4/1997.
- [6] C. Hoelper, P. Vary, "Mixed Pseudo Analogue-Digital Speech Transmission", *7. ITG Fachtagung Sprachkommunikation*, Kiel, Germany, 4/2006.
- [7] C. Hoelper, P. Vary, "Bandwidth-Efficient Mixed Pseudo Analogue-Digital Speech Transmission", *European Signal Processing Conference EUSIPCO*, Florence, Italy, 9/2006.
- [8] C. Hoelper, P. Vary, "Bandwidth-Efficient Mixed Pseudo Analogue-Digital Speech and Audio Transmission", *Int. Workshop On Multimedia Sig. Proc.*, Victoria, BC, Canada, 10/2006.
- [9] C. Hoelper, P. Vary, "A New Modulation Concept For Mixed Pseudo Analogue-Digital Speech and Audio Transmission", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Honolulu, Hawaii, 4/2007.

-
- [10] F. Hu, W. Henkel, "A Geometric Description of the Iterative Least-Squares Decoding of Analog Block Codes", *Turbo-Coding-2006*, Munich, 4/2006.
- [11] K. Järvinen, J. Vainio, P. Kapanen, T. Honkanen, P. Haavisto, R. Salami, C. Laflamme, B. Bessette, J-P. Adoul, "GSM Enhanced Full Rate Speech Codec", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Munich, 4/1997.
- [12] W.B. Kleijn, A. Ozerov, "Rate Distribution between Model and Signal", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, 10/2007.
- [13] H. Krüger, P. Vary, "A Partial Decorrelation Scheme for Improved Predictive Open Loop Quantization with Noise Shaping", *Interspeech 2005*, Lisbon, Portugal, 9/2005.
- [14] H. Krüger, P. Vary, "SCELP: Low Delay Audio Coding with Noise Shaping based on Spherical Vector Quantization", *European Signal Processing Conference EUSIPCO*, Florence, Italy, 9/2006.
- [15] X. Lin, W.C. Wong, C.-E.W. Sundberg, "Performance of Skewed Phase Shift Keying", *IEEE International Conference on Communications; Converging Technologies for Tomorrow's Applications*, Dallas, Texas, 6/1996.
- [16] D.T. Magill, C.K. Un, "Speech residual Encoding by Adaptive Delta Modulation with Hybrid Companding", *Proceedings of The National Electronics Conference*, 10/1974.
- [17] T. Miki, C.-E.W. Sundberg, N. Seshadri, "Pseudo-Analog Speech Transmission in Mobile Radio Communication Systems", *IEEE International Symposium on Information Theory and Its Applications*, Hawaii, 11/1990.
- [18] M. Mura, W. Henkel, L. Cottatellucci, "Iterative least-squares decoding of analog product codes", *Proceedings IEEE International Symposium on Information Theory*, 7/2003.
- [19] N. Phamdo, U. Mittal, "A CELP-Based Hybrid Digital-Analog (HDA) Joint Source-Channel Speech Coder", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Istanbul, 6/2000.
- [20] A.W. Rix, J.G. Beerends, M.P. Hollier, A.P. Hekstra, "Perceptual Evaluation of Speech Quality (PESQ) - A New Method For Speech Quality Assessment of Telephone Networks and Codecs", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, 4/2001.
- [21] M.R. Schroeder, B.S. Atal, "Code Excited Linear Prediction (CELP): High-Quality Speech at Low Bit Rates", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Florida, 4/1985.

- [22] M. Skoglund, N. Phamdo, F. Alajaji, "VQ-Based Hybrid Digital-Analog Joint Source-Channel Coding", *IEEE International Symposium on Information Theory*, Sorrento, Italy, 6/2000.
- [23] V.A. Vaishampayan, S.I.R. Costa, "Curves on a Sphere: Error Control for Continuous Alphabet Sources", *IEEE International Symposium on Information Theory*, Lausanne, Switzerland, 7/2002.
- [24] P. Vary, K. Hellwig, R. Hofmann, R.J. Sluijter, C. Galand, M. Rosso, "Speech Codec for the European Mobile Radio System", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, New York, 4/1988.
- [25] Y. Wang, F. Alajaji, T. Linder, "Hybrid Digital-Analog Coding of Memoryless Gaussian Sources over AWGN Channels with Bandwidth Compression", *IEEE 23rd Biennial Symposium on Communications*, 6/2006.
- [26] K. Watanabe, et al., "High Capacity Land Mobile Communication System in NTT", *GLOBECOM 87*, vol. 21.1, pp. 791-796, Tokyo, Japan, 11/1987.
- [27] B.S. Atal, S.L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave", *Journal of the Acoustical Society of America*, vol. 50, no. 2b, 8/1971.
- [28] A. Barron, T.M. Cover, "Minimum Complexity Density Estimation", *IEEE Transactions on Information Theory*, vol. 37, no. 4, 7/1991.
- [29] T. Berger, D.W. Tufts, "Optimum Pulse Amplitude Modulation Part I: Transmitter-Receiver Design and Bounds from Information Theory", *IEEE Transactions on Information Theory*, vol. IT-13, no. 2, 4/1967.
- [30] B. Chen, G.W. Wornell, "Analog Error Correcting Codes Based on Chaotic Dynamical Systems", *IEEE Transactions on Communications*, vol. 46, 7/1998.
- [31] J.P. Costas, "Coding with Linear Systems", *Proceedings of the I.R.E.*, pp. 1101-1103, 9/1952.
- [32] W.R. Gardner, B.D. Rao, "Theoretical Analysis of the High-Rate Vector Quantization of LPC Parameters", *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 5, 9/1995.
- [33] M. Gastpar, B. Rimoldi, M. Vetterli, "To Code, or Not to Code: Lossy Source-Channel Communication Revisited", *IEEE Transactions on Information Theory*, vol. 49, no. 5, 5/2003.
- [34] B. Geiser, P. Jax, H. Taddei, S. Schandl, M. Gartner, C. Guillaumé, S. Ragot, "Bandwidth Extension for Hierarchical Speech and Audio Coding in ITU-T Rec. G.729.1", *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 8, 11/2007.

- [35] T.J. Goblick, "Theoretical Limitations on the Transmission of Data From Analog Sources", *IEEE Transactions on Information Theory*, vol IT-11, no. 4, 10/1965.
- [36] L.M. Goodman, P.R. Drouilhet, "Asymptotically optimum pre-emphasis and de-emphasis networks for sampling and quantizing", *Proceedings of the IEEE*, vol 54, 5/1966.
- [37] A.H. Gray, J.D. Markel, "A Spectral Flatness Measure for Studying the Autocorrelation Method of Linear Prediction of Speech Analysis" *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol ASSP-22, no.3, 6/1974.
- [38] A.H. Gray, J.D. Markel, "Distance Measures for Speech Processing", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol ASSP-24, no.5, 10/1976.
- [39] R.M. Gray, A. Buzo, A.H. Gray, Y. Matsuyama, "Distortion Measures for Speech Processing", *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol ASSP-28, no.4, 8/1980.
- [40] F. Hu, W. Henkel, "Turbo-like iterative least-squares decoding of analogue codes", *Electronics Letters*, vol. 41, 10/2005.
- [41] F. Itakura, "Line Spectrum Representation of Linear Predictive Coefficients of Speech Signals", *Journal of the Acoustical Society of America*, vol. 57, no. s1, 4/1975.
- [42] Y. Linde, A. Buzo, R. Gray, "An Algorithm for Vector Quantizer Design", *IEEE Transactions on Communications*, vol. 28, 1/1980.
- [43] D.D. McRae, "Performance Evaluation of a New Modulation Technique", *IEEE Transactions on Communication Theory*, vol. COM-19, no. 4, 8/1971.
- [44] T. Miki, C.-E.W. Sundberg, N. Seshadri, "Pseudo-Analog Speech Transmission in Mobile Radio Communication Systems", *IEEE Transactions on Vehicular Technology*, 2/1993.
- [45] M. Nilsson, W.B. Kleijn, "On the Estimation of Differential Entropy from Data Located on Embedded Manifolds", *IEEE Transactions on Information Theory*, vol53, no. 7, 07/2007.
- [46] H. Nyquist, "Certain Topics in Telegraph Transmission theory", *Trans. AIEE*, vol. 47, 4/1928.
- [47] I. Peterson, "Messages in mathematically Scrambled Waves - Techniques for Scrambling Analogue Information Such As Telephone and television Signals", *Science News*, July 20, 1991.

- [48] N. Phamdo, U. Mittal, "A Joint Source-Channel Speech Coder Using Hybrid Digital-Analog (HDA) Modulation", *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 4, 5/2002.
- [49] P. Prandoni, M. Vetterli, "R/D Optimal Linear Prediction", *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 6, 11/2000.
- [50] I. Rosenhouse, A.J. Weiss, "Combined Analog and Digital Error-Correcting Codes for Analog Information Sources", *IEEE Transactions on Communications*, vol. 55, no. 11, 11/2007.
- [51] M.R. Schroeder, B.S. Atal, J. Hall, "Optimizing Digital Speech Codecs by Exploiting Masking Properties of the Human ear", *Journal of the Acoustical Society of America*, vol. 66, no.6, 12/1979.
- [52] S. Shamai (Shitz), S. Verdú, R. Zamir, "Systematic Lossy Source/Channel Coding", *IEEE Transactions on Information Theory*, vol. 44, no. 2, 3/1998.
- [53] C.E. Shannon, "Communication in the Presence of Noise", *Proceedings IRE*, vol. 47, 1/1949.
- [54] M. Skoglund, N. Phamdo, F. Alajaji, "Design & Performance of VQ-Based Hybrid Digital-Analog Joint Source-Channel Codes", *IEEE Transactions on Information Theory*, vol. 48, no. 3, 3/2002.
- [55] M. Skoglund, N. Phamdo, F. Alajaji, "Hybrid Digital-Analog Joint Source-Channel Coding for Bandwidth Compression/Expansion", *IEEE Transactions on Information Theory*, vol. 52, no. 8, 8/2006.
- [56] T. Thiede et al, "PEAQ - The ITU Standard for Objective Measurement of Perceived Audio Quality", *Journal of the Audio Engineering Society*, vol. 48, 1/2000.
- [57] C.M. Thomas, C.L. May, G.R. Welty, "Hybrid Amplitude-and-Phase Modulation for Analog Data Transmission", *IEEE Transactions on Communications*, vol. COM-23, no. 6, 6/1975.
- [58] C.K. Un, D.T. Magill, "The Residual-Excited Linear Prediction Vocoder with Transmission Rate Below 9.6 kbit/s", *IEEE Transactions on Communications*, vol. COM-23, no. 12, 12/1975.
- [59] V.A. Vaishampayan, S.I.R. Costa, "Curves on a Sphere, Shift-Map Dynamics, and Error Control for Continuous Alphabet Sources", *IEEE Transactions on Information Theory*, vol. 49, no.7, 7/2003.
- [60] A.J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm", *IEEE Transactions on Information Theory* vol. 13, no. 2, 4/1967.

- [61] J. Ziv, "The Behaviour of Analogue Communication Systems", *IEEE Transactions on Information Theory*, vol. IT-16, 9/1970.
- [62] T. Berger, *Rate Distortion Theory*, Prentice Hall, 1971.
- [63] I.N. Bronstein, K.A. Semendjajew, *Taschenbuch der Mathematik*, Teubner, 1991.
- [64] T. Brüggem, *Unequal Error Protection by Source-Controlled Digital Modulation*, Wissenschaftsverlag Mainz, 2006
- [65] J.J. Callahan, *The Geometry of Spacetime: An Introduction to Special and General Relativity*, Springer Verlag, 2000
- [66] T.M. Cover, J.A. Thomas, *Elements of Information Theory*, Wiley, 1991.
- [67] C. Erdmann, *Hierarchical Vector Quantization: Theory and Application to Speech Coding*, Wissenschaftsverlag Mainz, 2005.
- [68] S. Haykin, *An Introduction to Analog and Digital Communications*, Wiley, 1989.
- [69] J. Heitzer (ed.), *Spiralen*, Klett Verlag, 1998.
- [70] F. Hillebrand (ed.), *GSM and UMTS, The Creation og Global Mobile Communication*, Wiley, 2002.
- [71] P. Jax, *Enhancement of Bandlimited Speech Signals: Algorithms and Theoretical Bounds*, Mainz, 2002.
- [72] N.S. Jayant, P.Noll, *Digital Coding of Waveforms*, Prentice Hall, 1984.
- [73] J. Johann, *Modulationsverfahren*, Springer Verlag, 1992
- [74] J.D. Markel, A.H. Gray, *Linear Prediction of Speech*, Springer, 1976.
- [75] J. Makhoul, J. Wolf, *Linear Prediction and the Spectral Analysis of Speech*, Technical Report, Bolt, Beranek and Newman Inc., Cambridge, Massachusetts.
- [76] P. Kabal, "An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality", *TSP Lab Technical Report*, Dept. Electrical & Computer Engineering, McGill University, 5/2002.
- [77] K.-D. Kammeyer, *Nachrichtenübertragung*, 4. neu bearbeitete und ergänzte Auflage, Vieweg + Teubner, 2008.
- [78] W.B. Kleijn, K.K. Paliwal (ed.), *Speech Coding and Synthesis*, Elsevier, 1995.
- [79] V.A. Kotel'nikov, *The Theory of Optimum Noise Immunity*, McGraw-Hill Book Company Inc., 1959.

-
- [80] J.P. Odenwalder, "Optimal Decoding of Convolutional Codes", Ph.D. dissertation, UCLA, 1970.
- [81] J.G. Proakis, *Digital Communications*, McGraw-Hill, 2001.
- [82] M.S. Roden, *Analog and Digital Communication Systems*, 3rd ed., Prentice-Hall, 1991.
- [83] D.J. Sakrison, *Communication Theory: Transmission of Waveforms and Digital Information*, John Wiley & Sons Inc., 1968.
- [84] D.J. Sakrison, *Notes on Analog Communication*, Van Nostrand-Reinhold, 1970.
- [85] H. Schupp, H. Dabrock (ed.), *Höhere Kurven*, BI Wissenschaftsverlag, 1995.
- [86] C.E. Shannon, *A Mathematical Theory of Communication*, AT&T, 1948.
- [87] B. Sklar, *Digital Communications*, Pearson, 2001.
- [88] G. Steinberg (ed.), *Polarkoordinaten*, Metzler Verlag, 1993.
- [89] Y. Tsvividis, *Mixed Analog-Digital VLSI Devices and Technology*, World Scientific Pub Co, 2002.
- [90] J.M. Wozencraft, I.M. Jacobs, *Principles of Communication Engineering*, John Wiley & Sons Inc., 1965.
- [91] P. Vary, R. Martin, *Digital Speech Transmission - Enhancement, Coding and Error Concealment*, Wiley, 2006.
- [92] P. Vary, *Skriptum zur Vorlesung Nachrichtensysteme I*, IND, 2006.
- [93] P. Vary, *Skriptum zur Vorlesung Nachrichtensysteme II*, IND, 2007.
- [94] R.W. Yeung, *A First Course in Information Theory*, Kluwer Academic / Plenum Publishers, 2002.
- [95] 3GPP TS 26.090, *Adaptive Multi-Rate Speech Codec; Transcoding functions*, 6/1999.
- [96] 3GPP TS 26.190, *Adaptive Multi-Rate Wideband Speech Codec; Transcoding functions*, 3/2002.
- [97] DARPA-TIMIT, "Acoustic-Phonetic Continuous Speech Corpus", *NIST Speech Disc 1-1.1*, 1990.
- [98] ETSI ETS300575, *Digital cellular telecommunications system (Phase 2+)(GSM); Channel Coding (GSM 05.03)*.
- [99] ETSI Rec. GSM 06.10, *GSM Full Rate Speech Transcoding*, 1988.

-
- [100] ETSI Rec. GSM 06.55, *Digital cellular telecommunications system (Phase 2+); Performance characterization of the GSM Enhanced Full Rate (EFR) speech codec*, 2000.
- [101] ETSI Rec. GSM 06.60, *Digital cellular telecommunications system (Phase 2+); Enhanced Full Rate (EFR) speech transcoding*, 2000.
- [102] ETSI TR 26.975, *Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System; Performance Characterization of the Adaptive Multi-Rate (AMR) Speech Codec*, 6/2007.
- [103] ETSI TR 26.976, *Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System; Performance Characterization of the Adaptive Multi-Rate Wideband (AMR-WB) Speech Codec*, 6/2007.
- [104] ITU-R Rec. BS.1387, *Method for Objective Measurements of Perceived Audio Quality*, 12/1998.
- [105] ITU-T Rec. G.722.2, *Wideband Coding of Speech at Around 16 kbit/s Using Adaptive Multi-Rate Wideband Speech Transcoding*, Geneva, 1/2002.
- [106] ITU-T Rec. G.729.1, *G.729 based Embedded Variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729*, 2006.
- [107] ITU-T Rec. P.862.2, *Perceptual Evaluation of Speech Quality..*
- [108] ITU-T Rec. P.800, *Methods for Subjective Determination of Transmission Quality*, 1996.
- [109] ITU-T Rec. P.830, *Subjective Performance Assessment of Telephoneband and Wideband Digital Codecs*, 1996.
- [110] NIST *US Federal Information Processing Standard 197*, 2001

Curriculum Vitae

Name: Geburtstag: Geburtsort: Staatsangehörigkeit: Familienstand:	Carsten Hoelper 25. September 1973 Düsseldorf deutsch verheiratet, 2 Kinder
Schulbildung	1980 – 1984 Kath. Grundschule Pesch. 1984 – 1993 Gymnasium Korschenbroich, Abitur.
Wehrdienst	1993 – 1994 Grundausbildung und Fernmeldetätigkeit.
Studium der Elektrotechnik	Oktober 1994 – Dezember 1994 Rheinisch Westfälische Technische Hochschule Aachen ▷ <i>Vertiefungsrichtung:</i> Nachrichtentechnik. ▷ <i>Externe Studienarbeit:</i> NTNU Trondheim, Norwegen. ▷ <i>Diplomarbeit:</i> ‘Entwicklung und Untersuchung von Verfahren zur Sprachcodierung mit integrierter Störgeräuschreduktion’ bei Prof. Dr.-Ing. P. Vary.
Beruflicher Werdegang	Januar 2000 – Dezember 2008 Wissenschaftlicher Mitarbeiter am Institut für Nachrichtengeräte und Datenverarbeitung der RWTH Aachen (Prof. Dr.-Ing. Peter Vary) ▷ ’00 – ’02 Koordination und Bearbeitung eines R&D Projektes von Ericsson mit der RWTH. ▷ ’03 – ’07 Koordination des internationalen Masterstudienganges ‘Communications Engineering’. ▷ ’05 – ’09 Forschungsarbeit zur Dissertation. Seit März 2009 Automotive Innovation Center Aachen ▷ Aufbau einer terrestrischen Galileo Testumgebung für den bodengebundenen Verkehr.