Jan Dohl

Blind Estimation and Mitigation of Nonlinear Channels

Beiträge aus der Informationstechnik

Mobile Nachrichtenübertragung
Nr. 75

Jan Dohl

Blind Estimation and Mitigation of Nonlinear Channels

VOGT

Dresden 2015

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.dnb.de abrufbar.

Bibliographic Information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Zugl.: Dresden, Techn. Univ., Diss., 2014

Die vorliegende Arbeit stimmt mit dem Original der Dissertation „Blind Estimation and Mitigation of Nonlinear Channels“ von Jan Dohl überein.

Gesetzt vom Autor

ISBN 978-3-938860-92-2

Jörg Vogt Verlag
Niederwaldstr. 36
01277 Dresden
Germany

Phone: +49-(0)351-31403921
Telefax: +49-(0)351-31403918
e-mail: info@vogtverlag.de
Internet : www.vogtverlag.de

Technische Universität Dresden

Blind Estimation and Mitigation of Nonlinear Channels

Jan Dohl

von der Fakultät Elektrotechnik und Informationstechnik der
Technischen Universität Dresden

zur Erlangung des akademischen Grades eines

Doktoringenieurs

(Dr.-Ing.)

genehmigte Dissertation

Vorsitzender:	Prof. Dr.-Ing. habil. Renate Merker
Gutachter:	Prof. Dr.-Ing. Dr. h.c. Gerhard P. Fettweis
	Prof. Dr.-Ing. Dr.-Ing. habil. Johannes Huber
Tag der Einreichung:	01. September 2014
Tag der Verteidigung:	27. November 2014

Abstract

The worldwide demand for wireless data services is on the rise for several years now. With the introduction of smartphones and tablets, this trend has intensified. Many users now own multiple devices and data needs to be easily available across all of them. This has lead to an embrace of cloud services not only for documents, but also for photos, music and even video data, yielding another spike in traffic demand.

Multicarrier modulation is the current answer to the ever rising traffic demand. It allows the efficient usage of large bandwidths at relatively low computational complexity. It has been in use in digital video broadcasting and wireless local area networks for a while and is now introduced in cellular communications with the advent of LTE.

The downside of multicarrier modulation is its very high dynamic range which results in instantaneous power peaks. Amplifiers need to be driven with a large power reserve in order to cope with these peaks. This however reduces their energy efficiency dramatically which is especially bad for battery powered devices. Furthermore, development and production of these highly linear amplifiers is costly.

Driving the amplifiers with less reserve causes nonlinear distortion of the power peaks and hence a significant reduction in signal quality. In the past, both transmitter and receiver based methods have been presented to mitigate these distortions. However, very often perfect knowledge of the nonlinearity characteristic is assumed which is not realistic especially in receiver based methods.

The focus of this thesis is on methods that estimate the nonlinearity characteristic on the receiver side. Of special interest are so called blind algorithms because they don't require special pilot signals and hence can be used with existing standards. The main focus is on the formal derivation of blind estimation methods and their low-complexity implementation. It is shown that there is virtually no performance gap between estimated and perfect nonlinearity knowledge.

One of the methods has been implemented on a software defined radio platform. It is shown that significant performance gains can be reached for real nonlinear amplifiers. The system runs in real time on cheap off-the-shelf components proving the low complexity of the method which is ready for implementation on today's hardware.

Kurzfassung

Die weltweite Nachfrage nach drahtlosen Breitbanddiensten wächst seit mehreren Jahren kontinuierlich an. Dieser Trend hat sich mit der Einführung von Smartphones und Tablets weiter verstärkt. Viele Nutzer besitzen mehrere mobile Geräte und benötigen eine einfache Lösung zur Synchronisation. Die Cloud erfüllt dieses Bedürfnis und wird nicht nur für Dokumente, sondern auch für Fotos, Bilder und Videos benutzt. Dies treibt die Nachfrage nach Datendiensten zusätzlich an.

Die heutige Antwort liegt im Einsatz von Mehrträgermodulation. Sie verbindet die effiziente Nutzung breitbandiger Signale mit moderatem Rechenaufwand. Mehrträgermodulation ist seit mehreren Jahren im digitalen Rundfunk und drahtlosen Netzwerken im Einsatz und mit der Einführung von LTE auch im Mobilfunkbereich angekommen.

Ein Nachteil ist der hohe Dynamikumfang von Mehrträgersignalen, der sich in hohen kurzzeitigen Leistungsspitzen äußert. Um diese Spitzen nicht zu verzerren, müssen Verstärker mit hoher Leistungsreserve betrieben werden. Die Effizienz wird dadurch stark reduziert was besonders für batteriebetriebene Geräte von Nachteil ist. Außerdem sind diese hochgradig linearen Verstärker sehr teuer in Entwicklung und Produktion.

Werden die Verstärker mit weniger Leistungsreserve betrieben, wird die Signalqualität durch nichtlineare Verzerrung der Leistungsspitzen deutlich reduziert. In den letzten Jahren wurden viele Ansätze zur sender- und empfängerseitigen Korrektur nichtlinear verzerrter Signale vorgestellt. Oftmals wird in den jeweiligen Publikationen perfekte Kenntnis der Verstärkerkennlinie angenommen. Dies ist jedoch speziell für empfängerseitige Algorithmen unrealistisch.

Der Fokus dieser Dissertation liegt daher auf der empfängerbasierten Schätzung der Kennlinie der Nichtlinearität. Sogenannte blinde Algorithmen sind hierbei von besonderem Interesse weil sie keine speziellen Pilotsignale benötigen und daher kompatibel zu bestehenden Standards sind. Der Großteil der Arbeit besteht einerseits aus der formellen Herleitung blinder Schätzverfahren, andererseits aus dem Ableiten von Berechnungsverfahren mit geringer Komplexität. Es wird gezeigt, dass es praktisch keinen Unterschied zwischen der erreichbaren Systemperformance bei geschätzer und perfekter Nichtlinearitätskennlinie gibt.

Ein Algorithmus wurde auf einer Software Defined Radio Plattform implementiert. Es wird gezeigt, dass auch bei Einsatz realer Verstärker im Sättigungsbetriegb

eine deutliche Performancesteigerung möglich ist. Das System ist mit preiswerten standardkomponenten umgesetzt und ermöglicht Echtzeitbetrieb, was die niedrige Komplexität der Algorithmen und damit die Einsatzfähigkeit in heutigen Systemen beweist.

Contents

List of Figures

List of Tables

List of Acronyms

3DCSI 3D Chip Stack Interconnects

ADSL Asymmetric Digital Subscriber Line

AM/AM Amplitude to Amplitude

AM/PM Amplitude to Phase

AWGN Additive White Gaussian Noise

BER Bit Error Rate

CDF Cumulative Distribution Function

CP Cyclic Prefix

CRLB Cramér-Rao Lower Bound

CSS3 Cascading Style Sheets v3

DAB Digital Audio Broadcasting

DAC Digital-to-Analog Converter

DC Direct Current

DFG Deutsche Forschungs Gesellschaft

DFR Decision Feedback Receiver

DFT Discrete time Fourier Transformation

DVB-T Terrestrial Digital Video Broadcasting

FBMC Filter Bank Multi Carrier

FFT Fast Fourier Transformation

GFDM Generalized Frequency Division Multiplexing

GUI Graphical User Interface

HAEC Highly Adaptive Energy-Efficient Computing

HTML5 Hypertext Markup Language v5

i.i.d. independent and identically distributed

ICI Inter Carrier Interference

IDFT Inverse Discrete time Fourier Transformation

ISI Inter Symbol Interference

LTE Long Term Evolution

MAP Maximum a-posteriori

ML Maximum Likelihood

MLE Maximum Likelihood Estimator

MSE Mean Squared Error

OFDM Orthogonal Frequency Division Multiplexing

PAPR Peak to Average Power Ratio

PC Personal Computer

PDF Probability Density Function

PDP Power Delay Profile

PSD Power Spectral Density

PSK Phase Shift Keying

QAM Quadrature Amplitude Modulation

QPSK Quaternary Phase Shift Keying

SDR Software Defined Radio

SLM Selected Mapping

SNR Signal to Noise Ratio

SSPA Solid State Power Amplifier

TWTA Travelling Wave Tube Amplifier

USRP Universal Software Radio Platform

VDSL Very High Speed Digital Subscriber Line

WAN Wide Area Network

WLAN Wireless Local Area Network

1. Introduction

1.1. Development of Communications Systems

The trend towards transmitting data with ever increasing rates continues both in Wireless Local Area Network (WLAN) as well as celluar Wide Area Network (WAN) applications. Portable smartphones and tablet computers with ever growing computing power and high-definition displays are replacing traditional desktop and mobile Personal Computers (PCs) especially in many home applications. Cloud computing together with modern web standards such as Hypertext Markup Language v5 (HTML5), Cascading Style Sheets v3 (CSS3) and JavaScript now allows for rich web applications that increasingly replace classical desktop applications. Furthermore, the advent of high-definition audio and video streaming services create demand for always-on, low-latency internet connections with ever increasing bandwidth requirements.

One major technology behind this development is multicarrier modulation, especially in form of Orthogonal Frequency Division Multiplexing (OFDM). OFDM is utilized in current and future cellular networks such as Long Term Evolution (LTE) or LTE-advanced, local area networks (802.11a/g/n), digital broadcasting services like Digital Audio Broadcasting (DAB) or Terrestrial Digital Video Broadcasting (DVB-T) and internet access over copper wires like Asymmetric Digital Subscriber Line (ADSL) and more recently also Very High Speed Digital Subscriber Line (VDSL). One reason for todays dominance of OFDM is the computationally cheap implementation of transmitters and receivers by means of the Fast Fourier Transformation (FFT) as well as the relatively simple equalization of frequency selective channels when compared to other multi-carrier or even single carrier methods. Furthermore, OFDM enables fine grained resource allocation in time and frequency which is crucial for high speed cellular standards such as LTE. However, there are some practical issues when implementing OFDM.

1.2. Practical Implementation of OFDM Systems

OFDM can be described as a parallel transmission of multiple orthogonal single carrier signals. Each of the parallel transmissions is refered to as an OFDM subcarrier and modern OFDM systems employ 1024 or more subcarriers. Similar to single carrier systems, each subcarrier has a relatively limited dynamic range. However, due to constructive interference, the sum of all subcarriers exhibits large power spikes that increase the signal dynamic range. Power peaks in OFDM systems can reach 10 dB and more above the average power level. This imposes big issues for the design of power amplifiers in OFDM systems. Development and manufacturing of amplifiers operating linearly over a wide dynamic range is a challenging engineering problem. Furthermore, most amplifiers operate at peak energy efficiency when driven close to saturation. However, in order not to distort power peaks of OFDM signals, an amplifier operating point providing a large power reserve, the so called backoff, has to be chosen. This puts the amplifier into an energy inefficient mode of operation which is especially bad for battery powered devices such as laptops, tablets or smartphones.

In the past, this problem has attracted a lot of attention from the research community and many promising methods have been developed to alleviate nonlinear distortions. Most of these fall in one of three categories. Firstly, predistortion methods introduce a digital nonlinearity in the signal path such that the combined effects of digital and analog nonlinearity yield an overall linear characteristic. Secondly, Peak to Average Power Ratio (PAPR) reduction methods seek to reduce the dynamic range of the OFDM time domain signals. For example, a different data-to-carrier mapping or redundant coding can reduce the dynamic range of the resulting time signal. Finally, it has been shown that even strong nonlinear distortions do not have a large effect on the resulting channel capacity. Based on these results, receiver architectures have been developed that can mitigate the effects of nonlinear distortions to a large degree. There also exist hybrid methods such as clipping and filtering that combine digital predistortion to reduce the dynamic range of the signal and a receiver that can mitigate the resulting signal distortions.

1.3. State of the Art

As stated above, many methods have been developed to cope with the high dynamic range of OFDM signals and the problems that arise when amplifying these signals. These methods usually require knowledge of the nonlinearity characteristic of the amplifier. Special training sequences have been proposed to obtain this knowledge. However, employing such a pilot sequence would reduce the overall data rate. Therefore, it is desireable to obtain knowledge about the nonlinearity characteristic only by observing the data signal at the receiver. This approach is commonly known as blind estimation. To the best knowledge of the author, there exists no comprehensive analysis of blind nonlinearity estimation. The focus and the new contributions of this thesis are in this field.

1.4. Goals of this Work

The following are the three main goals of this work:

1. Show the feasibility of blind nonlinearity estimation by mathematical derivation of estimators and characterization of their performance.
2. Analyze the complexity of the estimators and evaluate their applicability to real-time environments. If required, derive low-complexity algorithms and give an overview of the tradeoffs.
3. Derivations are done using widely accepted nonlinearity models. Show that the resulting algorithms are suitable for use in the real world by testing them with actual hardware amplifiers operated in saturation mode.

1.5. Notation

The following notation is used throughout the thesis.

- h, H: Normal symbols represent scalar values. In cases where a signal could be in time or frequency domain, lower case symbols represent time domain values and upper case symbols represent frequency domain values.
- $\boldsymbol{h}, \boldsymbol{H}$: Bold symbols represent vectors. Upper and lower case rules are similar to scalars.

- $\underline{\boldsymbol{H}}$: Underlined bold symbols represent matrices. They are usually using upper case letters.

- s_k, $[\boldsymbol{s}_\phi]_k$: Both notations represent the k-th element of the vector $\boldsymbol{s}$ resp. $\boldsymbol{s}_\phi$. The first notation is used when the vector itself does not have a subscript as part of its name. It is often used together with a sum or product operator that defines the index variable. If no such operator is present it represents an arbitrary element of that vector, e.g. in cases where an equation is defined per-element. The second notation is used in ambiguous cases, e.g. when the vector itself has a subscript as part of its name.

- $[\underline{\boldsymbol{H}}]_{ij}$: Denotes the element in the i-th row and j-th column of the matrix $\underline{\boldsymbol{H}}$.

1.6. Structure of the Thesis

The thesis is structured as follows:

In **chapter 2**, an introduction to OFDM signals and memoryless nonlinearities is given. The effects of nonlinear distortion of OFDM signals is characterized and an overview of existing mitigation methods is presented.

In **chapter 3**, a blind feedforward nonlinearity estimator based on maximum likelihood estimation is formally derived. The estimation is based solely on the received signal. It is shown that it exhibits prohibitively high computational complexity and low complexity methods are derived for two special cases.

In **chapter 4**, a blind feedback nonlinearity estimator is derived. In addition to the received signal, it also uses the information in the received information bits through remodulation. It is shown that the method offers better estimation performance in a wider range of scenarios with lower computational complexity.

In **chapter 5**, a software defined radio implementation of a feed forward estimator working with signal distorted by real amplifiers is presented.

In **chapter 6**, a summary of the most important results as well as an outlook on future work is given.

2. Nonlinearly Distorted Orthogonal Frequency Division Multiplexing (OFDM) Signals

Multicarrier modulation using orthogonal subcarriers, i.e., OFDM, has become a wide-spread modulation scheme for many mobile radio applications. As such, it forms the basis for all methods derived in this thesis. In this chapter, a short introduction to OFDM, its relevant properties and the effects of nonlinear distortions are given. For a more general introduction to OFDM, there are numerous publications such as [Nee99, BSE04].

In section 2.1, the OFDM modulation scheme is defined and the most relevant statistical properties of the resulting signals are described. Memoryless nonlinearities and their effects on OFDM signals are investigated in section 2.2. Finally, an overview of current approaches to mitigate the negative effects of nonlinearities on OFDM signals is given in section 2.3.

2.1. Introduction to OFDM

2.1.1. From Orthogonality to OFDM

Orthogonality A set of functions ξ_m is called orthogonal if

$$\int_a^b \xi_m(t)\xi_n^*(t)\mathrm{d}t = \begin{cases} 0 & \text{if } m \neq n \\ \|\xi_m\|_2^2 & \text{if } m = n\,. \end{cases} \tag{2.1}$$

for appropriate integration boundaries [WJ65]. The set is called orthonormal if $\|\xi_m\|_2^2 = 1$, which can generally be achieved by means of normalization. Given a weighted sum of orthonormal functions

$$\xi(t) = \sum_{m=0}^{N} d_m \xi_m(t)\,, \tag{2.2}$$

each weight d_m can be recovered as follows:

$$d_m = \int_a^b \xi(t)\xi_m^*(t)\mathrm{d}t\,. \tag{2.3}$$

By setting d_m to represent information, this concept is widely used in communications and forms the basis for many modulation schemes and algorithms.

Single-carrier Many typical single-carrier transmission methods employ a Quadrature Amplitude Modulation (QAM) scheme. The time discrete data symbols $d_m \in \mathbb{C}$ are used to weight a train of unit impulses placed equidistantly with the symbol duration T_s. This weighted impulse train is then filtered with a band-limiting pulse shaping filter that has the impulse response $h_p(t)$ to form the baseband signal $s(t)$ as follows:

$$s(t) = \sum_{m=-\infty}^{\infty} d_m \cdot h_p(t - mT_s)\,. \tag{2.4}$$

If the system is designed to be orthogonal, comparing to (2.2) shows that the set of orthogonal functions is formed from time shifted variants of the pulse shape $h_p(t)$. This is achieved most easily by making the pulses non-overlapping, i.e., $h_p(t) = 0$ if $|t| > T_s/2$. However, there exist certain overlapping pulse shapes that still fulfill the orthogonality criteria, the so-called Nyquist pulses [Nyq28, PS07]. A system in which consecutive symbols do not interfere with one another is called Inter Symbol Interference (ISI)-free.

Multi-carrier Single-carrier systems spread the data symbols over time. Multi-carrier systems add another dimension by spreading out data symbols over time and frequency. A system with N_c subcarriers having center frequencies $f_0, \ldots, f_{N_c-1}$ is given by

$$s(t) = \sum_{m=-\infty}^{\infty} \sum_{l=0}^{N_c-1} S_{l,m} \cdot h_p(t - mT_s) \cdot \mathrm{e}^{j2\pi f_l t}\,, \tag{2.5}$$

where $S_{l,m}$ denotes the data symbol modulated onto the l-th subcarrier in the m-th time symbol. First it can be seen that a single-carrier system is just a special case

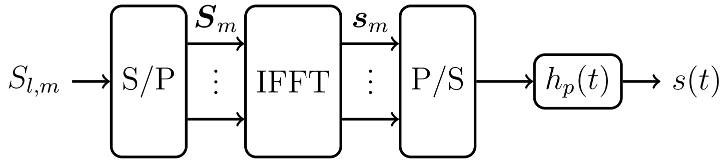

Figure 2.1.: Principal OFDM modulator

with $N_c = 1$ and $f_0 = 0$ Hz. Furthermore, the set of orthogonal functions now consists of the pulse shape $h_p(t)$ shifted in time by integer multiples of the symbol duration and in frequency by $f_0, \ldots, f_{N_c-1}$. If all subcarriers are orthogonal, the system is called Inter Carrier Interference (ICI)-free.

OFDM as an important special case of multi-carrier modulation is described now.

2.1.2. OFDM Modulation and Demodulation

OFDM is a very popular variant of multi-carrier modulation which is usually described and implemented in the digital domain. The simplest form of an OFDM modulator is shown in Figure 2.1. As indicated by the serial/parallel converter, an OFDM symbol $\boldsymbol{S}_m$ consists of N_c data symbols $S_{l;m}, 0 \leq l < N_c$. N_c is called the number of subcarriers of the OFDM symbol since it represents the number of simultaneously transmitted symbols in orthogonal frequency bins. The modulation to different frequencies is done by means of an Inverse Discrete time Fourier Transformation (IDFT). Hence, the time samples $\boldsymbol{s}_m$ belonging to the m-th OFDM symbol are given as:

$$s_m[k] = \frac{1}{\sqrt{N_c}} \sum_{l=0}^{N_c-1} S_{m;l} \cdot \mathrm{e}^{j2\pi \frac{lk}{N_c}} = \sqrt{N_c} \cdot \mathrm{IDFT}\left\{\boldsymbol{S}_m\right\} . \tag{2.6}$$

The factor $\frac{1}{\sqrt{N_c}}$ is a power normalization so that $\mathrm{E}\left\{|s(t)|^2\right\} = \mathrm{E}\left\{|S_{m;l}|^2\right\}$. Since the output of the IDFT consists of N_c time domain samples, the duration T_o of a single OFDM symbol is $T_o = N_c \cdot T_s$. (2.6) shows that the l-th data symbol belonging to each OFDM symbol is digitally mixed to the frequency $\frac{l}{N_c}$. The signal is then converted from the digital to the analog domain by pulse shaping with $h_p(t)$. The digital frequencies $\frac{l}{N_c}$ correspond to $\frac{l}{T_o}$ in the analog domain and hence, the frequency difference between two subcarriers is $\frac{1}{T_o}$ which has been shown [CTL12] to be the minimal required spacing for the subcarriers to be orthogonal. Hence, using the IDFT not only ensures orthogonality in the frequency domain avoiding inter-carrier interference, it also requires the least frequency resources possible to do so. Furthermore, using a Nyquist pulse for $h_p(t)$ ensures orthogonality in the time domain avoiding inter-symbol interference.

Assuming that there are no noise or other disturbances, i.e., the received signal $r_m(t) = s_m(t)$, the received samples

$$r_m[k] = s_m[k]\,. \tag{2.7}$$

OFDM demodulation is done by inverting (2.6), i.e.,

$$R_{m;l} = \frac{1}{\sqrt{N_c}} \sum_{k=0}^{N_c-1} r_m[k] \cdot \mathrm{e}^{-j2\pi \frac{lk}{N_c}}\,. \tag{2.8}$$

(2.8) can be expressed by means of the Discrete time Fourier Transformation (DFT) as follows

$$\boldsymbol{R}_m = \frac{1}{\sqrt{N_c}} \cdot \mathrm{DFT}\left\{\boldsymbol{r}_m\right\}\,, \tag{2.9}$$

where $\boldsymbol{r}_m$ and $\boldsymbol{R}_m$ represent vectorized notations of the received samples $r_m[k]$ and symbols $R_{m;l}$.

It shows, that OFDM modulation and demodulation can be achieved by means of IDFT and DFT. The Fast Fourier Transformation (FFT) and its inverse are very efficient algorithms for implementing these transforms and readily available on many development platforms.

2.1.3. Frequency selective channels

A mobile radio channel can be roughly characterized by two properties. The coherence time T_{coh} is the time in which the impulse response of the channel can be assumed to be constant. The coherence bandwidth B_{coh} is the bandwidth in which the frequency response of the channel can be assumed to be constant so that signals with smaller bandwidth experience only flat fading [Gar07].

One of the main reasons for using OFDM is the small bandwidth f_{sc} of each individual subcarrier. If $f_{sc} < B_{coh}$, each subcarrier is affected by approximately frequency flat fading that allows for relatively simple equalization. If the total bandwidth of the system is fixed to be B, $f_{sc} = \frac{B}{N_c}$ depends on the amount of subcarriers and can be made arbitrarily small by increasing the amount of subcarriers. Care should be taken that $\frac{N_c}{B} = T_o < T_{coh}$ or the fading will become time-selective during one OFDM symbol. There is a tradeoff between small subcarrier bandwidth and symbol duration given B is fixed. This means that there are channels for which $T_o < T_{coh}$ and $f_{sc} < B_{coh}$ cannot both be achieved. However, in many cases this is possible and in the following, fulfillment of both conditions is always assumed.

Since the channel coherence time is assumed to be longer than the duration of one OFDM symbol, the channel is modelled by a time-invariant impulse response $h(\tau)$ for the duration of one symbol. Subsequent symbols might see a different channel realization. For most practical channels the impulse response vanishes after some time, i.e., $h(\tau) = 0$ for some $\tau > \tau_c$. To satisfy $f_{sc} < B_{coh}$, it is required that $\tau_c < T_o$. Sampling the impulse response with the same sampling time $T_s = \frac{T_o}{N_c}$ as the OFDM system yields the vector

$$\boldsymbol{h} = \left[h(0), h(T_o \frac{1}{N_c}), h(T_o \frac{2}{N_c}), \ldots, h(T_o \frac{l_c - 1}{N_c}), 0, \ldots, 0 \right], \tag{2.10}$$

with l_c being the length of the impulse response counted in number of samples. Furthermore, let $\boldsymbol{H} = \mathrm{DFT}\{\boldsymbol{h}\}$.

Let $\boldsymbol{S}$ be a transmitted OFDM symbol and $\boldsymbol{H}$ the frequency response of the channel. Then, the resulting signal in frequency domain is

$$\boldsymbol{R} = \boldsymbol{S} \odot \boldsymbol{H}, \tag{2.11}$$

where $\odot$ denotes element-wise vector multiplication. For brevity, the time index m has been omitted. Each subcarrier of the source symbol is multiplied with just one complex element of the channel's frequency response, yielding frequency flat fading and allowing for very simple equalization. However, (2.11) is based on the convolution theorem of the Fourier Transform and in order to hold for the DFT, the convolution needs to be circular as follows:

$$\boldsymbol{r} = \boldsymbol{s} \circledast \boldsymbol{h}. \tag{2.12}$$

In a real system, convolution happens in a linear way. Due to channel memory, this causes OFDM symbols to leak energy into subsequent symbols yielding ISI. Furthermore, (2.12) does not hold anymore. In almost all OFDM systems, this problem is solved by introducing the Cyclic Prefix (CP).

Circular and linear convolution become indistinguishable after l_c samples for a channel with length l_c. Adding a cyclic prefix with length $l_{cp} > l_c$ as follows

$$\boldsymbol{s}_{cp} = [s[N_c - l_{cp}], \dots s[N_c - 1], s[0], \dots s[N_c - 1]] \tag{2.13}$$

ensures that the last N_c samples of $\boldsymbol{s}_{cp}$ after it was subject to linear convolution with $\boldsymbol{h}$ are equal to a circular convolution of $\boldsymbol{s}$ and $\boldsymbol{h}$. The cyclic prefix is typically designed to be larger than the maximum expected delay spread of the channel. On the receiver side, the cyclic prefix can be discarded or the additional information it contains can be used to aid in synchronization, frequency offset correction and other tasks [SvdBB95]. Using the cyclic prefix for synchronization works especially well with unique word OFDM where the cyclic prefix is replaced with an a-priori known sequence [HHH10]. In the following, the cyclic prefix is omitted from the derivations. It is implicitely assumed to be present, long enough to avoid ISI and discarded at the receiver. A more in-depth explaination of the cyclic prefix and especially the implications on data rates can be found in [DOB+09].

2.1.4. Signal Properties

For the coming analysis of nonlinearly distorted OFDM systems, the statistical signal properties of the time domain signals are crucial. The definition of an OFDM system (2.6) states that in time domain the signal is a superposition of N_c complex waves, i.e., the subcarriers of the system. Each complex wave is weighted by a complex-valued constant that represents a data symbol being transmitted on one subcarrier of one specific OFDM symbol. Firstly, the two theoretical extreme cases of just one subcarrier and infinitely many subcarriers are investigated.

One subcarrier system An OFDM system with only one subcarrier is similar to a single-carrier system that is shifted to a non-zero center frequency. The amplitude is constant during each symbol and possible amplitudes match the used symbol alphabet. Consequently, the Peak to Average Power Ratio (PAPR) is also defined by the symbol alphabet and can be 0 dB if a constant amplitude Phase Shift Keying (PSK)-type alphabet and rectangular pulse-shaping is used.

Many subcarrier system For $N_c \to \infty$, each time domain sample is formed by superposition of all weighted complex waves. Since each wave is randomly weighted with a symbol from the alphabet, this is equal to summing up N_c independent random variables. Therefore, using the central limit theorem, it follows that the probability distribution of the time domain samples $s[k]$ converges to a complex normal distribution for large N_c, i.e.,

$$s[k] \to CN(0, \sigma_s) \,. \tag{2.14}$$

$\sigma_s = \mathrm{E}\left\{|S[\cdot]|^2\right\}$ is the average symbol power of the modulation alphabet. If all subcarriers are used and no per-subcarrier power allocation (e.g., water-filling) is performed, then all time-domain samples are independent and identically distributed (i.i.d.). If there is a guardband (i.e., unused subcarriers) or unequal power allocation per subcarrier, the samples belonging to one OFDM symbol become correlated. Details on that are given later when required. Real and imaginary parts of the samples can also be considered independent which yields a Rayleigh distribution for the signal amplitudes and a uniform distribution for the signal phases.

Peak to Average Power Ratio (PAPR) Of special interest for this thesis is the ratio of peak power to average power κ which for a single OFDM symbol is defined as:

$$\kappa = \frac{\max\left(|s[k]|^2\right)}{\sigma_s^2} \,. \tag{2.15}$$

Specifically this is the PAPR of the time-discrete complex baseband samples $s[k]$. It is known that the PAPR of the continuous-time baseband signal $s(t)$ is usually higher. However, it has been shown [WPP08] that the difference is bounded and hence, the results apply similarly.

The main consequence of normally distributed time domain samples is a very high PAPR. In the case of no guardband, $\boldsymbol{s}$ consists of i.i.d. zero-mean complex normally distibuted samples with variance σ_s^2. Therefore, the amplitudes are i.i.d. Rayleigh distributed, i.e. $|s[k]| \to Rayleigh\left(\frac{\sigma_s}{\sqrt{2}}\right)$. Since the PAPR depends on the random vector $\boldsymbol{s}$, it is a random variable. The probability that the PAPR is smaller than a certain value κ_0 is given as

$$\begin{aligned} Pr\left(\kappa < \kappa_0\right) &= Pr\left(\forall \frac{|s[k]|^2}{\sigma_s^2} < \kappa_0\right) \\ &= \prod_{k=0}^{N_c-1} Pr\left(\frac{|s[k]|^2}{\sigma_s^2} < \kappa_0\right) \\ &= \left[F_{rayl}\left(\sqrt{\kappa_0}, \sigma = \frac{1}{\sqrt{2}}\right)\right]^{N_c} \,, \end{aligned} \tag{2.16}$$

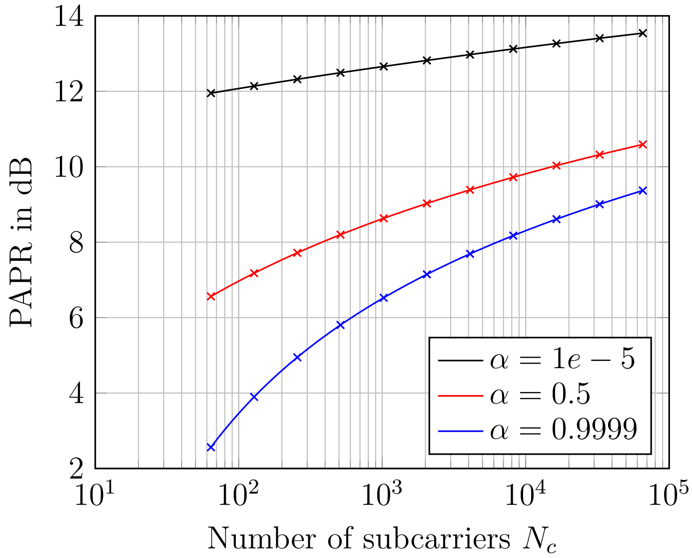

Figure 2.2.: PAPR of OFDM symbols, markers where N_c is power of two

where $F_{rayl}(\cdot)$ represents the Cumulative Distribution Function (CDF) of the Rayleigh distribution which is defined as

$$F_{rayl}(x, \sigma) = 1 - \mathrm{e}^{-\frac{x^2}{2\sigma^2}} . \tag{2.17}$$

Hence, the final result for $Pr\left(\kappa < \kappa_0\right)$ can be written as

$$Pr\left(\kappa < \kappa_0\right) = \left[1 - \mathrm{e}^{-\kappa_0}\right]^{N_c} \tag{2.18}$$

which has been shown before in [MBFH97]. Figure 2.2 gives an overview of the expected PAPR for different amounts of subcarriers. The three curves show the PAPR that is not exceeded by 0.001%, 50% and 99.99% of all OFDM symbols. Generally, the more subcarriers the system has, the higher the power peaks are. As an example, for Long Term Evolution (LTE) with 1200 active subcarriers, almost all symbols will have a PAPR between 6.5 dB and 13 dB.

As stated earlier, amplifiers need to be designed to exhibit a large linear range to amplify these signals without distortion. The following section describes the consequences when this condition is not met and OFDM signals are nonlinearly distorted.

Validity of Gaussian assumption Since the time-domain behaviour of OFDM signals differs greatly between the extreme cases of one and infinitely many subcarriers, the question about the amplitude distribution for a finite amount of subcarriers $N_c > 1$ arises. This is a well researched topic and the answers differ. In [Zil07], Zillmann gives empirical examples that show significant deviation of the densities of real and imaginary signal components from the normal distribution for a 64 subcarrier OFDM system with Quaternary Phase Shift Keying (QPSK) modulation. For higher-order modulation schemes such as 64-QAM the densities match more closely. Furthermore, for 1024 subcarriers the distributions seem to match very well even for QPSK modulation. In [DTV00], Dardari theoretically analyzed the influence of nonlinearities on OFDM systems. For 64 subcarriers, good matching between simulation and analysis with normally distributed samples was observed in terms of the resulting Bit Error Rate (BER). Recently in [AD12], Araujo investigated the validity of the Gaussian assumption for modelling the effects of nonlinearities on OFDM systems and found that for more than 100 subcarriers, the differences between the exact behaviour and the Gaussian approximation become insignificant. Many other publications on the topic implicitely assume the samples as Gaussian distributed and usually no significant discrepances between theory and simulation arise.

Based on this research the author considers it safe to assume that for $N_c >=$ 128 subcarriers, the Gaussian assumption is reasonably exact for investigating nonlinearities in an OFDM context. Since modern OFDM systems almost always have more subcarriers, this assumption is also realistic in the light of currently used communications systems. For the rest of this work, the Gaussian assumption is always assumed as valid.

2.2. Memoryless Nonlinearities and OFDM signals

It was pointed out in the previous sections that OFDM signals exhibit a large PAPR and why this is problematic for amplifier design. There exist a large amount of models for describing the behavior of amplifiers in the digital baseband that differ mostly by the characterization of frequency selective memory effects. A good overview is given in [PM05]. The relevant model class for this thesis is the memoryless nonlinearity which shall be described now.

2.2.1. Memoryless nonlinearities

Definition For a memoryless nonlinearity, the output at a certain time instant depends only on the input at exactly the same time instant. Furthermore, only the input signal's amplitude is relevant, not the phase. Since the thesis uses models in the equivalent complex baseband, the nonlinearities are specified as baseband models as well. To that end, a memoryless nonlinearity is defined as follows:

$$g(x) = g_A(|x|) \cdot \exp\left(\arg(x) + g_\phi(|x|)\right) . \tag{2.19}$$

The characteristic consists of an Amplitude to Amplitude (AM/AM) distortion $g_A(|x|)$ and an Amplitude to Phase (AM/PM) distortion $g_\phi(|x|)$.

Strict and quasi memoryless nonlinearities There is often some confusion about the exact meaning of memoryless. This arises from the AM/PM distortion that does not show up for strictly memoryless amplifers. Power amplifiers are usually the last element in a transmission chain before emitting the radio waves over the antenna. Hence, the amplifier operates on a passband signal centered around the carrier frequency f_c. Let $\widetilde{s}(t)$ and $\widetilde{r}(t)$ be the passband representations of $s(t)$ and $r(t)$:

$$\widetilde{s}(t) = \Re\left\{s(t) \cdot \exp(j2\pi f_c t)\right\} \tag{2.20}$$

$$r(t) = \frac{1}{\sqrt{2}} \cdot \left(\widetilde{r}(t) + j\mathcal{H}\left\{\widetilde{r}(t)\right\}\right) \cdot \mathrm{e}^{-j2\pi f_c t} \tag{2.21}$$

Here, $\mathcal{H}\{\cdot\}$ represents the Hilbert transform and $\left(\widetilde{r}(t) + j\mathcal{H}\left\{\widetilde{r}(t)\right\}\right)$ is the analytic representation of $\widetilde{r}(t)$ which is complex valued and has the same spectrum as $\widetilde{r}(t)$ for the positive frequencies but all negative frequencies removed. Hence, after down-conversion, a baseband signal around 0 Hz remains. Now, let $\widetilde{g}(\cdot)$ be a memoryless passband nonlinearity that is represented as a power series as follows:

$$\widetilde{r}(t) = \widetilde{g}(\widetilde{s}(t)) = \sum_{l=0}^{\infty} \widetilde{b}_l \widetilde{s}^{\,l}(t) . \tag{2.22}$$

Since the amplifier is in the analog/physical domain, all signals and coefficients $\widetilde{b}_l$ have to be real valued. It can then be shown that the baseband equivalent representation is given as [ZQDR05]:

$$r(t) = g(s(t)) = \sum_{l=0}^{\infty} b_{2l+1} |s(t)|^{2l} s(t) \tag{2.23}$$

with

$$b_{2l+1} = \frac{1}{2^{2l}} \binom{2l+1}{l} \tilde{b}_{2l+1} \,. \tag{2.24}$$

It shows that for a memoryless passband nonlinearity, the resulting power series representation of the baseband equivalent nonlinearity has only real coefficients, too. That means, for such a nonlinearity there is only AM/AM but no AM/PM distortion. Then, what real phenomenon does an AM/PM nonlinearity represent?

The answer is very similar to transferring a narrow-band signal over a frequency selective channel. As stated in section 2.1.3, when the signal bandwidth B is smaller than the channel coherence bandwidth B_{coh}, the fading becomes flat and can be represented as a multiplication with a single complex coefficient. Similarly, if the nonlinearity has a memory, but it's length τ_g is small compared to variations in the complex envelope of the signal, i.e., $s(t+\tau_g) \approx s(t)$, the resulting instantaneous effect on the signal is similar to frequency flat fading and can be represented as a single complex multiplication at each time instant. This is exactly what a combination of AM/AM and AM/PM nonlinearities does. A formal and detailed derivation of this is given in [RZ02].

In summary, a strictly memoryless nonlinearity will only exhibit AM/AM characteristics. However, if the amplifier exhibits a very short memory, it can be represented as memoryless with AM/AM and AM/PM distortions. Hence, this class of amplifiers is called quasi-memoryless in the literature.

2.2.2. Common models for memoryless nonlinearities

In the following, three very common models of memoryless nonlinearities are presented: The Solid State Power Amplifier (SSPA) model, the Travelling Wave Tube Amplifier (TWTA) model and the soft limiter.

SSPA Model The SSPA model was proposed in 1991 by Rapp [Rap91] as a model that characterizes the properties of solid state amplifiers better than the other models at that time. These properties are an improved linearity in the small signal region, the tendency of the transfer function to a maximum output amplitude for large

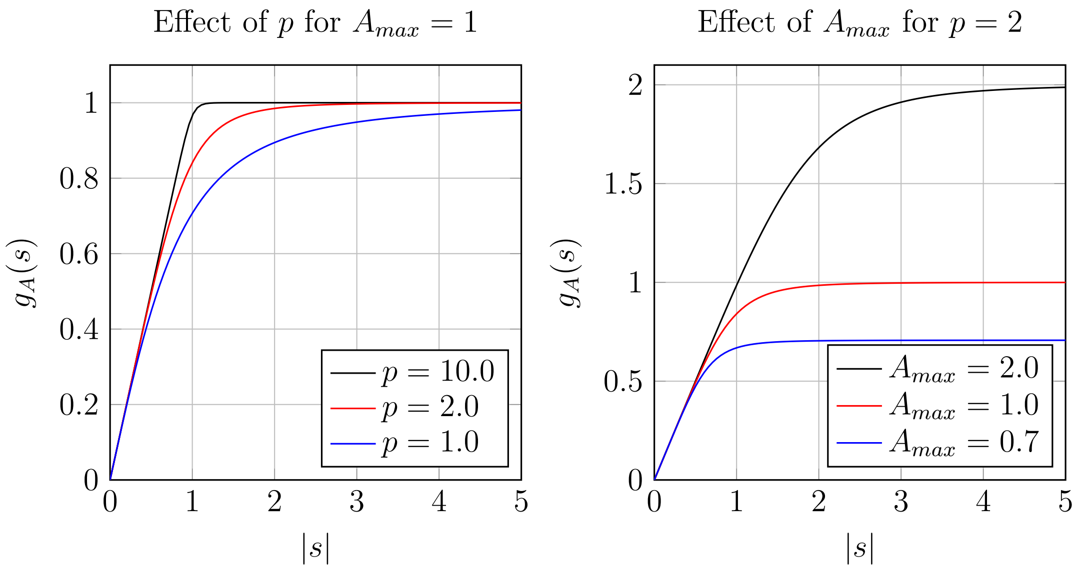

Figure 2.3.: AM/AM characteristic of SSPA model

inputs and the mostly negligible AM/PM distortion of solid state amplifiers. To that end, the model is defined as:

$$\begin{aligned} g_A(s, A_{max}, p) &= \frac{|s|}{\left[1 + \left(\frac{|s|}{A_{max}}\right)^{2p}\right]^{\frac{1}{2p}}} \\ g_\phi(s) &= 0\,. \end{aligned} \tag{2.25}$$

Figure 2.3 shows the resulting characteristic for different parameter combinations. The maximum output amplitude of the amplifier in saturation is A_{max} and p defines how smooth the transition between the linear and saturated part of the characteristic is. For small s, the model is linear and for large inputs it saturates towards A_{max}. The model does not exhibit AM/PM distortion to emphasize the negligible phase distortions of solid state amplifiers. It is therefore a strictly memoryless model.

TWTA Model The TWTA model was proposed in 1981 by Saleh [Sal81] to characterize the properties of travelling wave tube amplifiers. Today it is mainly used to model amplifiers in satellite communications. These amplifiers exhibit a strong AM/PM distortion and their amplitude characteristic has a maximum and tends to

	AM/AM	AM/PM
$\lvert s\rvert \to 0$	linear behavior	quadratic behavior
$\lvert s\rvert \to \infty$	$g_A(s) \to 0$	$g_\phi(s) \to \frac{\alpha_\phi}{\beta_\phi}$
Maxima	$\lvert s_0\rvert = \frac{1}{\sqrt{\beta_A}}$, $g_A(s_0) = \frac{\alpha_A}{2\sqrt{\beta_A}}$	None

Table 2.1.: Extreme cases of TWTA nonlinearity

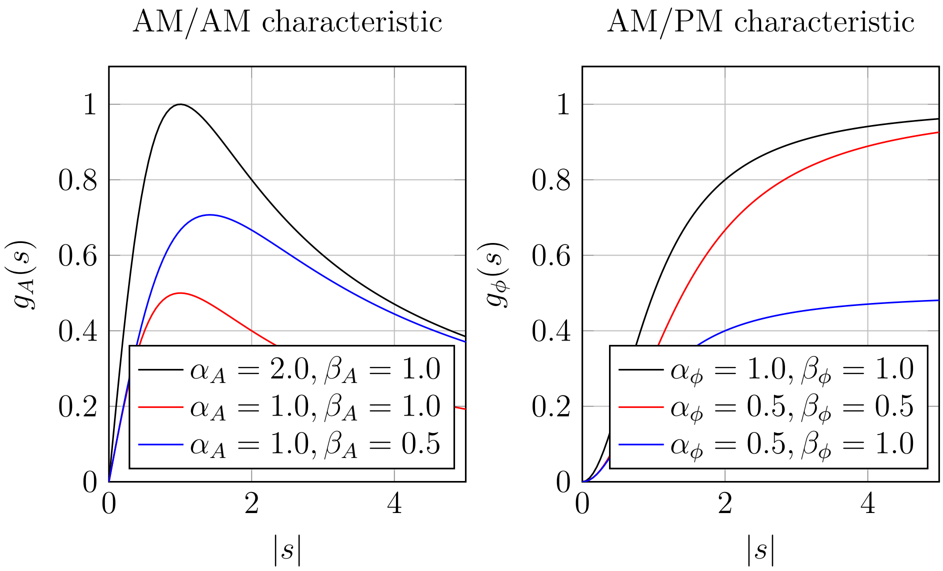

Figure 2.4.: Characteristic of TWTA model

decrease for larger inputs. The model is defined as follows:

$$g_A(s, \alpha_A, \beta_A) = \frac{\alpha_A |s|}{1 + \beta_A |s|^2} \tag{2.26}$$

$$g_\phi(s, \alpha_\phi, \beta_\phi) = \frac{\alpha_\phi |s|^2}{1 + \beta_\phi |s|^2} \, . \tag{2.27}$$

Since the relationship between the parameters and the shape of the resulting curves is not obvious, table 2.1 shows the behaviour for some special cases and Figure 2.4 shows exemplary AM/AM and AM/PM characteristics. The parameter α_A represents the small system gain of the model. In this thesis, normalized gain is always assumed without loss of generality, hence from now on $\alpha_A = 1$.

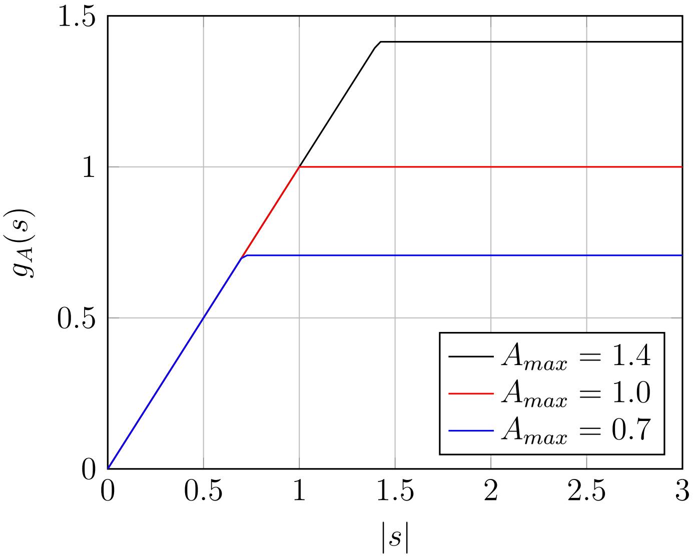

Figure 2.5.: Characteristic of soft limiter model

Soft Limiter The soft limiter is an idealized model for an amplifier with clipping. Its characteristic is perfectly linear until saturation A_{max} is reached. It is defined as:

$$g_A(s, A_{max}) = \begin{cases} |s| & \text{if } |s| \leq A_{max} \\ A_{max} & \text{if } |s| > A_{max} \end{cases} \tag{2.28}$$

$$g_\phi(s) = 0 \tag{2.29}$$

Figure 2.5 shows several exemplary characteristics. The model is an extreme case of the SSPA model for $p \to \infty$.

2.2.3. Nonlinearly distorted OFDM signals

In the last sections OFDM signals and memoryless nonlinearities have been introduced. In this section it is investigated how an OFDM signal is affected by a memoryless nonlinearity. Most of the following results are not limited to OFDM signals but apply in general to bandlimited normally distributed processes. As before, let $s(t)$ be the time domain representation of an OFDM symbol, $g(s)$ a memoryless nonlinearity and $r(t) = g(s(t))$ the signal after nonlinear distortion.

In-band interference and out-of-band leakage From the power series definition (2.23) follows that memoryless distortions have a linear multiplicative and a nonlinear

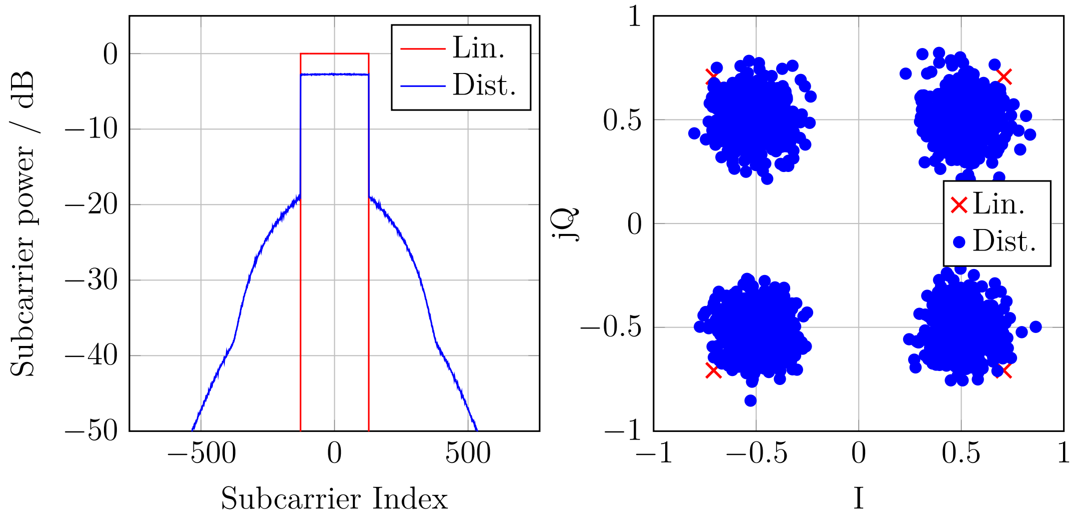

Figure 2.6.: Out-of-band radiation and in-band interference of a 256 subcarrier OFDM system distorted by an SSPA nonlinearity with $p = 2$ and $A_{max} = 0$ dB, QPSK modulation on each subcarrier

additive component. This becomes more visible by rewriting the equation as:

$$r(t) = g(s(t)) = b_1 s(t) + \underbrace{s(t) \cdot \sum_{l=1}^{\infty} b_{2l+1} |s(t)|^{2l}}_{s_g(t)}, \tag{2.30}$$

where the coefficients $b_l \in \mathbb{C}$ to encompass AM/PM distortions as well. Let $R(f)$, $S(f)$ and $S_g(f)$ be the Fourier Transforms of $r(t)$, $s(t)$ and $s_g(t)$. Then, the resulting spectrum is given as:

$$R(f) = b_1 S(f) + S(f) * S_g(f)\,. \tag{2.31}$$

Next to the scaled original spectrum, the output spectrum also contains the convolution of $S(f)$ and $S_g(f)$. Using the properties of the convolution reveals that the bandwidth of $R(f)$ is the sum of the bandwidths of $S(f)$ and $S_g(f)$.

Two effects can be observed. The total bandwidth of the signal is increased resulting in out-of-band leakage. If this happens on the transmitter side, the leakage may lead to violations of the spectral masks which limit the allowed emissions into neighboring channels. Furthermore, the in-band parts of the additive distortion term leads to ICI and therefore decreased Signal to Noise Ratio (SNR). Figure 2.6 shows the Power Spectral Density (PSD) and resulting constellation diagram of a 256 subcarrier OFDM system using QPSK to modulate each carrier that is distorted by a SSPA nonlinearity. To capture the out-of-band effects, the system was simulated with 32-fold oversampling to get a good approximation of the actual continuous time equivalent baseband signal passing through the nonlinearity model. As predicted, there is significant leakage in the adjacent frequency bands as well as a reduction in average energy. Furthermore, the constellation plot shows that the nonlinearity induced a large amount of interference.

While out-of-band leakage is an immensely important problem especially in mobile cellular communications, there are areas in wired and wireless communications where it is less of an issue. For example, the Deutsche Forschungs Gesellschaft (DFG) Highly Adaptive Energy-Efficient Computing (HAEC) project [FNL12] investigates wireless board-to-board communications in the 100-300 GHz band. Due to highly directed beams and very short range, interference is less of an issue and out-of-band distortion causes less disturbances in other links. Similarly, in the 3D Chip Stack Interconnects (3DCSI) project [FuHLF13], interconnects on a chip and between chips on a board are investigated where, again, out-of-band radiation is less of an issue while amplifier efficiency is crucial. The interested reader can find an in-depth analysis of out-of-band radiation in OFDM systems including an analytic derivation of the distorted signal's PSD in [CMP99,CP02]. For the in-band interference however, there is a very important theorem.

The Bussgang decomposition In 1952, Bussgang presented a very useful theorem [Bus52]. It states that if a zero-mean normally distributed process $s(t)$ is input into a memoryless nonlinearity $g(s)$, the cross-correlation $R_{sr}(\tau)$ of $s(t)$ with the output $r(t) = g(s(t))$ is proportional to the autocorrelation $R_{ss}(\tau)$ of the input process:

$$R_{sr}(\tau) = \alpha R_{ss}(\tau) \,. \tag{2.32}$$

In the original publication, the theorem was proven for time-continuous real-valued stationary normally distributed processes undergoing only AM/AM distortions. In that case, $\alpha = \mathrm{E}\left\{g'(s(t))\right\}$. The application to time-discrete processes is straight forward. Over time, different authors extended the scope of the theorem to more general applications. A major contribution was done by Minkoff in 1985. In his publication [Min85] he extended the scope not only to complex-valued normally distributed processes, but to any zero-mean complex valued process that exhibits circular symmetry, i.e., the value of the Probability Density Function (PDF) only depends on the amplitude but not the phase of the argument. Furthermore, Minkoff extended the theorem to encompass AM/PM nonlinearities and showed that it is impossible for an AM/PM nonlinearity to improve the post nonlinearity SNR. It has therefore been shown impossible that careful design of the AM/PM nonlinearity can ease the negative effects of an AM/AM nonlinearity.

It can be seen from (2.32) that by setting $\tau = 0$, α is given as:

$$\alpha = \frac{\mathrm{E}\{\boldsymbol{s}\boldsymbol{r}^*\}}{\mathrm{E}\{\boldsymbol{s}\boldsymbol{s}^*\}} = \frac{\mathrm{E}\{\boldsymbol{s}g^*(\boldsymbol{s})\}}{\sigma_s^2} \tag{2.33}$$

For nonlinearities that exhibit only AM/AM distortions, α will be real-valued whereas for AM/PM distortions, α will be complex valued. Equation (2.32) allows a stochastic description of the output process after the nonlinearity as follows:

$$\boldsymbol{r} = \alpha\boldsymbol{s} + \boldsymbol{n_d}\,. \tag{2.34}$$

The input signal is scaled with α and superimposed by a noise-like signal $\boldsymbol{n_d}$ that is uncorrelated to $\boldsymbol{s}$. $\boldsymbol{n_d}$ is usually not normally distributed and finding the distribution is very difficult in most cases. However, since both $\boldsymbol{r}$ and $\boldsymbol{s}$ are zero-mean, the same holds for $\boldsymbol{n_d}$. The average power σ_r^2 of the output process $\boldsymbol{r}$ is given as:

$$\sigma_r^2 = \mathrm{E}\{\boldsymbol{r}\boldsymbol{r}^*\} = \mathrm{E}\{g(\boldsymbol{s})g^*(\boldsymbol{s})\} = \mathrm{E}\left\{|g(\boldsymbol{s})|^2\right\}\,. \tag{2.35}$$

Assuming a normalized nonlinearity, it is a valid assumption that $|g(s)| \leq |s|$. Under this condition $\sigma_r^2 \leq \sigma_s^2$ so that the nonlinearity reduces the average output power. From (2.34) it also follows that the average output power is given by:

$$\sigma_r^2 = \mathrm{E}\{\boldsymbol{r}\boldsymbol{r}^*\} = \mathrm{E}\{(\alpha\boldsymbol{s} + \boldsymbol{n_d})(\alpha^*\boldsymbol{s}^* + \boldsymbol{n_d}^*)\} \tag{2.36}$$

$$= \mathrm{E}\left\{|\alpha\boldsymbol{s}|^2\right\} + \underbrace{\mathrm{E}\{\alpha\boldsymbol{s}\boldsymbol{n_d}^* + \alpha^*\boldsymbol{s}^*\boldsymbol{n_d}\}}_{0} + \mathrm{E}\left\{|\boldsymbol{n_d}|^2\right\} \tag{2.37}$$

$$\sigma_r^2 = |\alpha|^2\sigma_s^2 + \sigma_{nd}^2\,. \tag{2.38}$$

Combining both equations, the power of the additive distortion term $\boldsymbol{n_d}$ is given as:

$$\sigma_{nd}^2 = \sigma_r^2 - |\alpha|^2\sigma_s^2 = \mathrm{E}\left\{|g(\boldsymbol{s})|^2\right\} - |\alpha|^2\sigma_s^2\,. \tag{2.39}$$

Using this, the signal quality degradation of the nonlinearity alone can be characterized by the post-nonlinearity SNR ρ_{nd} as follows:

$$\rho_{nd} = \frac{|\alpha|^2 \sigma_s^2}{\sigma_{nd}^2} \, . \tag{2.40}$$

Comparing (2.34) with Figure 2.6 shows that the observations match the predictions again. Following a similar line of thought as in section 2.1.4, the frequency domain representation $\boldsymbol{N_d}$ of $\boldsymbol{n_d}$ is modeled as zero-mean normally distributed due to the central limit theorem.

While (2.34) looks very similar to (2.30), the important difference is that there is no correlation and hence no linear dependency between $\boldsymbol{n_d}$ and $\boldsymbol{s}$ which is not applying to (2.30). For that reason, $\alpha \neq b_1$ except for some specially crafted cases. Furthermore it should be noted that $\boldsymbol{n_d}$ is not a truly random variable. Instead, it is fully dependent on $\boldsymbol{s}$ and $g(s)$ and can be reproduced if both are known which will be useful for nonlinearity mitigation as discussed later.

2.2.4. Conclusions

In this section, the influence of memoryless nonlinear distortions on OFDM signals have been investigated. The main effects are:

- Out-of-band power leakage
- In-band distortions (ICI)
- Reduction of average output power

Missing from this is a derivation of how the PDF of the input signal changes after the nonlinearity. This will be investigated in chapter 3 since it is very important for the blind estimation methods. However, before jumping to that, a short overview of concepts for the mitigation of nonlinear distortions is given.

2.3. Concepts for nonlinearity mitigation

Over time, the high PAPR of OFDM signals and the resulting problems with nonlinear amplifiers have attracted a lot of attention from the research community. Many different algorithms have been developed to mitigate this problem and most of them can be categorized into one of the following classes:

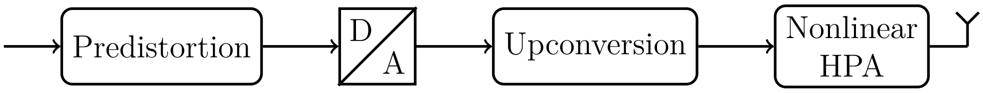

Figure 2.7.: Structure of a transmitter with digital predistortion

- Amplifier linearization through predistortion
- PAPR reduction
- Nonlinearity aware receivers

In the following, the principle of each class is presented.

2.3.1. Linearization through predistortion

The idea of predistortion is to use a preceeding nonlinearity to linearize the power amplifier's characteristic. By careful design it is possible that both nonlinearities combined inflict much less nonlinear distortions on the signals and hence the resulting characteristic is closer to a linear one. Predistortion can be implemented in the digital [D'A01] or analog [YYP00, SKH12] domain but digital predistortion is commonly agreed as more cost-effective since it does not require additional components in the analog frontend. Its basic structure is shown in Figure 2.7.

To design the predistorter, the characteristic of the power amplifier needs to be known. This is problematic since it is an analog component and subject to effects such at aging, production tolerances or circuit temperature which slightly change the electrical properties and therefore the characteristic. Using a static model for the nonlinear power amplifier can result in a residual nonlinearity. For that reason, some authors propose an analog calibration circuitry to account for these effects [BS11].

Figure 2.8 shows an example for the predistortion of a SSPA nonlinearity by means of characteristic inversion. The predistorter is able to extend the linear range of the amplifier significantly. However, the method has a downside. The maximum output level of the actual amplifier can never be surpassed. Therefore, predistortion alone cannot fully mitigate the nonlinear characteristic of amplifiers driven deep into saturation. Instead, residual nonlinearities will remain for very high power peaks. Furthermore, when linearizing a typical characteristic going from a linear to a saturation range, predistorters have the tendency to increase the dynamic range of the signal. Care must be taken not to drive other elements such as D/A-converters

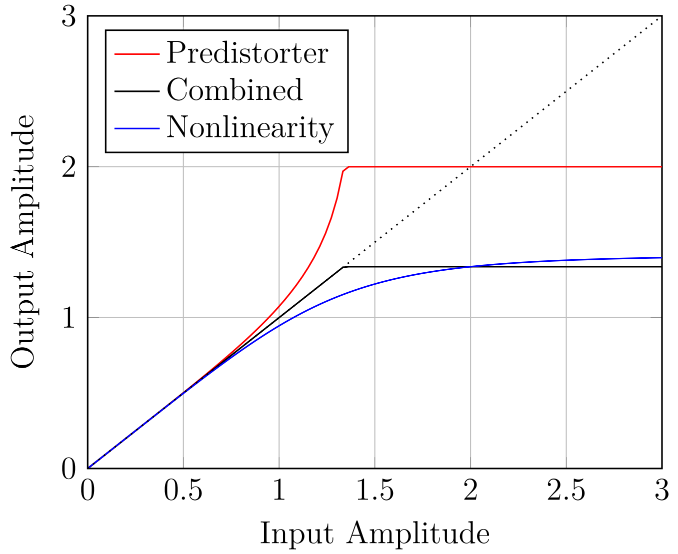

Figure 2.8.: Linearization of a SSPA nonlinearity with $A_{max} = \sqrt{2}$ and $p = 2$, maximum output amplitude of predistorter limited to 2

or mixers into saturation. In Figure 2.8 this is depicted by limiting the maximum output level of the predistorter.

In most cases, predistortion is used to linearize the high-power amplifier in a transmitter. Its big advantage is that it does neither require any changes on the receiver side nor does it generate out-of-band radiation if carefully designed. For these reasons it is a very popular technique for use in cellular base stations where it can significantly improve the power amplifier efficiency [KDJ05].

2.3.2. PAPR reduction

The idea of predistortion is to allow signals with higher PAPR by increasing the effective linear range of the nonlinearity. PAPR reduction techniques approach the problem by reducing the dynamic range of the OFDM signal.

Carrier Rearrangement Recalling Figure 2.2, the PAPR of each individual OFDM symbol is not constant but a random variable and almost all realizations fall into a band of roughly 4-8 dB between highest and lowest PAPR. The reason for the power peaks is the sum of random data weighted complex waveforms that can interfere constructively. Therefore, by changing the random weights it is possible to generate an alternative OFDM symbol with lower PAPR but still carrying the same information.

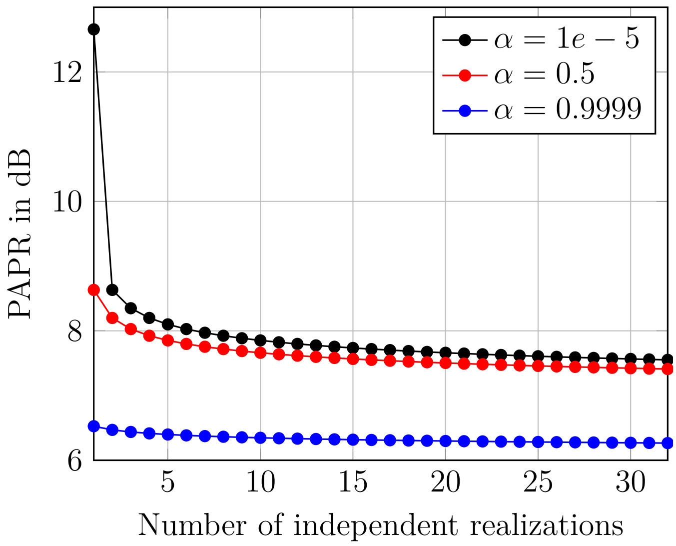

Figure 2.9.: Minimum PAPR of N OFDM symbols, $N_c = 1024$

For example, in the Partial Transmit Sequence method, the symbol is subdivided into carrier groups and each group is subject to a phase rotation which is then optimized to find the symbol with lowest peak power [MH97], reducing the peak by up to 4 dB. A similar method is the Selective Mapping method where N vectors, each containing a random phase rotation are generated and multiplied carrier-wise with the OFDM symbol. Again, the symbol with lowest peak power is transmitted and the method reaches a PAPR reduction of 3-4 dB [BFH96]. Another method employs interleaving the data symbols or data bits to generate a set of OFDM symbols [JT00].

These methods have in common that they do not introduce distortions to the signal. A set of N alternative OFDM symbols is usually generated and the one with the lowest PAPR is transmitted. Assuming that the time domain representations of all generated OFDM symbols can be considered to be independent, Figure 2.9 shows the distribution of the PAPR for different amounts of alternatives. It can be seen that the highest power peaks are reduced by over 4 dB for a very small set of symbols and further reduction requires much larger symbol sets. This matches the cited results that usually reach no more than 3-4 dB PAPR reduction.

The downside of this method is that both transmitter and receiver have to be aware of it. The information about which element of the possible set of symbols has been chosen needs to be transmitted, introducing overhead and reducing the information rate of the system. This transmission needs to happen with especially good error

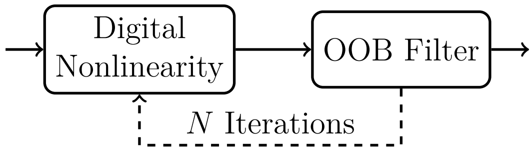

Figure 2.10.: Principle of iterative clipping and filtering

prevention measures since erroneous reception leads to the corruption of a whole OFDM symbol. For some methods, the receiver can estimate the transmitted variant at the expense of higher computational load. Finally, the Selected Mapping (SLM) method requires only very little redundant information which does not even to be specially protected [BMWH01].

Predistortion and PAPR reduction techniques can be combined for an even more efficient usage of the power amplifier [Ryu11].

Iterative Clipping and Filtering Rearrangement techniques reduce the PAPR of OFDM signals without introducing any distortions so that the receiver is not impaired as long as it knows how the rearrangement was done. However, as seen from Figure 2.9, the possible gains are limited and higher gains come at greater expense. The idea of clipping and filtering [Arm02] is to nonlinearly distort the signal but avoid the out-of-band radiation that occurs if an actual power amplifier is the cause of distortion.

The basic principle is shown in Figure 2.10. Firstly, the signal is distorted in the digital baseband. Contrary to digital predistortion, the goal is not to linearize another nonlinearity. Instead, signal peaks are cut off similar to a real amplifier driven into saturation. As shown in Figure 2.6, this induces out-of-band radiation. Therefore, a filter is used to remove all energy from the sidebands. In OFDM systems this is usually done by transforming the signal to the frequency domain and setting the guardband carriers to zero. The filtering will usually increase the PAPR of the resulting signal, diminishing the PAPR reduction. The process is then repeated iteratively until the resulting PAPR has converged to an acceptable level. Furthermore, optimized variants of the algorithm have been presented that reduce the required amount of iterations by means of convex optimization [WL11]. Also, there are approaches to combine this method with rearrangement techniques, see for example [YW06].

The advantages of this method are obvious. The dynamic range of the signal can be limited to a certain desired value and out-of-band radiation is avoided completely.

The downside of this method is that it generates a lot of in-band interference. Therefore, the receiver needs to be aware of the clipping and filtering and the algorithms required to successfully receive a signal which was clipped in this way are usually much more complex than algorithms for unscrambling rearranged OFDM symbols.

2.3.3. Nonlinearity aware receivers

Motivation As stated earlier, according to the Bussgang decomposition, a nonlinearly distorted OFDM time domain signal can be written as follows

$$\boldsymbol{r} = \alpha \boldsymbol{s} + \boldsymbol{n_d} + \boldsymbol{n}\,, \tag{2.41}$$

where $\boldsymbol{n}$ represents Additive White Gaussian Noise (AWGN) with average power σ_n^2 as present in any real radio transmission. The SNR ρ of this signal is given by

$$\rho = \frac{|\alpha|^2 \sigma_s^2}{\sigma_{nd}^2 + \sigma_n^2}\,. \tag{2.42}$$

It can be seen that even for $\sigma_n^2 \to 0$, the resulting ρ will not approach infinity. The result is an error floor. However, this is only true, when an OFDM receiver is used that is not aware of the presence of a nonlinearity and its major noise contribution. As stated earlier, $\boldsymbol{n_d}$ is not actually random, but fully deterministic if $\boldsymbol{s}$ and the nonlinearity $g(s)$ are known. This implies that a specially designed receiver should be able to achieve better performance.

This statement is supported in [ZF05]. In this paper, Zillmann presented a capacity analysis of nonlinearly distorted transmissions. Based on an evaluation of the mutual information between input and output of a memoryless nonlinearity, he showed that the theoretical performance loss is much smaller and that there is no error floor as compared to the case where $\boldsymbol{n_d}$ is assumed as noise.

In his thesis [Zil07], Zillmann provides a comprehensive overview of receiver techniques for nonlinearly distorted OFDM signals ranging from memoryless equalizers, Maximum Likelihood (ML)- and Maximum a-posteriori (MAP)-optimal approaches to iterative receivers based on the Bussgang decomposition. It turned out that for systems with more than $N_c > 64$ subcarriers, the Bussgang-based methods are preferred because they offer low complexity and comparable performance to the ML approaches. For that reason, only this class of receivers is analyzed in this thesis.

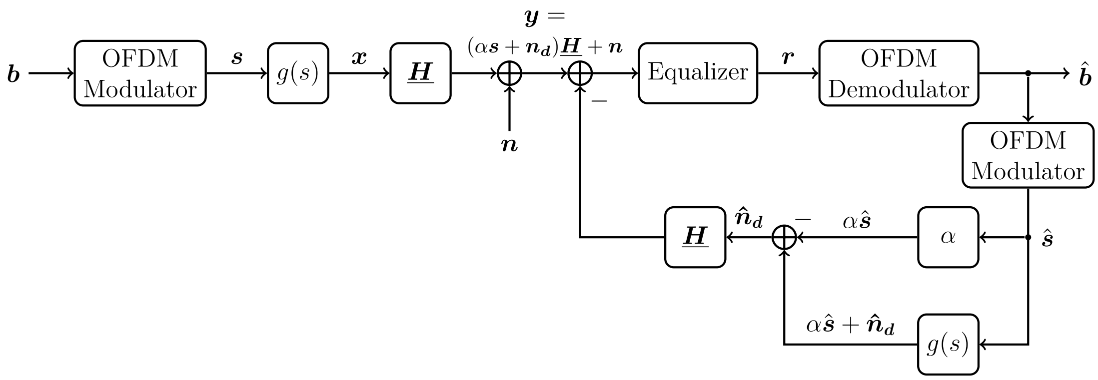

Figure 2.11.: OFDM transmission with Bussgang based decision feedback receiver

Decision Feedback Receiver The idea of a Decision Feedback Receiver (DFR) for nonlinearly distorted OFDM signals has been proposed and analyzed in [THC03]. It is based on the Bussgang decomposition which states that an OFDM time signal $\boldsymbol{s}$ with sufficiently many subcarriers that is subject to memoryless nonlinear distortion $g(s)$ can be decomposed as follows (see (2.34)):

$$\boldsymbol{x} = \alpha\boldsymbol{s} + \boldsymbol{n_d}\,. \tag{2.43}$$

The goal of the DFR is to remove the additive term $\boldsymbol{n_d}$ in order to improve the SNR. While often modeled as noise, $\boldsymbol{n_d}$ is fully deterministic and can be reproduced given that $\boldsymbol{s}$ and the nonlinearity characteristic $g(s)$ are known. Gaining knowledge of the nonlinearity characteristic is investigated in the next chapters and for now, $g(s)$ is assumed as perfectly known.

The basic principle is shown in Figure 2.11. A vector $\boldsymbol{b}$ of source data bits (assumed to be i.i.d. with equal probability of ones and zeros) is modulated using OFDM resulting in the time domain vector $\boldsymbol{s}$. This signal is then nonlinearly distorted, subject to frequency selective fading, modeled by multiplication with the circular convolution matrix $\underline{\boldsymbol{H}}$, and superimposed by AWGN $\boldsymbol{n}$. The signal at the input of the receiver is then given by

$$\boldsymbol{y} = g(\boldsymbol{s})\underline{\boldsymbol{H}} + \boldsymbol{n} = \alpha\boldsymbol{s}\underline{\boldsymbol{H}} + \boldsymbol{n_d}\underline{\boldsymbol{H}} + \boldsymbol{n} \tag{2.44}$$

and in frequency domain by

$$\boldsymbol{Y} = \alpha\boldsymbol{S} \odot \boldsymbol{H} + \boldsymbol{n_d} \odot \boldsymbol{H} + \boldsymbol{N}\,. \tag{2.45}$$

This signal is then equalized, demodulated and detected using standard OFDM techniques. The type of equalizer is not fixed for this receiver but it should be made sure that the complex factor α caused by the nonlinearity is equalized along with the channel. If the channel frequency response $\boldsymbol{H}$ is reconstructed from pilots in the OFDM symbol, this happens implicitely. If the channel is estimated by means of a training sequence that is not subject to nonlinear distortion, this has to be done separately. The result of detection is a vector $\hat{\boldsymbol{b}}$ that is an estimate on the transmitted bits $\boldsymbol{b}$ and may contain errors. These bits are again modulated with exactly the same parameters of the transmitter's OFDM modulator. The resulting $\hat{\boldsymbol{s}}$ is an estimate on the transmit signal $\boldsymbol{s}$ but again might be subject to errors. Using the knowledge of the nonlinearity characteristic $g(s)$ and α, an estimate of the additive distortion $\hat{\boldsymbol{n}}_{\boldsymbol{d}}$ is calculated:

$$\hat{\boldsymbol{n}}_{\boldsymbol{d}} = g(\hat{\boldsymbol{x}}) - \alpha\hat{\boldsymbol{x}} = \alpha\hat{\boldsymbol{x}} + \hat{\boldsymbol{n}}_{\boldsymbol{d}} - \alpha\hat{\boldsymbol{x}} \tag{2.46}$$

After multiplication with the channel matrix $\underline{\boldsymbol{H}}$ the result is subtracted from the input signal of the receiver. In the next iteration, the OFDM receiver will operate on the signal

$$\boldsymbol{y} = \alpha\boldsymbol{s}\underline{\boldsymbol{H}} + (\boldsymbol{n}_{\boldsymbol{d}} - \hat{\boldsymbol{n}}_{\boldsymbol{d}})\underline{\boldsymbol{H}} + \boldsymbol{n} \tag{2.47}$$

and performance will improve if the following inequality holds:

$$\mathrm{E}\left\{|\boldsymbol{n}_{\boldsymbol{d}} - \hat{\boldsymbol{n}}_{\boldsymbol{d}}|^2\right\} < \mathrm{E}\left\{|\boldsymbol{n}_{\boldsymbol{d}}|^2\right\} . \tag{2.48}$$

This process lends itself to iterative application since reducing the power of the additive distortion yields a better estimate of the bits $\hat{\boldsymbol{b}}$. This in turn will result in a better estimation of $\hat{\boldsymbol{n}}_{\boldsymbol{d}}$.

One important question is when (2.48) holds. When the amount of errors becomes too large, the estimate of the additive distortion might be so far off, that subtracting it introduces additional noise into the system. To the best knowledge of the author, an exact analytical answer to this problem is difficult and has not yet been given. Empirical results show that for common amplifier models presented earlier, (2.48) holds even if the backoff is reduced below 0dB which is much worse than most practical scenarios. In low SNR scenarios, the algorithm does not always converge to the best possible result due to non-recoverable errors that are caused primarily by AWGN contributions.

For the optimal case $\hat{\boldsymbol{b}} = \boldsymbol{b}$, the algorithm fully removes the additive distortion $\boldsymbol{n_d}$. The situation is then equal to linear transmission with an SNR loss of $10 \cdot \log_{10}(|\alpha|^2)$ dB. The DFR treats $\boldsymbol{n_d}$ completely as noise. This is not true since $\boldsymbol{n_d}$ depends fully on the transmitted signal $\boldsymbol{s}$ and the nonlinearity $g(s)$ and therefore contains information about $\boldsymbol{s}$ that this algorithm discards. There are algorithms [RZF06] that employ this information for higher gains. However, those algorithms usually suffer from a much higher computational complexity.

Receivers that are aware of nonlinear distortions have a broad spectrum of usage. In systems with predistortion, they can be used to remove residual nonlinearities. When clipping and filtering is employed or the signal is distorted by an actual nonlinearity in the transmitter, these receivers are required as well. And finally, if the nonlinearity is located in the receiver, it is the only way to remove the signal impairments it caused. Compared to predistortion and rearrangement techniques, the main implementation and computational complexity is on the receiver side. Therefore, these receivers can be used when the transmitter is too weak to implement complex PAPR reduction techniques. However, if the transmitter is simply distorting the signal, out-of-band radiation will occur and care must be taken to not violate spectral masks.

The decision feedback receiver is used throughout the rest of this thesis to benchmark the performance of the estimation algorithms which will provide the DFR with the knowledge about the nonlinearity characteristic $g(s)$.

2.4. Conclusions

In this chapter, a short overview of OFDM transmission and reception methods has been given. The statistical properties of OFDM time domain signals have been analysed especially in terms of their PAPR. Furthermore, an overview of the general properties of memoryless nonlinearities has been given and some common models have been presented. It was shown that OFDM signals that are subject to nonlinear distortion suffer from in-band noise and out-of-band radiation. The resulting signals can be decomposed into a scaled variant of the original signal and an uncorrelated additive noise term by means of the Bussgang decomposition.

In the second part of the chapter, an overview of techniques to deal with the high PAPR of OFDM signals has been given. On the transmitter side, those can be separated into predistortion techniques that try to linearize the nonlinearity characteristic and PAPR reduction techniques. The latter can be divided into rearrangement techniques that try to find alternative signal representations with lower PAPR without distorting the signal and clipping and filtering where the signal is nonlinearly distorted on purpose and the out-of-band radiation is removed by filtering. Finally, a nonlinearity aware receiver has been reviewed that is based on the Bussgang decomposition and removes the additive noise term of the nonlinearity.

With the exception of rearrangement techniques, all methods require knowledge of the nonlinearity. Up to now, this was usually assumed as perfectly known which is not true in most realistic cases. The following chapters will deal with novel methods of gaining knowledge about the nonlinearity characteristic.