Ilmenau University of Technology
International Graduate School on Mobile Communications

Faculty of Electrical Engineering and Information Technology
Institute for Information Technology
Electronic Measurement Research Lab

Dirty RF Signal Processing for Mitigation of Receiver Front-end Non-linearity

Dissertation

to obtain a doctoral degree in electrical engineering (Dr.-Ing.)

submitted by

Dipl.-Ing. Michael Grimm

born in Weimar on January 7th, 1986.

Reviewers:	1. Univ.-Prof. Dr.-Ing. habil. Reiner S. Thomä (TU Ilmenau)
	2. Univ.-Prof. Dr. rer. nat. habil. Matthias Hein (TU Ilmenau)
	3. Prof. Dr. Tech. Mikko Valkama (TU Tampere, Finland)
Date of submission:	04/10/2013
Date of defence:	27/06/2014
URN:	urn:nbn:de:gbv:ilm1-2014000134

Acknowledgements

The help of various people have made this thesis possible or complete. First, I would like to thank my supervisor Prof. Reiner Thomä for the continuous support, encouragement, and his human way throughout the last years. Without him, I would not have chosen this topic and this thesis would not exist. Second, I would like to thank my master students Diego Dupleich and Florian Schlembach for their valuable and fruitful cooperation and their contribution to this thesis through their master theses and student assistant jobs. Third, I thank all colleagues of the department "Electronic Measurement Research Lab" for discussions on all matters and for the friendly working atmosphere. I would like to mention especially my office colleagues Gerd Sommerkorn, Peter Rauschenbach, Robert Müller, and Stephan Häfner for their friendly way and support in the everyday work. Throughout the doctoral studies, feedback is indispensable to successfully complete a PhD. I would like to thank Prof. Matthias Hein, Wim Kotterman, Ralf Herrmann, and Rajesh Sharma for proof-reading my papers and thesis. I thank Kurt Blau for fruitful discussions, especially during lab experiments.

The work presented in this thesis has been carried out within the "International Graduate School on Mobile Communications", that is supported by the German Research Foundation (DFG GRK1487). I would like to thank Prof. Andreas Mitschele-Thiel for giving me the chance to work on an interesting research project, and Mirko Kirschkowski for his kind administrative support throughout my doctoral studies.

The cooperation with external partners has made this thesis complete. I feel special gratitude for Prof. Mikko Valkama for reviewing my thesis and giving me the opportunity to visit Tampere University of Technology in Finland for three months. This stay abroad enriched my social and technical skills significantly. My special thanks go to Markus Allén and Jaakko Marttila for the greatest cooperation one can think of, even in the time after my visit, where we prepared a joint journal article for IEEE Trans. on Microwave Theory and Techniques [GAM+14]. Furthermore, I would like to thank Reda Zemmari from Fraunhofer FKIE for the great cooperation in the field of GSM and passive radar. Besides, I am very grateful for our guest researcher Andrey Budilov for his cooperation and for being a great host during a vacation in Russia in 2011.

During my doctoral studies, I found many new friends who are colleagues in the Graduate School. I would like to thank Alexander Krah, André Puschmann, Noman Murtaza, Anastasia Lavrenko, Tobias Simon, Christopher Schirmer, and Jonas König for their cooperation at work and the nice time at workshops, conferences, and any social event beside work. You made the last years to an unforgettable time in my life.

Finally, I would like to thank my parents, my brother and sister-in-law, and my grandparents for their everlasting support and encouragement on my way. Furthermore, I have to thank my girlfriend Katharina for her love, trust and motivation throughout the last years since my diploma time.

Ilmenau, October 2013 Michael Grimm

Zusammenfassung

Moderne drahtlose Kommunikationssysteme stellen hohe und teilweise gegensätzliche Anforderungen an die Hardware der Funkmodule, wie z.B. niedriger Energieverbrauch, große Bandbreite und hohe Linearität. Die Gewährleistung einer ausreichenden Linearität ist, neben anderen analogen Parametern, eine Herausforderung im praktischen Design der Funkmodule. Der Fokus der Dissertation liegt auf breitbandigen HF-Frontends für Software-konfigurierbare Funkmodule, die seit einigen Jahren kommerziell verfügbar sind. Die praktischen Herausforderungen und Grenzen solcher flexiblen Funkmodule offenbaren sich vor allem im realen Experiment. Eines der Hauptprobleme ist die Sicherstellung einer ausreichenden analogen Performanz über einen weiten Frequenzbereich. Aus einer Vielzahl an analogen Störeffekten behandelt die Arbeit die Analyse und Minderung von Nichtlinearitäten in Empfängern mit direkt-umsetzender Architektur. Im Vordergrund stehen dabei Signalverarbeitungsstrategien zur Minderung nichtlinear verursachter Interferenz – ein Algorithmus, der besser unter "Dirty RF"-Techniken bekannt ist. Ein digitales Verfahren nach der Vorwärtskopplung wird durch intensive Simulationen, Messungen und Implementierung in realer Hardware verifiziert. Um die Lücken zwischen Theorie und praktischer Anwendbarkeit zu schließen und das Verfahren in reale Funkmodule zu integrieren, werden verschiedene Untersuchungen durchgeführt. Hierzu wird ein erweitertes Verhaltensmodell entwickelt, das die Struktur direkt-umsetzender Empfänger am besten nachbildet und damit alle Verzerrungen im HF- und Basisband erfasst. Darüber hinaus wird die Leistungsfähigkeit des Algorithmus unter realen Funkkanal-Bedingungen untersucht. Zusätzlich folgt die Vorstellung einer ressourceneffizienten Echtzeit-Implementierung des Verfahrens auf einem FPGA. Abschließend diskutiert die Arbeit verschiedene Anwendungsfelder, darunter spektrales Sensing, robuster GSM-Empfang und GSM-basiertes Passivradar. Es wird gezeigt, dass nichtlineare Verzerrungen erfolgreich in der digitalen Domäne gemindert werden können, wodurch die Bitfehlerrate gestörter modulierter Signale sinkt und der Anteil nichtlinear verursachter Interferenz minimiert wird. Schließlich kann durch das Verfahren die effektive Linearität des HF-Frontends stark erhöht werden. Damit wird der zuverlässige Betrieb eines einfachen Funkmoduls unter dem Einfluss der Empfängernichtlinearität möglich. Aufgrund des flexiblen Designs ist der Algorithmus für breitbandige Empfänger universal einsetzbar und ist nicht auf Software-konfigurierbare Funkmodule beschränkt.

Abstract

Today's wireless communication systems place high requirements on the radio's hardware that are largely mutually exclusive, such as low power consumption, wide bandwidth, and high linearity. Achieving a sufficient linearity, among other analogue characteristics, is a challenging issue in practical transceiver design. The focus of this thesis is on wideband receiver RF front-ends for software defined radio technology, which became commercially available in the recent years. Practical challenges and limitations are being revealed in real-world experiments with these radios. One of the main problems is to ensure a sufficient RF performance of the front-end over a wide bandwidth. The thesis covers the analysis and mitigation of receiver non-linearity of typical direct-conversion receiver architectures, among other RF impairments. The main focus is on DSP-based algorithms for mitigating non-linearly induced interference, an approach also known as "Dirty RF" signal processing techniques. The conceived digital feedforward mitigation algorithm is verified through extensive simulations, RF measurements, and implementation in real hardware. Various studies are carried out that bridge the gap between theory and practical applicability of this approach, especially with the aim of integrating that technique into real devices. To this end, an advanced baseband behavioural model is developed that matches to direct-conversion receiver architectures as close as possible, and thus considers all generated distortions at RF and baseband. In addition, the algorithm's performance is verified under challenging fading conditions. Moreover, the thesis presents a resource-efficient real-time implementation of the proposed solution on an FPGA. Finally, different use cases are covered in the thesis that includes spectrum monitoring or sensing, GSM downlink reception, and GSM-based passive radar. It is shown that non-linear distortions can be successfully mitigated at system level in the digital domain, thereby decreasing the bit error rate of distorted modulated signals and reducing the amount of non-linearly induced interference. Finally, the effective linearity of the front-end is increased substantially. Thus, the proper operation of a low-cost radio under presence of receiver non-linearity is possible. Due to the flexible design, the algorithm is generally applicable for wideband receivers and is not restricted to software defined radios.

Contents

1 Introduction

The demand for broadband wireless communications is increasing every year and continuously accelerates the pace of engineering and science. There is an exponential growth in data traffic due to the increasing number of users subscribing broadband packages and the emergence of new devices, data services, and applications. A current trend is to apply multi-mode and multi-band transceivers based upon simple radio frequency (RF) front-ends that integrate a variety of wireless standards. Conventional communication systems have fixed radio parameters such as carrier frequency, bandwidth, and modulation scheme. Instead, software defined radio (SDR) technology is experiencing increasing popularity where most of the signal processing is performed in the digital domain. Thereby, simple and flexible RF hardware can be realised and applied in many different use cases, operating at different frequencies, power levels, and with different waveforms. Suitable areas of application for this technology lie not only at the mobile terminal side, where costs have to be kept low for mass production, but also the base station side that have to accommodate for various wireless systems in a large frequency range. For instance, cognitive radio (CR) has been proposed as a novel concept to solve the spectrum scarcity problem by utilising flexible SDR hardware. Most of the CR research deals with theoretical analyses or is based on simulations that often do not consider the physical limitations of the radio. Since SDR platforms became commercially available, the community is now focusing on real implementations and practical evaluation of proposed algorithms and concepts. Thereby, practical limitations of transceiver electronics in RF front-ends for SDR turned out to cause a performance degradation of the radio. Moreover, future emerging wireless communication systems demand for higher bandwidth, higher-order constellation diagrams, and multiple antennas. These increases in modem performance, bandwidth, and carrier frequency provoke dirt effects that have reached a new problem level. In brief, systems will definitely become more sensitive to front-end non-idealities. Therefore, it is worthwhile to analyse these effects and to look for solution strategies.

1.1 Research Problem

Novel concepts in communications and radar have high demands on the RF transceiver electronics. Among all, wide bandwidth and frequency-agility play a crucial role. In addition, radios should operate under extreme dynamic range conditions, allowing analysis and synthesis of different waveforms with different power levels. In fact, many of these requirements reach far beyond the state-of-the-art technology. These demands may be fulfilled by measurement and prototype devices, but not by practical terminals in mass production. The analogue stages of a radio front-end are inherently imperfect and are likely to cause errors that deteriorate the signal quality. Add to this, RF technology is already pushed to its operation boundaries, making it to a serious problem for future technology [Sch08]. In general, there are different RF impairments, such as mirror-frequency interference due to in-phase/quadrature (I/Q) imbalance, non-linear distortions, timing jitter, and phase noise [VSH10].

The focus in thesis is on non-linear distortions at the receiver side as a subset of the impairments. Nowadays, simple direct-conversion receivers (DCRs) are employed that have a general lack of selectivity at the intermediate frequency (IF) [Rou09]. Weak and strong signals may be simultaneously present at the RF input, causing high demands on the receiver dynamic range and linearity. In traditional narrowband RF receivers, strong signals are greatly suppressed by surface acoustic wave (SAW) filters. Typically for SDRs, there is no pre-selection filter at the front-end input at all, allowing strong interferers to enter amplification and mixing stages within the whole receiver bandwidth. Hence, non-linear distortions due to these strong signals are easily generated by non-linear receiver components, such as the low-noise amplifier (LNA), the mixer, and additional baseband (BB) amplification stages. Finally, distortion products may fall in free frequency bands or hit weak useful signals.

In contrast to the *transmitter* side, the perspective on non-linear distortions at the *receiver* side is fundamentally different. At the transmitter, there is only one well-known signal to be transmitted and interference is created in adjacent bands due to transmitter non-linearity. However, non-linear distortions at the receiver are more challenging due to presence of multiple unknown signals, with different power levels and dynamics due to specific wave propagation. In other words, the original input signal is unknown at the receiver, whereas desired and distorted output can be easily measured at the transmitter. In particular, highly unbalanced signals appear in frequency-division multiple access (FDMA) systems and finally sum up in the receiver. Thereby, distortion products are easily generated by different unknown strong signals. On the other hand, crucial interference conditions seldom appear

in conventional radio systems, such as in cellular networks, but are more probable in CR or in coexistence scenarios with heterogeneous radio systems. From the mitigation point of view, handling of non-linear distortions at the receiver always have to start from a guess of the input compared to pre-distortion techniques at the transmitter [BCG+10].

1.2 Solution Strategies

The main motivation is to consider non-linear distortions, created by the receiver front-end non-linearity, in the operation of the radio. Finally, mitigating the deteriorating effects and enhancing the reliability of the radio are desirable. Thereby, only partial mitigation of non-linear distortions compared to full cancellation or compensation might be achieved.

There are several possible solutions that can be classified into three main categories:

- Optimising the front-end design,
- Preventing the generation of non-linear distortion, or
- Mitigating non-linear distortion.

Following the first approach, highly linear components should be employed in the front-end design so the receiver can even cope with high-power input signals. By using the second approach, specific interfering signals are to be modified or suppressed to prevent non-linear distortion products created by them during reception. The cancellation might be achieved through attenuation, analogue filtering, analogue subtraction of an interferer replica, or spatial filtering using an antenna array. Alternatively, waveforms may be adjusted towards relaxed linearity requirements for the front-end, such as peak-to-average power ratio (PAPR) reduction techniques. The third approach, being on focus in this thesis, includes all pre- or post-correction techniques to clean the signal from distortions after they have already occurred. To this end, digital signal processing is adopted, e.g. by analysing specific training symbols in the transmit signal, or by using a reference model for distortion regeneration and subsequent subtraction. This approach is better known as dealing with "Dirty RF" effects in the digital domain, a term, that has been first coined in [FLP+05]. However, several synonyms can be found in literature, such as RF impairment mitigation, or alleviating RF imperfections. Digital mitigation techniques are preferred, especially due to their easy integration in the digital back-end. Moreover, digital circuits are growing faster and are less sensitive to process variations compared to their analogue counterparts [HB08; Chi11]. Up to now, analogue front-end and digital back-end have been designed independently, assuming that each block works ideally. Digital mitigation of non-linear distortions can relax the

linearity and dynamic range requirements of the front-end, enabling the design of low-cost receivers. That requires a close cooperation between RF and signal processing engineers.

1.3 State-of-the-Art and Research Contribution

Extensive research has been undertaken on pre-distortion techniques for mitigating power amplifier (PA) non-linearity at the transmitter, such as in [BBC88; BCG+10]. However, there is also some prior work considering non-linear distortions at the receiver side. Proposed solutions vary in the complexity of implementation and also in the use cases considered. Mitigating non-linear distortions by means of digital signal processing has been first proposed in [VSHGA+06], and subsequently followed in [KH08b; ZMS09; AMV10b; GSH+12a], among others. In [VSHGA+06], a reference model of the non-linear receiver is applied in the digital domain to regenerate distortion products, and to subtract them from the received signal using adaptive filtering. Custom receiver designs are proposed in [KH08b] and [ZMS09], to overcome the limitations of the entirely digital approach. Sometimes, pilot signals are integrated in the transmit signal, enabling the subsequent estimation of the receiver non-linearity at symbol level [DS07].

However, there are some issues in the current state-of-the-art that have not been considered so far:

- Specifics of DCRs,
- Application to modulated signals,
- Impact of fading conditions, and
- Real-time implementation.

The main contributions of this thesis are seen in the consideration of these issues that bridge the gap between theory and practical applicability of digital mitigation algorithms. Real front-end signals are analysed and processed throughout the thesis. First, an experimental characterisation of the non-linear behaviour and the effects of non-linear distortions has been conducted based on a wideband receiver for SDR. Second, a state-of-the-art mitigation algorithm has been extended for DCRs and applied to real communication signals. Thereby, the performance of "Dirty RF" algorithms was analysed in real-world scenarios. Third, an advanced reference model for the non-linearity of a DCR has been developed that jointly considers distortions created at RF and BB. Finally, practical issues have been covered, such

as the impact of fading conditions on the performance of the algorithm, or implementation issues for embedding the algorithm onto a field-programmable gate array (FPGA).

Three main use cases of the algorithm have been investigated in more detail:

- Spectrum sensing in CR,
- GSM downlink reception, and
- GSM-based passive radar.

1.4 Organisation of the Thesis

The introduction is followed by six main chapters. Chapter 2 deals with characterisation of receiver non-linearity, including sources, modelling, and measurement-based analysis. In Chapter 3, a detailed overview of possible mitigation techniques is given. Concepts on cancelling distortion-producing signals are briefly discussed. Then, the digital feedforward mitigation algorithm, on focus in this thesis, is discussed in detail in Chapter 4. Practical aspects of this technique, namely the impact of fading conditions and the challenges of its real-time implementation, are discussed in the Chapters 5 and 6. Chapter 7 presents three use cases, underlining the practical relevance of the digital feedforward mitigation. Finally, conclusions and an outlook on future work are given in Chapter 8.

2 Non-linear Distortions

This chapter deals with typical front-end architectures adopted in state-of-the-art receivers as well as with sources of non-linearity. In addition, BB models of non-linear front-ends are developed that are used later in the mitigation part. To end the chapter, results of an experimental characterisation of a typical SDR platform with narrowband and wideband test signals are presented.

2.1 Typical Receiver Architectures

There is a variety of receiver architectures that are classified by the number of frequency conversions to translate the signal from RF to BB. Most common types are heterodyne, homodyne, or low-IF, which have benefits and problems in certain applications [HS11]. Among all, the homodyne or zero-IF architecture fits best for SDR. Problems of this simplified front-end have been better understood in recent years, thereby enabling the design of a low-cost front-end [HS11]. First, most common architectures with their pros and cons are presented. Finally, the concept of SDR is picked up and suitable candidate architectures are discussed.

2.1.1 Heterodyne Architecture

The heterodyne architecture, being the conventional design, first translates the RF signal to a lower IF by using a down-conversion mixer [GCB08]. Thereby, possible out-of-band interferers are filtered at IF, hence, a high selectivity is achieved. Therefore, less selective filters may be adopted before conversion to the digital domain. Add to this, there is a minimum received signal power level that the receiver can resolve above its noise level. For example, adding a pre-selection filter in front of an amplifier adds noise and reduces the sensitivity. Hence, there is always a trade-off between sensitivity and selectivity in designing a heterodyne receiver [Raz97]. Typically, a dual-IF architecture is utilised as illustrated in Figure 2.1. A high first IF is used to easily perform image rejection filtering by the band-pass filter BP_2 that can have a lower quality factor at high frequencies. A second low IF

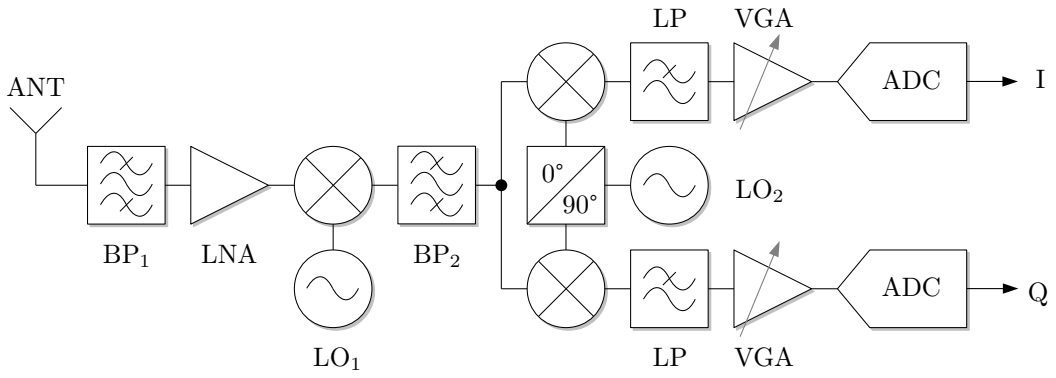


Figure 2.1: Block scheme of a typical dual-IF heterodyne receiver.

is used to relax requirements for the channel selection filter. At the second IF, quadrature down-conversion is adopted, that separates the signal into in-phase component (I) and quadrature component (Q) through multiplication with two sinusoidal frequencies that are exactly 90° out of phase. After digitisation by two analogue-to-digital converters (ADCs), demodulation and detection are performed at BB by means of digital signal processing (DSP). Alternatively, digitisation may happen already at IF, moving the quadrature mixing and lowpass filtering into the digital domain.

The heterodyne architecture can be easily adapted to many different standards, achieving very good sensitivity and selectivity. However, this structure requires a large number of external components, therefore, making integration for battery-powered terminals difficult. Moreover, this front-end is less flexible and thereby not particularly well suited for SDR purposes.

2.1.2 Homodyne or Zero-IF Receiver

The homodyne receiver, shown in Figure 2.2, directly converts the RF signal down to BB, i.e. direct current (DC) frequency (zero-IF or DCR) [Raz97; GCB08]. Instead of bandpass filtering at IF, all interferers are suppressed by a lowpass at I and Q branch. This receiver architecture is simple, cheap, easy to integrate, and causes only a low power consumption. Therefore, it became very popular in communications. However, the simplicity of this front-end also causes some major drawbacks, such as DC offsets, even-order distortions, I/Q mismatch, and flicker noise. DC offset is caused by mixing of local oscillator (LO) leakage with the LO itself, either by the feedthrough from the mixer LO port to the mixer input, or by a large signal leaking from the LNA to the LO port. Distortions in DCRs will be discussed in the next section in more detail. Any amplitude or phase mismatch between the

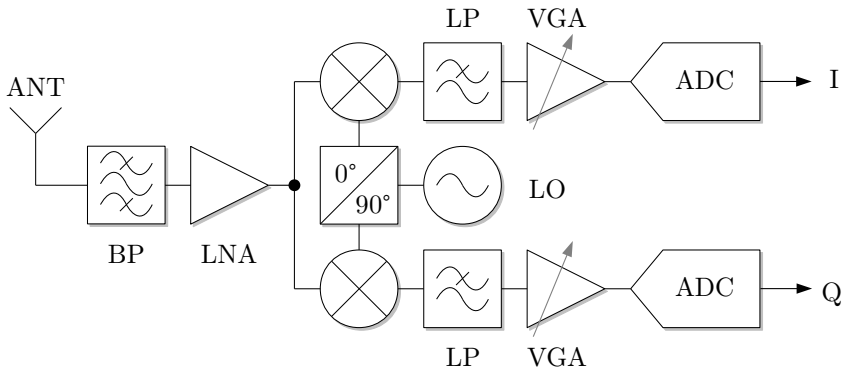


Figure 2.2: Block scheme of a homodyne (zero-IF) receiver.

two receiver branches for I and Q causes interference that may fall on top of other signals. In addition, flicker noise or $1/f$ noise, typically negligible in heterodyne receivers, increases the overall noise figure of the receiver, especially due to the low signal frequencies around DC at the stages after the mixer. However, many of these issues have been solved by integrated circuit technology in the past years.

2.1.3 Low-IF Receivers

An interesting approach is to combine heterodyne and homodyne architecture, where the received signal is down-converted to a low non-zero IF in range of several hundred kHz up to several MHz. After amplification at BB, the signal is digitised by an ADC. The final down-conversion to DC is performed digitally by using a numerically controlled oscillator (NCO) and a complex multiplier, as illustrated in Figure 2.3. As the signal is centred around a low

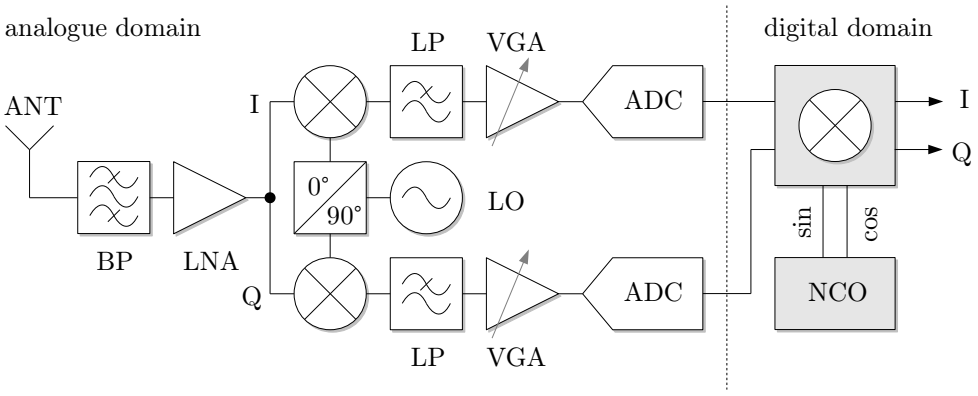


Figure 2.3: Block scheme of a low-IF receiver.

IF, DC offset is negligible and can be easily filtered by averaging [LHG+10]. Beside the

preferences as in the zero-IF architecture, a large image rejection ratio (IRR) is required for this receiver topology.

There are various other receiver types that are out of scope here, such as bandpass sampling or direct RF sampling, that apply discrete-time analogue signal processing by subsampling [GCB08]. These concepts can save the mixer through a subsampling operation, but impose very tough bandpass filter specifications in order to avoid aliasing. Hence, these architectures are not suited for flexible SDR front-ends that are on focus in this thesis.

2.1.4 Software Radio and Software Defined Radio

In 1995, Mitola proposed a novel concept of a software radio (SR) as shown in Figure 2.4 [Mit95; GCB08]. The idea is to move the digitisation as close as possible to the antenna

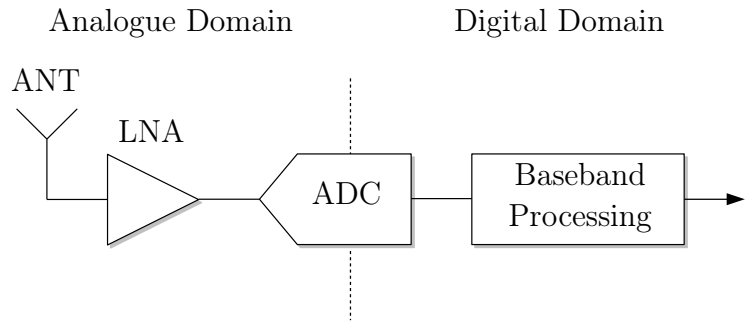


Figure 2.4: Ideal SR architecture proposed in [Mit95].

without any frequency conversion. Thus, most of the signal processing tasks are performed in the digital domain. The remaining analogue components are the LNA with a built-in automatic gain control (AGC) and the ADC. This receiver design is most flexible as it can operate over a wide range of frequencies and process arbitrary waveforms. However, the SR concept is far ahead from practical implementation with the state-of-the-art technology. For instance, designing an omnidirectional antenna that serves a bandwidth ranging over several decades of frequencies is challenging. In addition, the requirements for the ADC are very difficult to achieve, such as the large input bandwidth, high sampling rate, and high dynamic range, especially when considering direct RF sampling according to the Nyquist theorem.

A practical implementation of the SR concept with today's technology is represented by a SDR that keeps the frequency conversion stage and some of the analogue signal processing. Eventually, the analogue RF front-end's components are designed in a way that they are reconfigurable by software, therefore it is called a software defined radio [Rou09]. In contrast

to traditional single-mode front-ends, a SDR may adapt itself to various air standards and waveforms at different frequencies, with flexible channel bandwidth, and channel conditions. To realise the large operating bandwidth, multiple receive chains may be required, each working in different frequency bands. The best candidate to implement a SDR is the zero-IF or low-IF topology due to their reconfigurability over a broad frequency range. Taking the heterodyne architecture, no common IF will be found that fits for all input frequency bands [GCB08]. The reconfigurable parameters of an SDR are, e.g., the carrier frequency, the channel bandwidth (sampling rate), the amplifier gain, and the filter characteristics. Thereby, the SDR can fulfil the requirements of different standards.

In contrast to the receiver block schemes illustrated in Figure 2.2 and Figure 2.3, the pre-selection bandpass between the antenna and the LNA is omitted in typical SDR front-ends in favour of the flexibility. Typically, a SAW filter is employed as a bandpass, filtering out most of the harmful out-of-band interference. However, this filter has a fixed centre frequency and is thereby very application-specific, being a major disadvantage in SDR. Tunable bandpass filters are very difficult to implement and are not available off the shelf [VSHGA+06; ZMS09]. The missing pre-selection in such front-ends is a critical issue as it reduces the receiver immunity to blockers, thereby dramatically increasing the linearity requirements for the following amplification and mixing stages. Add to this, aliasing effects in the down-conversion due to signals at mirror frequency are typically accepted in favour of flexibility. The only filtering remaining in the analogue domain is performed by the lowpass at BB. However, the analogue front-end components have, in any case, a limited bandwidth. These are, for example, the input bandwidth of the LNA and the IF bandwidth of the mixer.

In 1999, Mitola coined the concept of CR that employs the SDR technology [MM99]. Whereas a SDR targets certain standards and their allocated bands, a CR can be seen as an extension that adds cognition and makes adjustments of the radio based on algorithms in a cognitive engine [Raz09; Hay05; Law10]. The CR use case is discussed in detail in Section 7.1.

2.2 Sources of Distortions

Non-linear distortions can appear at the transmitter or receiver side of a communication system, while the receiver side is on focus in this thesis. Electrical systems can be classified into linear and non-linear systems that either exhibit memory or not [VR03]. In mathematics, a non-linear system is given if the principle of superposition is violated, or

if there is no proportional relationship between the input and output. Figure 2.5 depicts examples of different transfer functions, a linear and two non-linear curves derived from different input-referred third-order intercept points (IIP3s), a measure for the strength of non-linearity. The lower the value of IIP3, the stronger is the non-linearity. A linear device

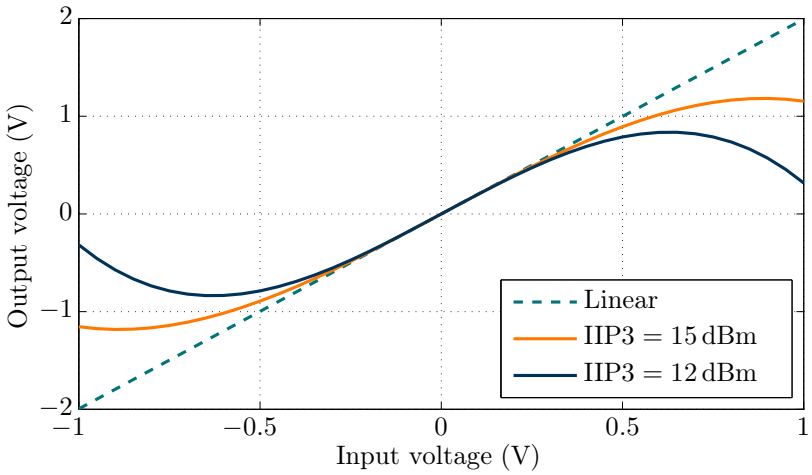


Figure 2.5: Linear and non-linear transfer functions.

outputs a voltage directly proportional to the input (green dashed line in Figure 2.5). The output voltage is two times higher than the input, indicating an amplification factor of 2 (equals 6 dB gain). A non-linear device, as illustrated by the orange and blue curves, operates as a linear device under small-signal conditions. However, with increasing input signal strength, the gain decreases due to saturation effects in the electronics. This phenomena is also called compression and results in a non-linear function [Ken00]. In fact, real circuits do not exhibit such an exaggerated non-linear behaviour as indicated by the blue curve in Figure 2.5. It is indeed a result of the limited order of the non-linearity model used for this simulation. Typically, a non-linear device is going from linear behaviour ($[-0.25, +0.25]$ V), to mildly non-linear ($\pm[0.25, 0.5]$ V), and then to strongly non-linear behaviour (± 0.5 V onwards). Amplifiers manifesting such a behaviour are called "linear", i.e. the amount of distortion can be reduced when the input power is decreased [Sch09]. This is typical for class A and AB amplifiers, but not for switching-mode amplifiers that are designed for highly efficient operation.

Non-linear behaviour is caused by parasitic effects in any active analogue circuit, such as a diode or a transistor. On the contrary, passive elements such as resistors, capacitors, and inductors are linear components. However, cables, antennas, or connectors may cause non-linear junctions under special circumstances, better known as passive intermodulation [Anr12]. From the circuit design point of view, non-linearities are typically annoying as they

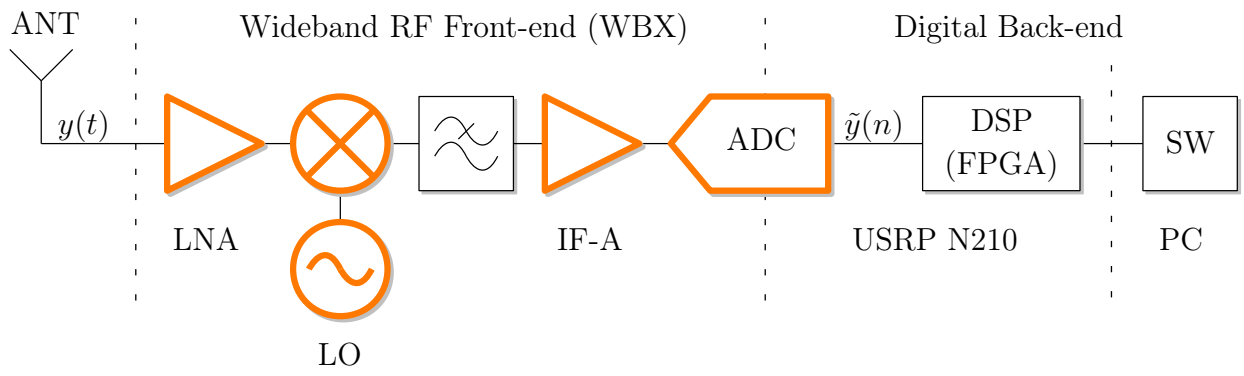


Figure 2.6: Block scheme of a typical SDR receiver, taking the example of USRP N210+WBX [Ett].

distort electrical signals, but can be sometimes useful for certain applications, such as for designing a mixer by using generated harmonics of the LO.

Now, sources of distortions in a typical RF front-end for SDR are discussed, taking the example of Universal Software Radio Peripheral (USRP) version N210 with a wideband RF front-end (WBX) [Ett]. Details on this receiver, that is used for many of the RF measurements in this thesis, are given in Appendix B. Figure 2.6 depicts a block scheme, where all components with non-linear characteristics are highlighted. The WBX front-end is built as a low-IF architecture without SAW pre-selection filter, converting the RF signal down to a low user definable IF (see also Appendix B). Non-linear components in this front-end are the LNA and the IF amplifier, the mixer and the LO, as well as the ADC.

The LNA provides substantial gain of the RF signal and fixes the noise floor [HB08]. As a classical amplifier, it suffers from gain compression under large-signal excitation and causes non-linear distortions.

The mixer translates the RF signal to BB with a non-linear conversion characteristic. A single frequency generated by the LO is required for the mixing process. However, a non-ideal LO may also create harmonics beside the fundamental frequency that can mix with potential interferers, if the receiver bandwidth is wide enough [Raz97; Raz09]. This effect is negligible for higher harmonics as their magnitude is inversely proportional to the frequency.

The BB or IF amplifier is used for optimal analogue/digital (A/D) conversion and behaves in a same non-linear manner as the LNA. However, in terms of linearity performance, the BB amplifier is typically worse than the RF amplifier (LNA).

The ADC is responsible for digitisation of the analogue BB signal, however, suffers from non-linearity and other impairments, such as quantisation noise, sampling clock off-

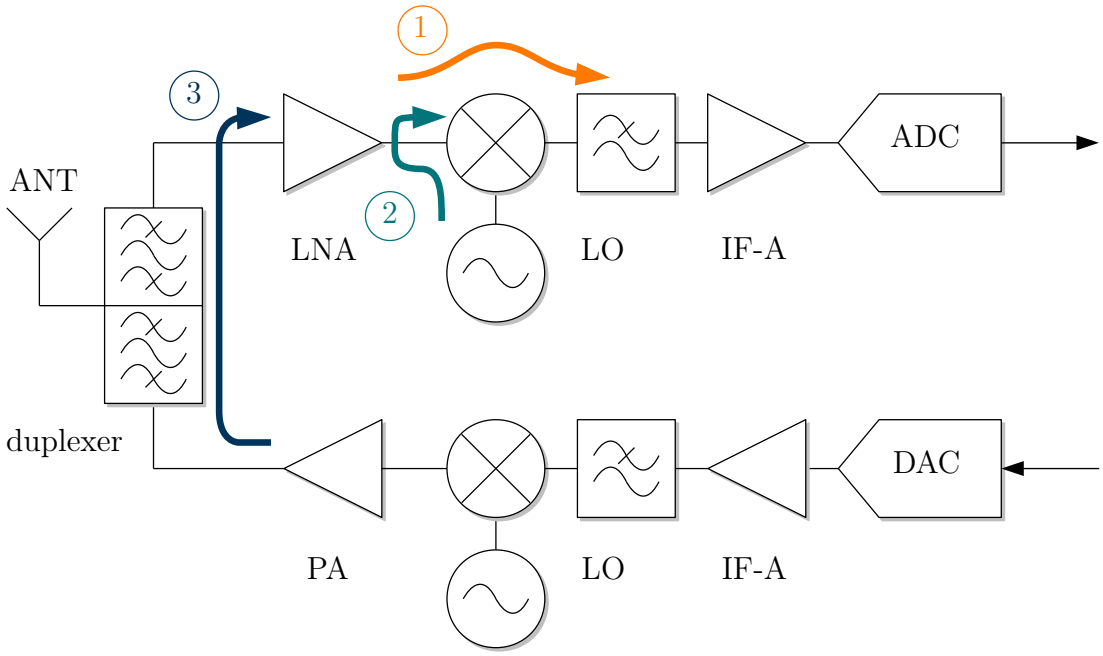


Figure 2.7: Leakage effects in DCR architectures as another source of distortions.

set, and sampling jitter [AMV10a]. Non-linearities are caused by clipping, as well as integral non-linearity (INL) and differential non-linearity (DNL), denoting unintentional deviations of the ideal quantisation levels. Whereas the DNL denotes the relative difference between the actual and the ideal code, the INL means the difference between the actual and ideal transition threshold (integral of DNL). Clipping distortions (CDs) are created due to improper signal conditioning, i.e. the input signal exceeds the full scale voltage of the ADC.

Another source of distortions are caused by leakage effects, especially in DCRs. Figure 2.7 depicts three different phenomena: LNA-mixer feed-through ①, LO-mixer feed-through ②, and duplexer feed-through ③ [KAV13; KAV14; CGH+11; Raz97]. Considering an ideal mixer, low-frequency even-order distortions produced by the LNA are translated to high frequencies and become unimportant. However, they can feed through BB without mixing due to circuit asymmetry [CGH+11; Rou09]. In addition, DC offsets may be generated due to feed-through between LO and mixer input [Raz97]. Moreover, a strong transmit signal can leak into the receiver path via the duplexer that typically switches the antenna between transmitter or receiver. This signal gets further amplified by the LNA and may cause additional distortions, especially in frequency division duplexing (FDD) operation [Rou09]. Although leakage phenomena are side effects, they are important to know when dealing with DCR architectures and may contribute to the overall distortion seen in the digitised BB signal.

Besides, scaling in technology promotes the non-linear effects by reducing the dimensions of transistors. Thus, decreasing supply voltages limit the dynamic range and increase the level of non-linearities [HB08]. Likewise, behaviour of the components is less predictable due to increased variability in the process.

Above all, non-linear circuits can be either with or without memory. Memory effects of the receiver components can have a strong impact on the nature of non-linear distortions [ZWC10; Rou09]. They are generated by inductors and capacitors that store energy, as well as by their frequency-dependent impedance. As a result of memory, node voltage and current depend not only on the instantaneous input, but also on historical signals. That is, the output of the component depends on the current input and the previous input. The higher the bandwidth of the input signal, the higher the memory effects will raise. Memory effects can be distinguished into two types: electrothermal memory effects appearing at low modulation frequencies below 100 kHz, and electrical memory effects appearing at modulation frequencies of above some MHz [VR03].

2.3 Behaviour of Non-linear Receivers

This section discusses the effects of receiver non-linearity and their metrics for describing the strength of non-linear distortion.

2.3.1 Effects

General results of circuit non-linearity are twofold. First, new frequency components are generated by the non-linear device, thereby adding unwanted signal energy to the received signal [Ken00]. Second, there is an amplitude-dependency of the fundamental signal gain [VR03]. Phenomena of non-linear systems can be classified in *order* and *degree* [VR03]. The higher the order of non-linearity, the more non-linear signal content will be generated. Degree is a property of the non-linear device and defines the shape of the non-linearity. Order is related to the amplitude and frequency of the distortion terms, that depend on the degree of non-linear device and the amplitude and frequency of the input. For instance, the order of a distorted signal can be higher than the degree of the device in case the input was already distorted, e.g. in cascaded non-linear devices in a DCR chain (see Section 2.5.5). Typically, non-linear devices are classified into linear, mildly non-linear, and strongly non-linear according to their degree of non-linearity (see also Figure 2.5) [Sch09].

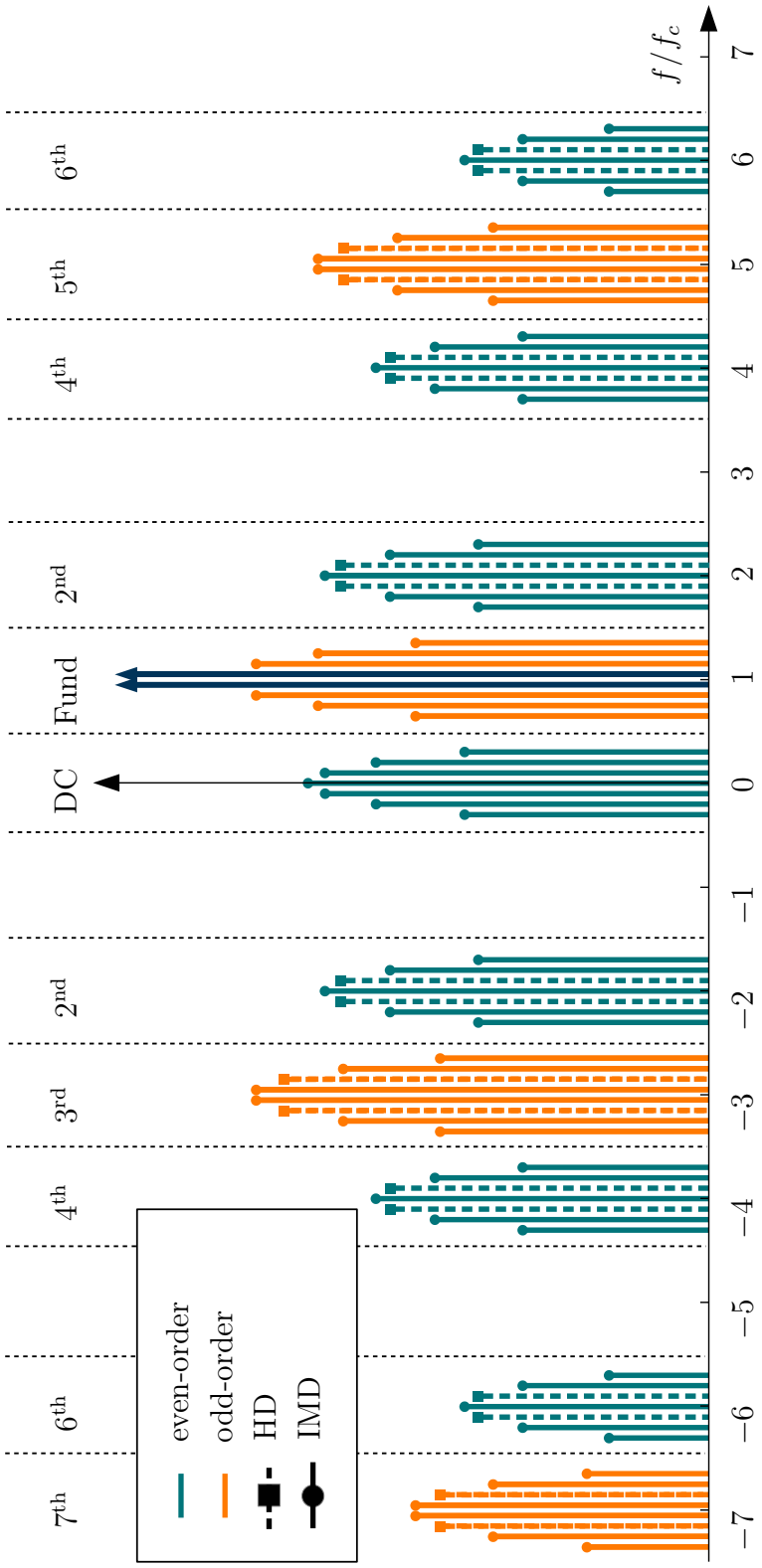


Figure 2.8: Spectral components induced by non-linear device with degree 7 and two-tone input.

Five different types of non-linear distortions at the receiver can be distinguished and are now discussed in detail. Figure 2.8 illustrates most of the different distortions induced by non-linear device of degree 7 and with two-tone input. In contrast to typical figures in textbooks, even and odd-order distortions are illustrated at complex BB after passing a low-IF receiver with quadrature mixer. The two-tone input signal, the blue arrows at $f/f_c = 1$, are assumed to be ideally converted to BB and then distorted by a BB amplifier of degree 7 in the I and Q branch, respectively. In addition, DC, fundamental zone (Fund), and harmonic zones up to order 7 are sketched in Figure 2.8 [Ken00].

Compression or in-band distortion denotes the impact of the receiver non-linearity on the original input signal that basically results in gain reduction as illustrated in Figure 2.5.

Harmonic distortion (HD) denotes integer multiples of the input frequency f_c. That is, harmonic signals are generated at DC, $2f_c$, $3f_c$, up to $N \cdot f_c$, where the factor N is the order. HDs are already generated with a single input tone, which is especially a crucial effect for non-linear LOs. In Figure 2.8, HDs are depicted with dashed lines and squared markers.

Intermodulation distortion (IMD) or non-harmonic distortion describe linear combinations of the input frequency components. They can be interpreted as the response of the non-linear system to multiple signals at the input that comprises at least two frequencies [Rou09]. Odd-order IMD are the most annoying effects and describe distortion terms falling into the desired frequency band nearby the original input. Multiple odd-order IMD terms cause a spreading of the original input signal, also called spectral regrowth or out-of-band distortion, especially in the field of transmitter non-linearity. It is a special term for IMD caused by complex-envelope or non-constant envelope modulated signals and is referred to the bandwidth of the input signal [Sch09]. Even-order IMD shows up around DC and plays a key role in DCRs, where even-order distortion induced by the LNA easily leak through BB via the mixer (see Section 2.2), or directly fall on top of the desired signal if the receiver has a high bandwidth over several decades. The IMD products show up in each harmonic zone, illustrated with straight lines and circled markers in Figure 2.8, surrounding the harmonic frequencies of the pure tones. For an input of two tones at f_1 and f_2, the new frequencies generated due to HD and IMD will be of the form

$$f = m \cdot f_1 \pm n \cdot f_2, \tag{2.1}$$

where m and n are positive integers and $m + n$ equals the order of distortion [Ken00]. If either m or n is zero, the resulting frequency denotes HD, otherwise IMD. Following the example of a non-linear device with degree 7, as illustrated in Figure 2.8, m and n will run from 0 to 7 in (2.1), while obeying the condition $m + n \leq 7$.

Cross-modulation distortion (XMD) is a specific type of intermodulation between two relatively strong signals. Here, the stronger signal given to any odd-order non-linearity is transferring its amplitude modulation to a weaker signal, even though the opposite case is also possible. The amplitude modulation can be a result of a filtered digital modulation scheme or a result of fast fading of the radio channel [Ken00]. The generation of XMD effects depends on the power of all signals at hand [VSHGA+06]. Figure 2.9 illustrates this problem for a Gaussian shaped modulated strong interferer and a weak continuous wave (CW) wanted signal.

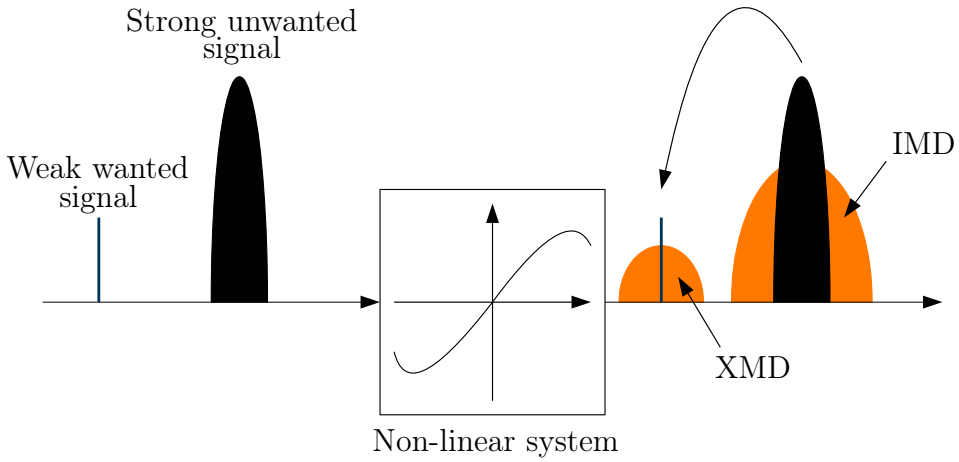


Figure 2.9: Impact of cross-modulation distortion (XMD) onto a weak carrier.

Phase Distortion or amplitude-to-phase (AM/PM) distortion denotes any modification of the phase of the input according to the non-linear transfer function [Ken00]. The preceding types just denote amplitude-to-amplitude (AM/AM) distortion, i.e. modification of the amplitude of the signal according to a non-linear transfer function. Indeed, AM/AM and AM/PM distortion superimpose in a real receiver, i.e. both amplitude and phase of the input are shaped according to the non-linear transfer function. For this reason, upper and lower IMD products may differ if AM/PM is out of phase for one of the products (e.g. 180° would cause disappearance of the distortion product).

Clipping distortion (CD) is caused due to limiting amplifiers and ADCs. These are severe distortions that cover the whole BB signal and appear almost randomly due to modulated signals [VSH10; AMV10b]. Whereas the first four types describe mild distortions that can be mitigated at reasonable complexity level, mitigating CD can be more challenging and is out of focus in this thesis.

Memory effects cannot directly be measured from the input and/or output signal of the non-linear circuit [ZWC10]. Memory itself does not modify the steady-state signal waveform, but introduces a phase shift between the input and the output. Thereby, the largest time delay of the output determines the depth or order of the memory. Indicators for memory effects in practical measurements are, e.g. imbalances between the upper and lower IMD products in case of multi-tone excitation, or spread of AM/AM and AM/PM [Sch09]. In the latter case, the conversion is broadened, i.e. the input/output values are not identical to the mean amplification. In general, memory effects appear if the bandwidth of the excitation signal is comparable with the inherent bandwidth of the analogue receiver components.

To sum up, non-linearities limit the receiver sensitivity and dynamic range through aforementioned distortion phenomena.

2.3.2 Metrics

There are several practical metrics describing the non-linear behaviour that differentiate between narrowband and wideband modulated signals [VR03; Ken00]. Figure 2.10 is illustrating, in logarithmic scale, the output vs. the input power of a non-linear device of third degree and supports the derivation of typical metrics for circuit non-linearity.

1dB compression point (P1dB) denotes the input power level where the large signal gain has dropped by 1 dB due to the compressive and saturating nature of the circuit [Rou09]. The input/output relationship of the device is no longer linear, hence, the real output drops off from the linear output power curve. Non-linear devices operating around the P1dB are already heavily non-linear and cause severe distortions. Therefore, receivers should typically operate below the P1dB with a certain back-off in order to avoid spectral regrowth and signal distortion.

Intercept Points are one of the most common metrics describing the strength of circuit non-linearity and became standard figures of merit in the RF and microwave community. These are fictitious points, e.g. IP2 and IP3 in the power level diagram Figure 2.10, where extrapolated linear and distortion products cross [VR03]. That is, the IMD product equals that of the fundamental signal amplitude. In Figure 2.10, power of IMD products of second and third order are illustrated in red and blue colour, respectively. The slope of these curves depends on the order of the distortion term. Hence, the slope of the fundamental term is $1:1$ (linear curve), and $2:1$ and $3:1$ for the IMD2 and IMD3 terms, respectively. In other words, with increasing input power, the IMD products increase 2 (or 3) times faster than the power of the fundamental frequency. Intercept points can be defined according to the input or output power. For instance, the IIP3 denotes the input-referred third-order intercept point, the OIP2 denotes the output-referred second-order intercept point. It is noteworthy that the system power would compress before the input power reaches the power of the intercept point. In case of memoryless non-linearity under weak signal conditions, i.e. with a high back-off to the P1dB, there are simple relationships between the IMD power and the corresponding IIP, as given in (2.2) and (2.3).

$$P_{\mathrm{IMD2}} = 2P_{\mathrm{in}} - P_{\mathrm{IIP2}} \tag{2.2}$$

$$P_{\mathrm{IMD3}} = 3P_{\mathrm{in}} - 2P_{\mathrm{IIP3}} \tag{2.3}$$

For memoryless systems, IIP3 is typically 10 dB higher than the P1dB [Rou09]. Also, a system intercept point can be defined, which depends on the receiver line-up and describes the overall intermodulation characteristics, or in general sense, the linearity of the system. It is determined by various block's linearity and gain, as well as their contribution in the receiver line-up [Rou09].

Basic Performance Specifications are typically found in data sheets of non-linear devices and base on raw-level numbers. They are either expressed in dBc, if there is some dependency on the LO or interferer power level, otherwise in dB, dBm, or % [Rou09]. For instance, HD2 given in unit dBc means the power difference between the fundamental (or carrier) frequency to the second harmonic. Similarly, IMD3 in dB denotes the power difference between the fundamental and the one of the third-order IMD products. Moreover, there are other figures, such as the total harmonic distortion, often used in acoustics, that describes the harmonic performance. It is defined as the ratio of the sum of powers of all harmonic components to the power of the fundamental frequency in %.

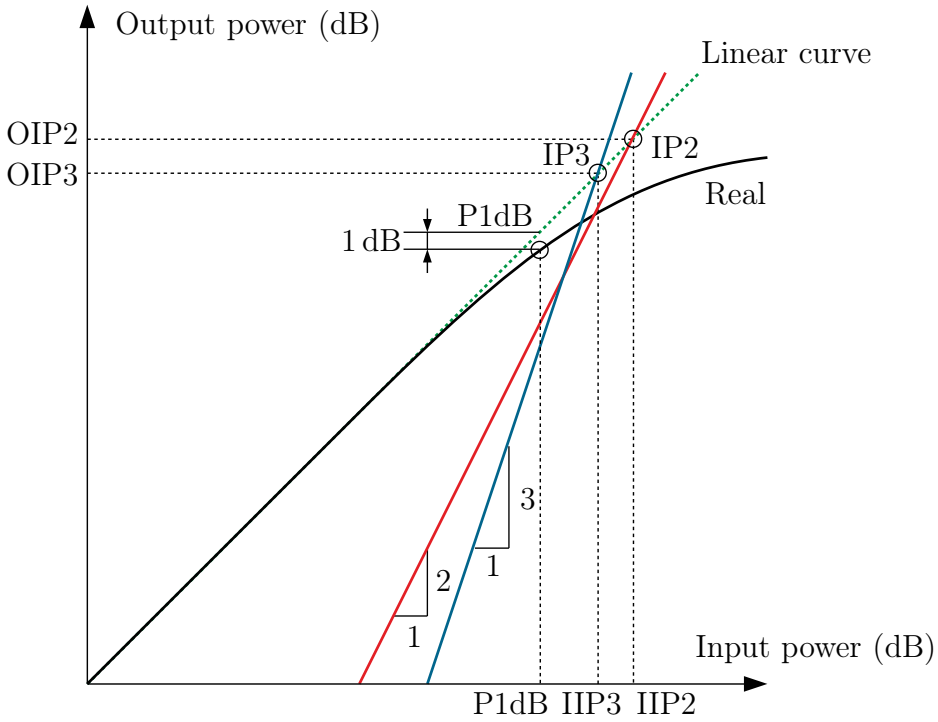


Figure 2.10: Power level diagram illustrating typical metrics for non-linear circuits.

AM/AM and AM/PM are transfer functions modelling the amplitude and phase of the fundamental signal with increasing input [VR03]. AM/AM measures the non-linearity on basis of the fundamental signal that comprises a strong linear term. However, very small non-linear effects will be seen on top of it due to common good linearity of systems. Similarly, the AM/PM is typically very small and at the order of 1° to 2° at full power and approaches zero with decreasing power. Therefore, both metrics are highly sensitive to measurement errors and seldom used in practice. It is easier and more robust to measure metrics based on the generated distortion terms, such as the intercept points.

Metrics for Wideband Signals are, e.g., adjacent channel power ratio (ACPR), noise power ratio (NPR), and multi-tone intermodulation ratio (M-IMR) [Ken00]. Previous metrics were originally defined based on narrowband signals, therefore, representing only a narrowband approximation of real bandwidth-dependent systems. However, metrics describing the effects of non-linearity for wideband modulated signals are available. ACPR denotes the degree of signal spreading, i.e. spectral regrowth by any odd-order IMD, into an adjacent channel. NPR is a measure of unwanted in-channel distortion power, found by examining the level of distortion filling a spectral gap in the input signal, that has been generated with a notch filter before. The M-IMR, often used in

orthogonal frequency-division multiplexing (OFDM) systems, is defined as the ratio of the wanted tone power to the highest IMD power.

Error vector magnitude (EVM) is a performance metric of a digital radio demodulator in general. It measures how far the received constellation points are from the ideal locations [ZMB12]. Sometimes, EVM is mentioned in the context of non-linear distortions. However, the radio channel and any RF impairment may degrade the EVM performance. Thereby, different signal space constellations are subject to different EVM values, a ratio of the power of the error vector to the root mean square (RMS) power of the reference (typically the outermost constellation point or average power).

2.4 Crucial Signal Scenarios

Beside the non-linear behaviour of the receiver circuits, signal power at the receiver input must be sufficiently high such that the non-linearity gets excited and essentially generates distortion. Hence, the amount of distortion depends on the strength of non-linearity and the input power, and therefore, on the actual signal configuration at hand.

Due to the missing pre-selection filters before the LNA in typical SDRs (see Section 2.1.4), strong out-of-band blocking signals may enter the front-end amplification and mixing stages [VSH10]. Coexistence scenarios with weak wanted and strong unwanted signals in heterogeneous radio systems are of major concern. For instance, these strong signals may cause signal desensitisation and blocking of the weak signal [Rou09], and are often referred to as blockers. With increasing signal power, the compressive behaviour of the circuit appears and the desired signal gets attenuated. This effect is known as desensitisation [WSW+10]. The power dynamics between the two signals may be up to 60–80 dB, as illustrated in Figure 2.11 [Raz10]. These conditions are quickly achieved due to near-far effects caused by

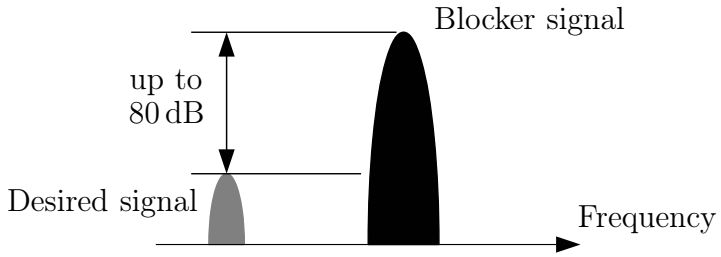


Figure 2.11: Typical signal configuration with weak desired and strong blocker signal.

wave propagation in a mobile environment, especially at the mobile terminal in cellular networks [Ken00]. Here, weak links to a serving base transceiver station (BTS) in presence of strong interfering BTSs are susceptible to non-linear distortions (see Section 7.2.2). In contrast, there may be also very high linearity requirements for the BTS receiver hardware that has to accommodate for various wireless system standards in a large frequency range [Etsa; Etse]. Further scenarios on focus in this thesis, where such conditions occur, are spectrum sensing in CR and passive radar (see Chapter 7).

However, in general, receivers need to operate under such crucial signal configurations only for a small fraction of time of roughly 10 % [KH08a; KH09b]. Though, in many applications, even such short-time problematic blocker conditions may cause a failure of the radio and should therefore be handled properly.

2.5 System-level Modelling of Non-linear Front-ends

This section deals with behavioural modelling of non-linear devices that is required later on for the mitigation processing. The use of memoryless polynomials is justified and modelling of a typical DCR topology is discussed in detail. The section is closed with some remarks on modelling of non-linear circuits with memory.

2.5.1 Behavioural vs. Physical Models

From the system identification point of view, physical and behavioural models can be distinguished [Sch09].

Physical models, based on equations related to voltage and current, require significant knowledge about the front-end's analogue components, as well as the relationship and theoretical rules describing their interaction. They are used in circuit-level simulations to simulate the behaviour under all kinds of excitations and environmental conditions, but are very time-consuming as they need to handle true RF (bandpass) modulated signals. Thereby, circuit-level simulations account for possible harmonic content and distinct time scales for RF carrier and BB information. For example, according to the Nyquist theorem, a quasi-analogue signal representation of a 20 MHz wide signal in a 2 GHz RF band will require sampling with at least 4 GHz, causing a huge amount of samples to be processed in the circuit-level simulation. Two types of physical models can be further distinguished: (i) equivalent circuit models and (ii) circuit-level behavioural models, both working with voltages and currents or incident and

reflected waves. The latter one also includes possible interstage mismatches in the input and output ports. The most recent circuit-level model is the X-parameter model as a non-linear extension of S-parameters, invented by Agilent Technologies Inc. [Vye10; VRW+05].

Behavioural models are extracted from input-output observations and enable complete system-level simulations. They are preferred in most cases, especially when no equivalent circuit description is available [Sch09]. However, simulation results are sensitive to the model structure and the parameter extraction method, thus, behavioural models have a drawback of generalisation. In other words, they are accurate for the data set used for its extraction and most likely for the same excitation class, but are not guaranteed to provide useful results for a different data set [TGT+96]. For example, applying a broadband input to a model that has been extracted with narrowband excitation intuitively produces wrong results and is equivalent to assume the superposition principle for non-linear systems which is of course wrong [Sch09]. It is likely that there are unknown aspects of response-based behavioural models as technological attributes are omitted in the model. The closer the model matches the real front-end architecture, the more robust and accurate will be the prediction of the behavioural activity under all kinds of excitations and environmental conditions. A detailed classification of behavioural models according to application, model structure, and amplifier physics is given in [Sch09].

A major distinction of behavioural models is made between lowpass or bandpass equivalent models [Sch09]. Physical (circuit) models work directly with the actual RF signal which contains the full RF circuit's bandpass nature. Lowpass or BB equivalent modelling processes only the complex envelope of the actual information signal, its carrier frequency is basically neglected. This type of system-level modelling is computationally efficient and matches also to the application of mitigating non-linear distortion at system level.

A further classification of behavioural models is conducted based upon the memory characteristics of the underlying non-linear component into:

- memoryless (only AM/AM),
- quasi-memory (AM/AM and AM/PM), and
- memory models [Sch09].

Memoryless models consider only static and frequency-independent AM/AM distortion, thus, no previous values are considered. Quasi-memory models additionally include AM/PM distortion that manifest instantaneous memoryless non-linear phase distortion effects. Examples for memoryless models are complex power series (polynomials), Saleh model and modifications of it, Fourier series model, and Bessel-Fourier model, among others [Sch09].

Memory models consider also the dependency of current output by previous input and are often applied for wideband excitation signals. Typical candidates for models with memory are Volterra series, memory polynomial, and the Wiener/Hammerstein model [Sch09].

2.5.2 Modelling by Memoryless Polynomials

The memoryless polynomial is applicable for many non-linearity related issues [AMV10a] and often provides an acceptable level of accuracy in many applications. It allows for easy computational implementation and very effective system simulations, thereby it found good use in engineering design and analysis. An extension to model strong or higher-order non-linearities is easily achieved.

However, the polynomial model is subject to various limitations. First, it is only valid for a small fitting range due to the Taylor series development around a specific operating point with small-signal excitation. Thus, the front-end should operate in the linear or mildly non-linear regions. Applying the model for strongly non-linear regions will result in a high modelling error [Sch09]. In fact, the limitation for large-signal operation is very problematic, as complex input signals with a significant envelope PAPR are often to be processed in modern RF air interfaces. Second, the memoryless assumption is valid only for one frequency at a time, i.e. for narrowband systems where the RF carrier frequency is much larger than the desired signal bandwidth. Although the bandwidth dependency may be negligible, long-term memory effects can still appear [MKH+10; Sch09]. By strict definition, the memoryless assumption holds only for circuits without inductors or capacitors, so that voltages and currents are not depending on previous values. In practice, all active circuits manifest memory, e.g. due to transistors equivalent schemes and matching and bias networks [Sch09]. The memoryless polynomial approximation is acceptable under the assumption that the front-end manifests no or few memory effects, whose duration is close to the period of the RF carrier [Sch09]. In addition, the input signal should be narrowband and comprise a constant envelope in the band of interest.

The choice of memoryless polynomials for modelling weakly non-linear systems is mainly application-driven. From the mitigation point of view, a simple model is required that identifies the origin of new frequency components. After the source and process of the unwanted signal is known, a mitigation algorithm can be formulated. That is, reproducing the most important non-linear behaviour of the front-end is sufficient for the algorithm, thus, a simple power series is chosen.

From the system-level point of view on memoryless non-linear models, the output envelope simultaneously reacts to the input envelope, represented by amplitude and phase. That is, there is a static non-linear relation between the input $x(t)$ and the output $y(t)$ given by

$$y(t) = G\left[x(t)\right]. \tag{2.4}$$

Here, G denotes the complex transfer function of the non-linear device and can be written as

$$G(A) = g(A)e^{j\Phi(A)} = P(A) + jQ(A), \tag{2.5}$$

where $g(A)$ is the AM/AM distortion and $\Phi(A)$ represents the AM/PM distortion. The complex BB signal is represented by

$$x(t) = A(t)e^{j\phi(t)} = x_{\mathrm{I}}(t) + j\, x_{\mathrm{Q}}(t), \tag{2.6}$$

where $A(t)$ and $\phi(t)$ are the general envelope amplitude and phase components, and x_{I} and x_{Q} are the I and Q-component of the excitation signal. Equation (2.5) is also referred to as quadrature model, constituting a general form of a memoryless non-linear model [Sch09]. A functional schematic of such a quadrature model is depicted in Figure 2.12.

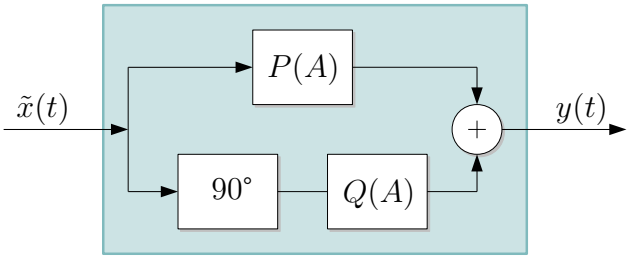


Figure 2.12: Quadrature model of a memoryless non-linear model according to [Sch09].

An Nth order complex power series memoryless model is expressed as

$$y(t) = \sum_{n=1}^{N} c_n \tilde{x}^n(t) = c_1\tilde{x}(t) + c_2\tilde{x}^2(t) + c_3\tilde{x}^3(t) + \ldots + c_N\tilde{x}^N(t), \tag{2.7}$$

where c_n are complex-valued coefficients and $\tilde{x}(t)$ being a real-valued input (e.g. I- or Q component of complex envelope $x(t)$). Equation (2.7) represents a memoryless polynomial at RF passband. Here, the output signal as polynomial expansion of the instantaneous input has been truncated after N coefficients. With complex coefficients c_n in (2.7), phase shift in narrowband systems is also modelled, thus, also memory effects are considered to some

extent (quasi-memory model) [VR03]. If these coefficients are real, only AM/AM distortion would be considered.

Equation (2.7) can be transformed to its equivalent BB representation with the complex envelope input $x(t)$ to [KK01; Teh09]

$$y(t) = \sum_{n=1}^{N} \tilde{c}_n x(t) \, |x(t)|^{n-1}, \tag{2.8}$$

by assuming that some zonal filtering is adopted, hence, only odd-order distortion terms around the carrier frequency f_c contribute to the output signal [MMK+06]. The coefficients are then given by

$$\tilde{c}_n = \frac{1}{2^{n-1}} \binom{n}{\frac{n+1}{2}} c_n. \tag{2.9}$$

A detailed derivation is given in Section A.1.

Typically, real-valued coefficients are assumed for the polynomial (2.7), where relations with the intercept points can be easily deduced [ZMS09; Che11]. The coefficients can be derived by applying the definition of the intercept points, namely the equality between the linear term and the IMD2 and IMD3 term, respectively. The following derivation is based on a two-tone excitation of a non-linear device of degree 3. Assuming a linear gain of

$$\mathrm{G}|_{\mathrm{dB}} = 20 \log_{10} c_1, \tag{2.10}$$

the corresponding relations of input-referred second-order intercept point (IIP2) and IIP3 are

$$c_1 \, A_{\mathrm{IIP2}} = c_2 \, A_{\mathrm{IIP2}}^2, \tag{2.11}$$

$$c_1 \, A_{\mathrm{IIP3}} = \frac{3}{4} c_3 \, A_{\mathrm{IIP3}}^3, \tag{2.12}$$

A_{IIP2} and A_{IIP3} are the peak voltage levels at the corresponding power levels of the intercept points that cannot be reached in practice. The coefficients on the right result from the square and cubic operation (cp. Table 2.2). Converting IIP2 and IIP3 powers in dBm to voltages and inserting (2.11) and (2.10) yields the desired relation between the intercept points and

the polynomial coefficients

$$\mathrm{IIP2}|_{\mathrm{dBm}} = 10\log_{10}\left(\frac{A_{\mathrm{IIP2}}^2}{2\cdot R\cdot 10^{-3}}\right) = 10\log_{10}\left(\frac{c_1^2}{c_2^2\cdot 2\cdot R\cdot 10^{-3}}\right), \tag{2.13}$$

$$\mathrm{IIP3}|_{\mathrm{dBm}} = 10\log_{10}\left(\frac{A_{\mathrm{IIP3}}^2}{2\cdot R\cdot 10^{-3}}\right) = 10\log_{10}\left(\frac{4\cdot c_1}{c_3\cdot 3\cdot 2\cdot R\cdot 10^{-3}}\right), \tag{2.14}$$

where R is the reference resistance for converting W to V and is assumed to be $1\,\Omega$ for sake of simplicity (system-theoretic resistance). For extracting the model coefficients out of the intercept points, inverse formulas are desirable

$$c_1 = 10^{\frac{\mathrm{G}}{20}}, \tag{2.15}$$

$$c_2 = \sqrt{\frac{c_1^2}{2\cdot R\cdot 10^{-3}\cdot 10^{\frac{\mathrm{IIP2}}{10}}}}, \tag{2.16}$$

$$c_3 = (-)\frac{4\cdot c_1}{3\cdot 2\cdot R\cdot 10^{-3}\cdot 10^{\frac{\mathrm{IIP3}}{10}}}. \tag{2.17}$$

Note that the third-order coefficient c_3 is always negative. This kind of model extraction based on intercept points will be discussed in Section 2.7.2.

In the following subsections, BB equivalent modelling with memoryless polynomials for individual receiver components as well as the total RF chain of a typical DCR, as illustrated in Figure 2.6, is discussed. This approach has not been considered in the state-of-the-art literature so far. Indeed, the RF front-end is generally modelled to be ideal when evaluating the performance of a communication system, and the receiver non-linearity is often just modelled as an additive white Gaussian noise (AWGN) source [Sch08]. Moreover, all the models are originally introduced for PAs, because the PA non-linearity at the transmitter is considered to be the major source of non-linearity in the whole transmission chain [Sch08]. As seen in the next sections, modelling of receiver RF front-ends can be fairly different to a single transmitter amplifier. For instance, the impact of frequency conversion by the mixer is considered therein for developing appropriate BB representations of the distortion terms. This novel approach of modelling the complete DCR chain has been published in [GAM+14].

2.5.3 Modelling of RF Non-linearity

Next, a BB representation of the distortion caused by the RF amplifier is developed, i.e. distortion generated by the LNA of a DCR with a low-IF architecture as introduced in Section 2.1.3. The LNA constitutes the only non-linear component in such an RF chain that

operates with the true RF bandpass signal and is highlighted in blue colour in Figure 2.13. The real-valued RF bandpass signal can be written as

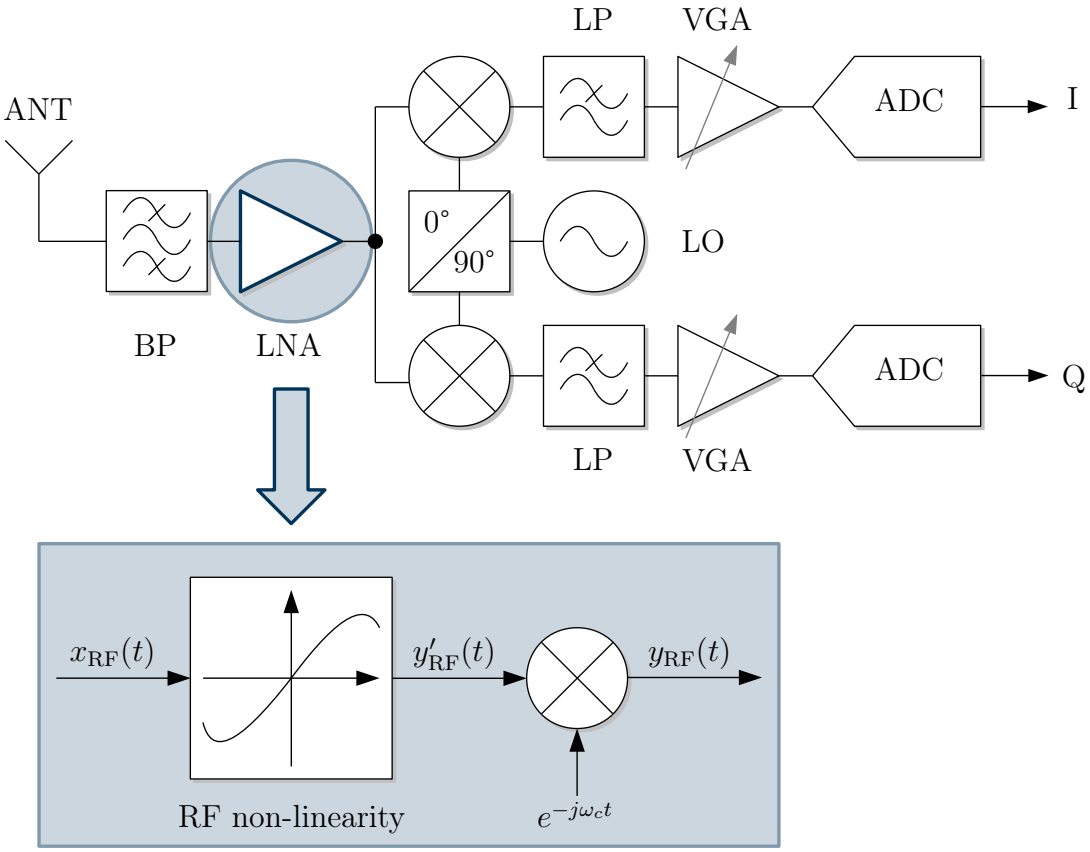


Figure 2.13: Modelling of a non-linear LNA adopted in a low-IF receiver.

$$x_{\mathrm{RF}}(t) = 2\,\mathrm{Re}\left[x(t)e^{j\omega_c t}\right] = x\,(t)\,\mathrm{e}^{j\omega_c t} + x^*(t)\mathrm{e}^{-j\omega_c t}, \tag{2.18}$$

where $\omega_c = 2\pi f_c$ is the angular RF carrier frequency and $(.)^*$ denotes complex conjugate. Note that $x(t)$ may contain multiple carriers of multiple co-existing radio access technologies at different complex IFs, especially the desired signal among other strong interferers (see Section 2.4). The RF non-linearity is then modelled by a memoryless polynomial according to (2.7), outputting the signal

$$y_{\mathrm{RF}}(t) = c_1 x_{\mathrm{RF}}(t) + c_2 x^2_{\mathrm{RF}}(t) + c_3 x^3_{\mathrm{RF}}(t) + \ldots + c_N x^N_{\mathrm{RF}}(t) \tag{2.19}$$

that is still including HD and IMD at passband [Sch09]. Now, some significant simplifications are made that simplify the derivation of a BB representation of (2.19).

First, the RF carrier frequency f_c is typically significantly higher than the maximum envelope frequency of $x(t)$ [Sch09]. That is, harmonics and even-order distortion are likely to be filtered out when translated to BB, obtained by the so called "zonal filters" (typically lowpass) adopted in the analogue or digital domain at BB [Sch09]. Moreover, signals may also be out of the down-conversion bandwidth of the mixer.

Second, due to typically low IIP2 values, even-order distortion is very weak and generates signal components near DC or at frequencies very far from f_c [Raz97; Raz09]. They are not harmful, except if the RF front-end is extremely wideband and x_{RF} consists of multiple strong signals that occupy a large bandwidth. Even-order distortion in the DC zone may leak through the mixer to BB without frequency translation, as discussed in Section 2.2. However, these issues have been reduced significantly due to integrated circuit design and differential signalling [Kar01]. They can also be effectively filtered out by alternating current coupling or filtering. Instead, odd-order distortion products can fall into the original signal band (fundamental zone) as depicted in Figure 2.8. In practice, the third-order non-linearity is usually the strongest and the only one that generates distortion that clearly appears above the noise level. Thus, (2.19) can be simplified to

$$y'_{\mathrm{RF}}(t) = a_1 x_{\mathrm{RF}}(t) + a_2 x^3_{\mathrm{RF}}(t), \tag{2.20}$$

where $a_1 = c_1$ denotes the linear gain and $a_2 = c_3$ the third-order coefficient of the RF amplifier. In fact, operating at large-signal conditions demand for larger number of terms, however, higher-order coefficients are difficult to extract as they cannot be separated out from lower-order coefficients. In other words, many terms of the complex power series (2.19) give simultaneous spectral contribution. The coefficients a_1 and a_2 are chosen to be complex to model also static AM/PM distortion. By inserting (2.18) into (2.20), a further analysis of the new frequency components can be performed:

$$\begin{aligned} y'_{\mathrm{RF}}(t) = &\left[a_1 x(t) + 3\, a_2 x^2(t) x^*(t)\right] \mathrm{e}^{j\omega_c t} \\ &+ \left[a_1 x^*(t) + 3\, a_2 x\,(t)\, x^{*2}(t)\right] \mathrm{e}^{-j\omega_c t} \\ &+ a_2 x^3(t) \mathrm{e}^{j\,3\omega_c t} \\ &+ a_2 x^{*3}(t) \mathrm{e}^{-j\,3\omega_c t}. \end{aligned} \tag{2.21}$$

Only the first term $e^{j\omega_c t}$ will remain after mixing and zonal filtering at BB, i.e. by multiplying (2.21) with $e^{-j\omega_c t}$ as illustrated in Figure 2.13. The BB representation of (2.20) after zonal

filtering is therefore

$$\begin{aligned} y_{\mathrm{RF}}(t) &= a_1 x(t) + 3a_2 x^2(t) x^*(t) \\ &= a_1 x(t) + 3a_2 A^2(t) x(t) \\ &= a_1 x(t) + 3a_2 z(t). \end{aligned} \tag{2.22}$$

A block scheme of the RF non-linearity model, derived from (2.22), is depicted in Figure 2.14.

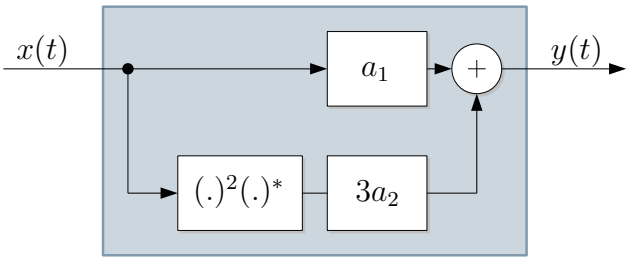


Figure 2.14: Block scheme of the RF non-linearity model derived from (2.22).

The second form comprises a notation with the envelope $x(t)x^*(t) = |x(t)|^2 = A^2(t)$. This exactly matches to the complex polynomial notation of (2.8) for $n = 1, 3$. Indeed, same assumptions about filtering have been assumed in the derivation (cp. Section A.1). In the last form, $z(t)$ is introduced to enhance readability for the following joint modelling of RF and BB non-linearity. Expanding (2.22) with (2.6) and separating in-phase and quadrature component leads to

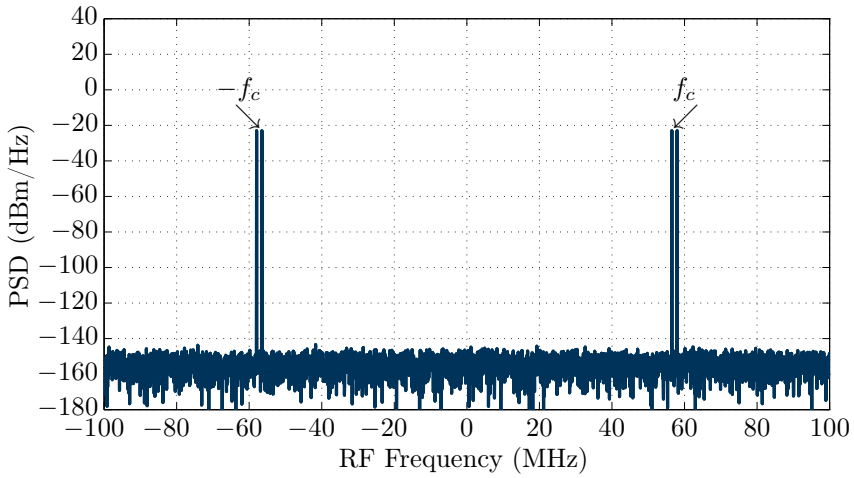
$$y_{\mathrm{RF,I}}(t) = a_1 x_{\mathrm{I}}(t) + 3a_2 \left[x_{\mathrm{I}}^3(t) + x_{\mathrm{I}}(t) x_{\mathrm{Q}}^2(t) \right] \tag{2.23}$$

$$y_{\mathrm{RF,Q}}(t) = a_1 x_{\mathrm{Q}}(t) + 3a_2 \left[x_{\mathrm{Q}}^3(t) + x_{\mathrm{Q}}(t) x_{\mathrm{I}}^2(t) \right]. \tag{2.24}$$

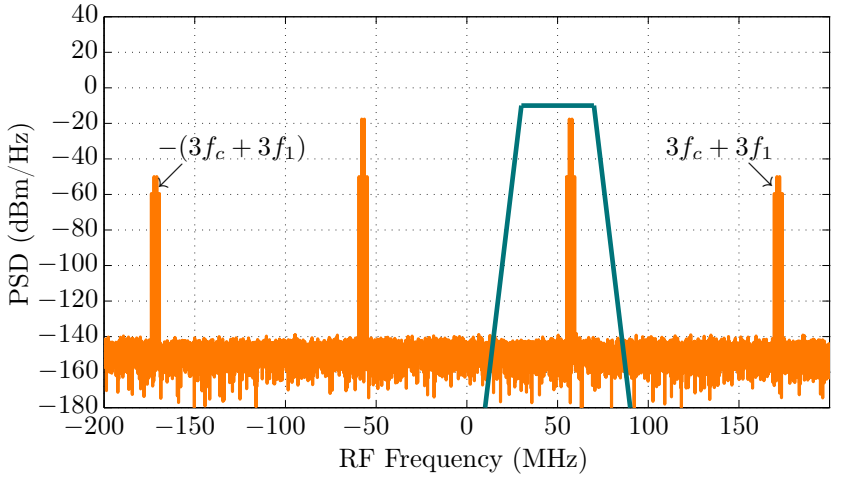
Next, a two-tone simulation has been conducted to illustrate the derivation. The complex-valued BB signal $x(t)$ according to (2.6) is

$$x(t) = e^{j\omega_1 t} + e^{j\omega_2 t}, \tag{2.25}$$

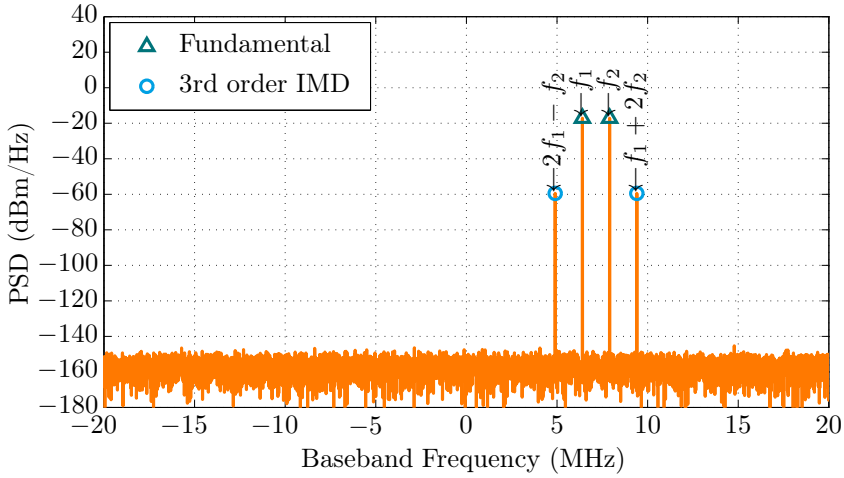
with the tones frequencies $f_1 = 6.4\,\mathrm{MHz}$ and $f_2 = 7.9\,\mathrm{MHz}$ relative to the carrier frequency $f_c = 50\,\mathrm{MHz}$. The RF spectrum is depicted by Figure 2.15(a), where only the RF carrier of the two-tone signal is marked. Both tones have an equal power of $-23\,\mathrm{dBm}$, resulting in a total power of $-20\,\mathrm{dBm}$. The noise power has been set to a particularly small value of about $-200\,\mathrm{dBm}$ in order to make also weak IMD after subsequent distortion at BB visible.



(a) Real-valued RF passband signal with two-tone signal at both sides of the spectrum.



(b) RF passband distortion with zonal filter marked in green colour.



(c) BB equivalent representation of RF distortion after zonal filtering.

Figure 2.15: RF distortion caused by two-tone input.

Note that the noise power distributes over the full BB bandwidth. By substituting (2.25) in (2.22) and sorting the terms yields the signal components in Table 2.1.

Table 2.1: Frequency components generated by RF non-linearity.

Coefficient	**Frequency**	**Type**
$9\,a_2 + a_1$	$+\omega_1$	Fundamental
$9\,a_2 + a_1$	$+\omega_2$	Fundamental
$3\,a_2$	$2\omega_2 - \omega_1$	3^{rd} order IMD
$3\,a_2$	$2\omega_1 - \omega_2$	3^{rd} order IMD

Note that the amplitude of the fundamental frequencies is, beside linear gain a_1, weighted by the third-order coefficient a_2 which is always negative and causes the compression behaviour. Distortion on top of the actual input signal is also denoted as in-band distortion from the perspective/bandwidth of the input signal. The RF distortion at passband is illustrated in Figure 2.15(b), assuming $a_1 = 2$ and $a_2 = -1000$, that corresponds to $\text{G}_{\text{RF}} = 6\,\text{dB}$ and $\text{IIP3}_{\text{RF}} = +6\,\text{dBm}$ according to Equations (2.10) and (2.14). A bit exaggerated value is chosen for the third-order coefficient a_2 of the RF non-linearity for the sake of better visualisation. The band of interest after zonal filtering (green frame in Figure 2.15(b)) is illustrated in Figure 2.15(c). Hence, only the third-order IMD around the fundamental frequencies is left from the distortion introduced by a single RF amplifier (LNA).

2.5.4 Modelling of BB Non-linearity

Now, modelling of sole BB non-linearity is discussed. Contrary to RF non-linearity, there are physically separated I and Q branches at BB that manifest a different non-linear behaviour. Physical sources of BB non-linearity in both I and Q branches are the mixer, BB amplifier (variable-gain amplifier (VGA)) or ADC driver, and the ADC , as illustrated in Figure 2.16. The total distorted BB signal

$$y_{\text{BB}}(t) = y_{\text{BB,I}}(t) + j y_{\text{BB,Q}}(t) \tag{2.26}$$

can be modelled by two individual memoryless polynomial models

$$y_{\text{BB,I}}(t) = c_{1,\text{I}}\, y_{\text{RF,I}}(t) + c_{2,\text{I}}\, y^2_{\text{RF,I}}(t) + c_{3,\text{I}}\, y^3_{\text{RF,I}}(t) + \ldots + c_{N,\text{I}}\, y^N_{\text{RF,I}}(t) \tag{2.27}$$

$$y_{\text{BB,Q}}(t) = c_{1,\text{Q}}\, y_{\text{RF,Q}}(t) + c_{2,\text{Q}}\, y^2_{\text{RF,Q}}(t) + c_{3,\text{Q}}\, y^3_{\text{RF,Q}}(t) + \ldots + c_{N,\text{Q}}\, y^N_{\text{RF,Q}}(t), \tag{2.28}$$

where $y_{\mathrm{RF,I}} = a_1 x_{\mathrm{I}}$ and $y_{\mathrm{RF,Q}} = a_1 x_{\mathrm{Q}}$, i.e. fully linear amplification at the RF amplifier is assumed for a moment. Indeed, the non-linear characteristics of the I and Q channels can be modelled as a sum of two bandpass non-linearities [LHG+10]. Following the modelling approach for the RF non-linearity, similar simplifications are made. First, even-order distortion is assumed to be very weak due to efficient circuit design techniques such as differential signalling [Kar01]. In typical DCRs, symmetrical electric wiring is adopted behind the LNA, i.e. by using two signals V_+ and V_- that are ideally 180° out of phase, i.e. $V_+ = -V_-$. Following [Kar01], the output or differential voltage can be written as

$$V_{\mathrm{diff}} = V_+ - V_-, \tag{2.29}$$

where the time-dependency of the voltages is omitted for a moment. Then, assuming perfectly balanced signals, both voltages can be written as

$$V'_+ = c_1 V_+ + c_2 V_+^2 + c_3 V_+^3 + \ldots + c_N V_+^N \tag{2.30}$$

$$V'_- = c_1 (-V_+) + c_2 (-V_+)^2 + c_3 (-V_+)^3 + \ldots + c_N (-V_+)^N, \tag{2.31}$$

according to (2.27) or (2.28). Note that the pair (2.30) and (2.31) have the same form for both I and Q component, so appropriate subscripts are omitted here. Eventually, the odd-order terms keep their sign, but the even-order distortion is always positive. Computing V'_{diff} according to (2.29) yields

$$V'_{\mathrm{diff}} = V'_+ - V'_- = 2c_1 V_+ + 2c_3 V_+^3 + \ldots + 2c_{2n+1} V_+^{2n+1}. \tag{2.32}$$

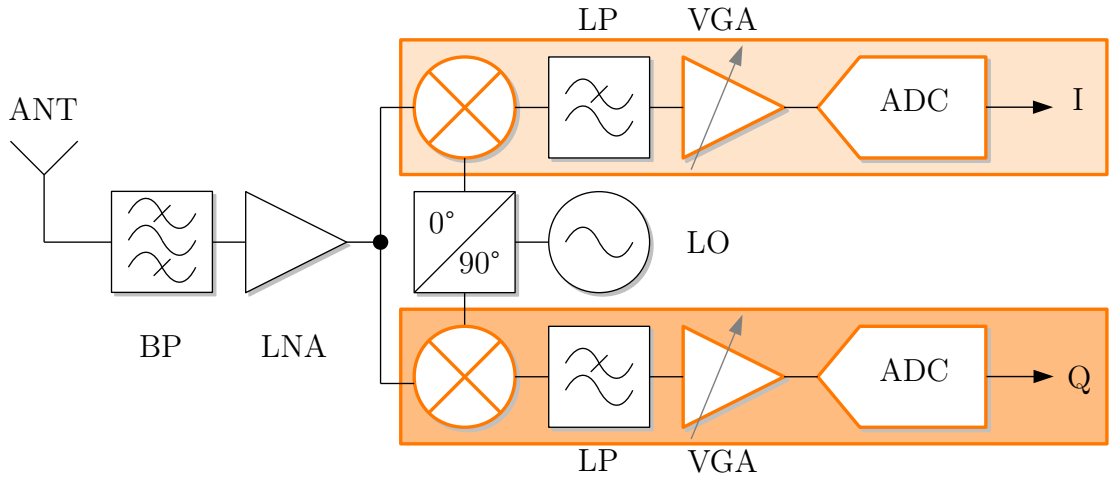


Figure 2.16: Modelling of non-linear BB components in a low-IF receiver.

That is, the even-order terms cancel out, but the odd-order terms increase by a factor two [Kar01]. However, perfect even-order cancellation by differential signalling requires symmetrical layout and matched amplifiers to ensure perfect balance among the signals. The minor importance is also justified by measurements presented in Section 2.7. Moreover, higher-order terms are typically masked by noise. Therefore, it is assumed that third-order non-linearity is describing the most significant distortion, hence, (2.27) and (2.28) can be simplified to:

$$y'_{\mathrm{BB,I}}(t) = a_{3,\mathrm{I}}\, y_{\mathrm{RF,I}}(t) + a_{4,\mathrm{I}}\, y^3_{\mathrm{RF,I}}(t) \tag{2.33}$$

$$y'_{\mathrm{BB,Q}}(t) = a_{3,\mathrm{Q}}\, y_{\mathrm{RF,Q}}(t) + a_{4,\mathrm{Q}}\, y^3_{\mathrm{RF,Q}}(t), \tag{2.34}$$

where $a_{3,\mathrm{I}} = c_{1,\mathrm{I}}$, $a_{3,\mathrm{Q}} = c_{1,\mathrm{Q}}$, $a_{4,\mathrm{I}} = c_{3,\mathrm{I}}$, and $a_{4,\mathrm{Q}} = c_{3,\mathrm{Q}}$ for notational convenience in the following derivation including both RF and BB distortion. A block scheme of the BB non-linearity model is depicted in Figure 2.17.

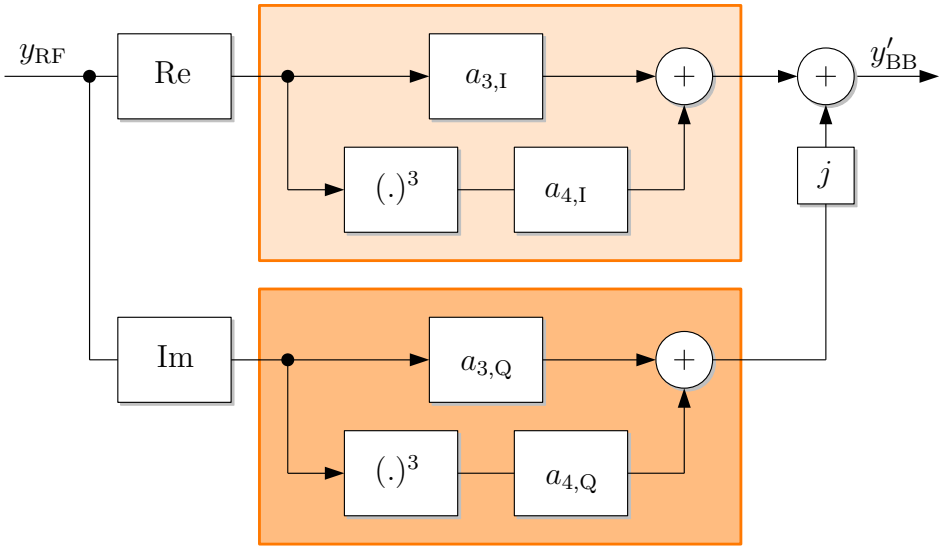


Figure 2.17: Block scheme of the BB non-linearity model employing a cubic term separately for the I and Q part.

Assuming a two-tone down-converted signal according to (2.25), the output components after passing BB non-linearity are as listed in Table 2.2.

The BB distortion is illustrated in Figure 2.18, assuming $a_{3,\mathrm{I}} = a_{3,\mathrm{Q}} = 2$ and $a_{4,\mathrm{I}} = a_{4,\mathrm{Q}} = -1000$ similar to the RF distortion case. Beside IMD, the BB non-linearity generates also HD directly in the BB on the left side of the spectrum at $-3\omega_{1,2}$, if the strong complex carriers fall onto low (non-zero) IFs. Thus, HD cannot be neglected in the discussion as for the RF non-linearity. Moreover, it is noteworthy that the BB non-linearity causes third-

Table 2.2: Frequency components generated by BB non-linearity.

Coefficient	Frequency	Type
$a_3 + \frac{9}{4}\,a_4$	$+\omega_1$	Fundamental
$a_3 + \frac{9}{4}\,a_4$	$+\omega_2$	Fundamental
$\frac{1}{4}\,a_4$	$-3\omega_1$	3rd order HD
$\frac{1}{4}\,a_4$	$-3\omega_2$	3rd order HD
$\frac{3}{4}\,a_4$	$2\omega_1 - \omega_2$	3rd order IMD
$\frac{3}{4}\,a_4$	$2\omega_2 - \omega_1$	3rd order IMD
$\frac{3}{4}\,a_4$	$-(2\omega_2 + \omega_1)$	3rd order IMD
$\frac{3}{4}\,a_4$	$-(2\omega_1 + \omega_2)$	3rd order IMD

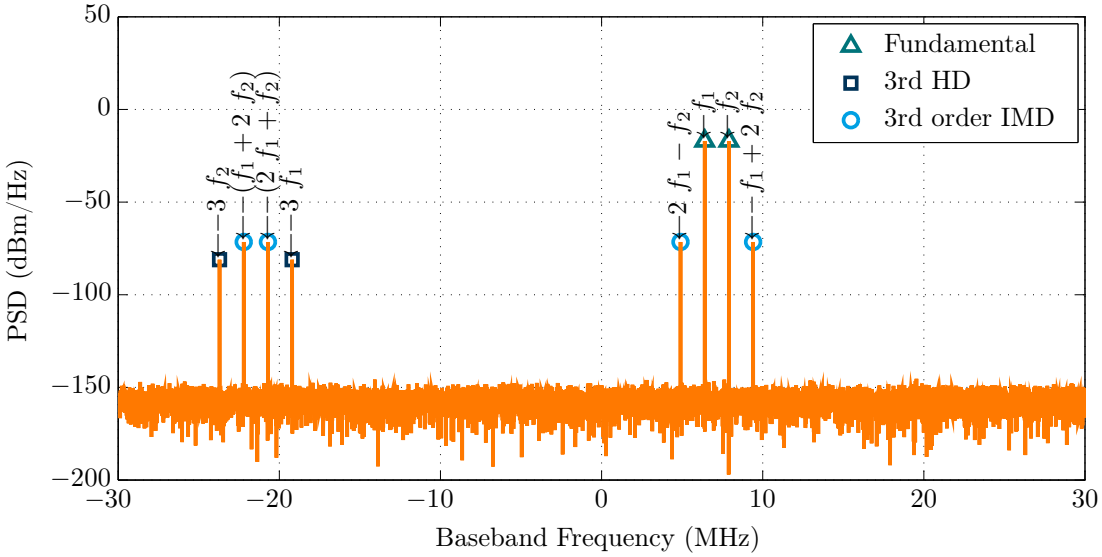


Figure 2.18: BB distortion caused by two-tone input generates additional frequency components in the fundamental and third harmonic zone.

order IMD in the fundamental zone, at exactly the same frequencies of the RF distortion. Considering the total receiver line-up of a DCR, both RF and BB distortion will sum up in the fundamental zone, as will be shown later.

Considering only the third-order of the received signal is the main approach followed in the state-of-the-art literature, such as in [GSH+12a; VSHGA+06]. However, it is restricted for modelling non-linear systems at equivalent BB. Down-conversion by the mixer and subsequent distortion of RF distortions at BB are not considered therein that is one of the research gaps tackled in this thesis and being on focus in the next section.

2.5.5 Joint Modelling of RF and BB Non-linearity

Now, a cascade of RF and BB models is considered, i.e. both models depicted in Figure 2.14 and Figure 2.17 are connected in series. In practice, the non-linearities of the RF stage are transferred to the BB receiver stages, thereby the order of the final distortion and the number of IMD products is increased. In fact, adopting a cascade of two third-order polynomials matches to the physical architecture of a DCR. By splitting (2.22) into I and Q component and substituting them in (2.33) and (2.34) yields

$$\begin{aligned} y'_{\mathrm{BB,I}}(t) &= a_{3,\mathrm{I}}a_1x_{\mathrm{I}}(t) + 3a_{3,\mathrm{I}}a_2z_{\mathrm{I}}(t) \\ &\quad + a_{4,\mathrm{I}}{a_1}^3x_{\mathrm{I}}^3(t) + 9a_{4,\mathrm{I}}{a_1}^2a_2x_{\mathrm{I}}^2(t)z_{\mathrm{I}}(t) \\ &\quad + 27a_{4,I}a_1{a_2}^2x_{\mathrm{I}}(t)z_{\mathrm{I}}^2(t) + 27a_{4,\mathrm{I}}{a_2}^3z_{\mathrm{I}}^3(t), \end{aligned} \tag{2.35}$$

and

$$\begin{aligned} y'_{\mathrm{BB,Q}}(t) &= a_{3,\mathrm{Q}}a_1x_{\mathrm{Q}}(t) + 3a_{3,\mathrm{Q}}a_2z_{\mathrm{Q}}(t) \\ &\quad + a_{4\mathrm{Q}}{a_1}^3x_{\mathrm{Q}}^3(t) + 9a_{4\mathrm{Q}}{a_1}^2a_2x_{\mathrm{Q}}^2(t)z_{\mathrm{Q}}(t) \\ &\quad + 27a_{4\mathrm{Q}}a_1{a_2}^2x_{\mathrm{Q}}(t)z_{\mathrm{Q}}^2(t) + 27a_{4\mathrm{Q}}{a_2}^3z_{\mathrm{Q}}^3(t). \end{aligned} \tag{2.36}$$

Equations (2.35) and (2.36) can be further opened by applying (2.23) and (2.24). Hence, a BB representation of the total RF and BB distortion is obtained with relation to the input signal $x(t)$ (2.6):

$$\begin{aligned} y'_{\mathrm{BB,I}}(t) &= a_{3,\mathrm{I}}a_1x_{\mathrm{I}}(t) + 3\,a_{3,\mathrm{I}}a_2x_{\mathrm{I}}^3(t) + 3\,a_{3,\mathrm{I}}a_2x_{\mathrm{I}}(t)x_{\mathrm{Q}}^2(t) \\ &+ 27\,a_{4,\mathrm{I}}a_2^3x_{\mathrm{I}}^9(t) + 81\,a_{4,\mathrm{I}}a_2^3x_{\mathrm{I}}^7(t)x_{\mathrm{Q}}^2(t) \\ &+ 27\,a_{4,\mathrm{I}}a_2^2a_1x_{\mathrm{I}}^7(t) + 81\,a_4a_2^3x_{\mathrm{I}}^5(t)x_{\mathrm{Q}}^4(t) \\ &+ 54\,a_{4,\mathrm{I}}a_2^2a_1x_{\mathrm{I}}^5(t)x_{\mathrm{Q}}^2(t) + 9\,a_{4,\mathrm{I}}a_2a_1^2x_{\mathrm{I}}^5(t) \\ &+ 27\,a_{4,\mathrm{I}}a_2^3x_{\mathrm{I}}^3(t)x_{\mathrm{Q}}^6(t) + 27\,a_{4,\mathrm{I}}a_2^2a_1x_{\mathrm{I}}^3(t)x_{\mathrm{Q}}^4(t) \\ &+ 9\,a_{4,\mathrm{I}}a_2a_1^2x_{\mathrm{I}}^3(t)x_{\mathrm{Q}}^2(t) + a_4a_1^3x_{\mathrm{I}}^3(t), \end{aligned} \tag{2.37}$$

and

$$\begin{aligned} y'_{\mathrm{BB,Q}}(t) &= a_{3,\mathrm{Q}}a_1x_{\mathrm{Q}}(t) + 3\,a_{3,\mathrm{Q}}a_2x_{\mathrm{Q}}^3(t) + 3\,a_{3,\mathrm{Q}}a_2x_{\mathrm{Q}}(t)x_{\mathrm{I}}^2 \\ &+ 27\,a_{4,\mathrm{Q}}a_2^3x_{\mathrm{Q}}^9(t) + 81\,a_{4,\mathrm{Q}}a_2^3x_{\mathrm{Q}}^7(t)x_{\mathrm{I}}^2(t) \\ &+ 27\,a_{4,\mathrm{Q}}a_2^2a_1x_{\mathrm{Q}}^7(t) + 81\,a_{4,\mathrm{Q}}a_2^3x_{\mathrm{Q}}^5(t)x_{\mathrm{I}}^4(t) \\ &+ 54\,a_{4,\mathrm{Q}}a_2^2a_1x_{\mathrm{Q}}^5(t)x_{\mathrm{I}}^2(t) + 9\,a_4a_2a_1^2x_{\mathrm{Q}}^5(t) \\ &+ 27\,a_4a_2^3x_{\mathrm{Q}}^3(t)x_{\mathrm{I}}^6(t) + 27\,a_{4,\mathrm{Q}}a_2^2a_1x_{\mathrm{Q}}^3(t)x_{\mathrm{I}}^4(t) \\ &+ 9\,a_{4,\mathrm{Q}}a_2a_1^2x_{\mathrm{Q}}^3(t)x_{\mathrm{I}}^2(t) + a_{4,\mathrm{Q}}a_1^3x_{\mathrm{Q}}^3(t). \end{aligned} \tag{2.38}$$

Note that due to the cascade of two third-order polynomials, i.e. two non-linear devices of degree three in series, up to ninth order IMD is considered. The equations (2.37) and (2.38) clearly show the interaction between the RF and BB stages. Figure 2.19 depicts a simplified block scheme of the joint model considering RF and BB distortion as a cascade of two third-order polynomials.

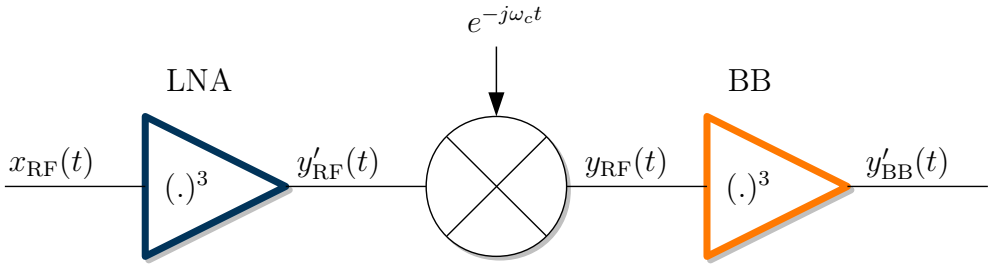


Figure 2.19: Model of joint RF and BB distortion using a cascade of two third-order polynomials.

Again, a two-tone excitation is assumed and signal components generated by the cascaded non-linearity model are analysed. The signal $y'_{\mathrm{BB}}(t)$ after passing the cascaded non-linearity in the complex BB are listed in Table 2.3.

Figure 2.20 illustrates the complex BB spectrum with a two-tone input, where $\mathbf{a} = [a_1, a_2, a_3, a_4] = [2, -1000, 2, -1000]$. In fact, there are 20 signal components in total al-

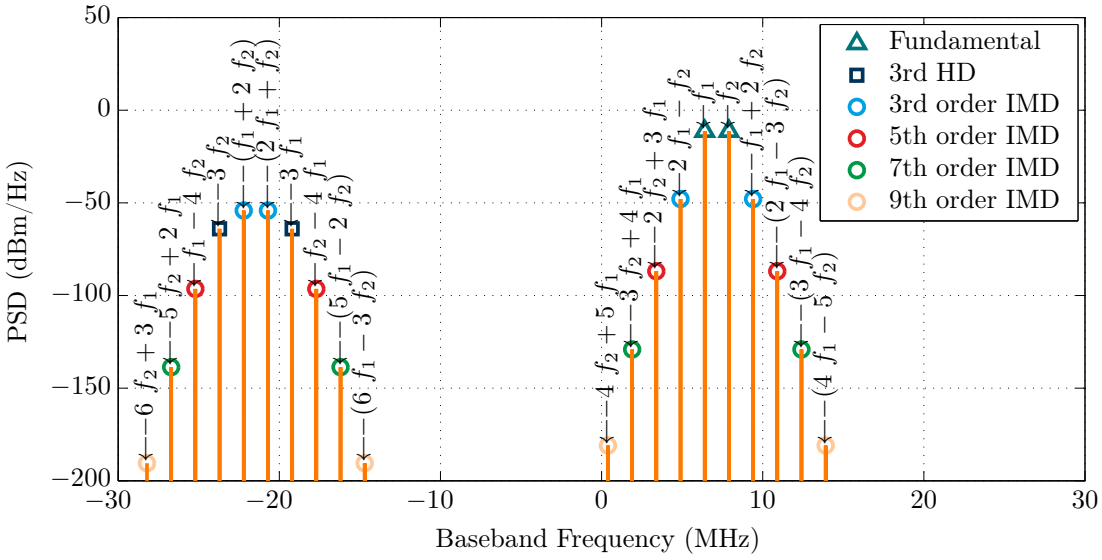


Figure 2.20: Signal components of cascaded non-linearity with two-tone input.

though there was only a two-tone (two signal components) input. However, the magnitude of the IMD products decrease with increasing order of non-linearity. Noise has

Table 2.3: Frequency components generated by cascaded RF and BB non-linearity.

Coefficient	**Frequency**	**Type**
$\frac{9}{4}\,a_4{a_1}^3+\frac{5103}{2}\,a_4{a_2}^3+a_3a_1+9\,a_3a_2+\frac{135}{2}\,a_4{a_1}^2a_2+\frac{2835}{4}\,a_4a_1{a_2}^2$	$+\omega_1$	Fundamental
$\frac{9}{4}\,a_4{a_1}^3+\frac{5103}{2}\,a_4{a_2}^3+a_3a_1+9\,a_3a_2+\frac{135}{2}\,a_4{a_1}^2a_2+\frac{2835}{4}\,a_4a_1{a_2}^2$	$+\omega_2$	Fundamental
$\frac{1}{4}\,a_4{a_1}^3+567\,a_4{a_2}^3+\frac{567}{4}\,a_4a_1{a_2}^2+\frac{45}{4}\,a_4{a_1}^2a_2$	$-3\omega_1$	3rd order HD
$\frac{1}{4}\,a_4{a_1}^3+567\,a_4{a_2}^3+\frac{567}{4}\,a_4a_1{a_2}^2+\frac{45}{4}\,a_4{a_1}^2a_2$	$-3\omega_2$	3rd order HD
$\frac{3}{4}\,a_4{a_1}^3+1701\,a_4{a_2}^3+\frac{135}{4}\,a_4{a_1}^2a_2+\frac{1701}{4}\,a_4a_1{a_2}^2+3\,a_3a_2$	$2\omega_1-\omega_2$	3rd order IMD
$\frac{3}{4}\,a_4{a_1}^3+1701\,a_4{a_2}^3+\frac{135}{4}\,a_4a_1^2a_2+\frac{1701}{4}\,a_4a_1{a_2}^2+3\,a_3a_2$	$2\omega_2-\omega_1$	3rd order IMD
$\frac{1701}{2}\,a_4{a_2}^3+\frac{3}{4}\,a_4{a_1}^3+\frac{45}{2}\,a_4{a_1}^2a_2+\frac{945}{4}\,a_4a_1{a_2}^2$	$-(2\omega_2+\omega_1)$	3rd order IMD
$\frac{1701}{2},a_4{a_2}^3+\frac{3}{4}\,a_4{a_1}^3+\frac{45}{2}\,a_4{a_1}^2a_2+\frac{945}{4}\,a_4a_1{a_2}^2$	$-(2\omega_1+\omega_2)$	3rd order IMD
$\frac{9}{4}\,a_4{a_1}^2a_2+\frac{189}{4}\,a_4a_1a_2^2+243\,a_4{a_2}^3$	$\omega_1-4\omega_2$	5th order IMD
$\frac{9}{4}\,a_4{a_1}^2a_2+\frac{189}{4}\,a_4a_1{a_2}^2+243\,a_4{a_2}^3$	$\omega_2-4\omega_1$	5th order IMD
$\frac{27}{4}\,a_4{a_1}^2a_2+\frac{567}{4}\,a_4a_1{a_2}^2+729\,a_4{a_2}^3$	$-2\omega_1+3\omega_2$	5th order IMD
$\frac{27}{4}\,a_4{a_1}^2a_2+\frac{567}{4}\,a_4a_1{a_2}^2+729\,a_4{a_2}^3$	$-2\omega_2+3\omega_1$	5th order IMD
$\frac{81}{4},a_4a_1{a_2}^2+\frac{729}{4}\,a_4{a_2}^3$	$-3\omega_1+4\omega_2$	7th order IMD
$\frac{81}{4}\,a_4a_1{a_2}^2+\frac{729}{4}\,a_4{a_2}^3$	$-3\omega_2+4\omega_1$	7th order IMD
$\frac{27}{4}\,a_4a_1{a_2}^2+\frac{243}{4}\,a_4{a_2}^3$	$-5\omega_2+2\omega_1$	7th order IMD
$\frac{27}{4}\,a_4a_1{a_2}^2+\frac{243}{4}\,a_4{a_2}^3$	$-5\omega_1+2\omega_2$	7th order IMD
$\frac{81}{4}\,a_4{a_2}^3$	$-4\omega_1+5\omega_2$	9th order IMD
$\frac{81}{4}\,a_4{a_2}^3$	$-4\omega_2+5\omega_1$	9th order IMD
$\frac{27}{4}\,a_4{a_2}^3$	$-6\omega_1+3\omega_2$	9th order IMD
$\frac{27}{4}\,a_4{a_2}^3$	$-6\omega_2+3\omega_1$	9th order IMD

been excluded in this particular simulation to make also high-order distortion visible. In practice, 7th- and 9th-order non-linearity are below the receiver noise level or beyond the signal-to-noise ratio (SNR) of the ADC. The SNR of a typical ADC is approx. 60–80 dB, hence, making all signal components below −80 dBFS to disappear due to the quantisation noise.

Moreover, it is visible that the level of third-order IMD in the fundamental and third harmonic zone do not match. That is, RF non-linearity initially generated some additional third-order distortion power around the fundamental frequencies as already seen in Figure 2.15(c).

2.5.6 Interaction with I/Q Imbalance

Beside RF and BB non-linearity, I/Q imbalance of the mixer and the individual BB branches cause mirror signals that may interfere with other useful signals. Especially the interaction between non-linear distortion and the I/Q imbalance effects plays a special role when dealing with DCRs [GAM+14]. More details on I/Q imbalance effects and appropriate mitigation techniques can be found in [Ant11; AVR08].

Up to now, ideal down-conversion at the mixer has been assumed. In practice, a quadrature mixer cannot provide exactly 90° phase shift, but sustains a phase mismatch of ϕ_m (rad). In addition, the I and Q branches manifest a certain amplitude mismatch g_m. The I/Q imbalance of the down-conversion stage can be modelled as

$$\tilde{y}_{\mathrm{RF}}(t) = k_1 y_{\mathrm{RF}}(t) + k_2 y^*_{\mathrm{RF}}(t), \tag{2.39}$$

where the complex mismatch coefficients are

$$k_1 = (1 + g_m e^{-j\phi_m})/2, \tag{2.40}$$
$$k_2 = (1 - g_m e^{+j\phi_m})/2. \tag{2.41}$$

Figure 2.21 depicts the mixer I/Q imbalance model. In the previous derivations, perfect I/Q

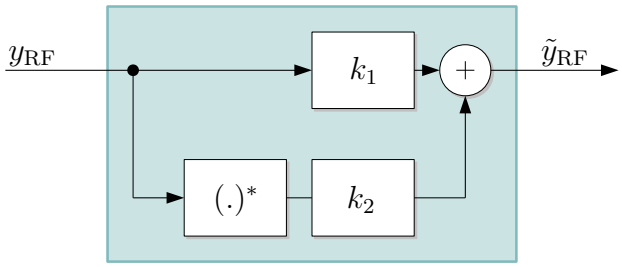


Figure 2.21: Block scheme of the mixer I/Q imbalance model.

balance has been assumed, i.e. $g_m = 1$, $\phi_m = 0$, hence $k_1 = 1$, $k_2 = 0$ and $\tilde{y}_{\mathrm{RF}}(t) = y_{\mathrm{RF}}(t)$. Next, a complex representation of the BB distortion under presence of I/Q imbalance is derived [GAM+14].

Equations (2.33) and (2.34) can be combined to

$$\begin{aligned} y'_{\mathrm{BB}}(t) &= y'_{\mathrm{BB,I}}(t) + j\, y'_{\mathrm{BB,Q}}(t) \\ &= a_{3,\mathrm{I}}\, \tilde{y}_{\mathrm{RF,I}}(t) + a_{4,\mathrm{I}}\, \tilde{y}^3_{\mathrm{RF,I}}(t) + j \left\{ a_{3,\mathrm{Q}}\, \tilde{y}_{\mathrm{RF,Q}}(t) + a_{4,\mathrm{Q}}\, \tilde{y}^3_{\mathrm{RF,Q}}(t) \right\}, \end{aligned} \tag{2.42}$$

with $\tilde{y}_{\mathrm{RF,I/Q}}$ including RF non-linearity and I/Q imbalance. Real and imaginary part of $\tilde{y}_{\mathrm{RF,I/Q}}$ can be written as

$$\tilde{y}_{\mathrm{RF,I}} = \frac{\tilde{y}_{\mathrm{RF}}(t) + \tilde{y}^*_{\mathrm{RF}}(t)}{2},$$
$$\tilde{y}_{\mathrm{RF,Q}} = \frac{\tilde{y}_{\mathrm{RF}}(t) - \tilde{y}^*_{\mathrm{RF}}(t)}{2j},$$

yielding the BB distortion in complex notation:

$$\begin{aligned} y'_{\mathrm{BB}}(t) = {} & \frac{a_{3,\mathrm{I}} + a_{3,\mathrm{Q}}}{2}\tilde{y}_{\mathrm{RF}}(t) + \frac{a_{3,\mathrm{I}} - a_{3,\mathrm{Q}}}{2}\tilde{y}^*_{\mathrm{RF}}(t) \\ & + \frac{a_{4,\mathrm{I}} + a_{4,\mathrm{Q}}}{8}\tilde{y}^3_{\mathrm{RF}}(t) + \frac{3a_{4,\mathrm{I}} - 3a_{4,\mathrm{Q}}}{8}\tilde{y}^2_{\mathrm{RF}}(t)\tilde{y}^*_{\mathrm{RF}}(t) \\ & + \frac{3a_{4,\mathrm{I}} - 3a_{4,\mathrm{Q}}}{8}\tilde{y}_{\mathrm{RF}}(t)\left[\tilde{y}^*_{\mathrm{RF}}(t)\right]^2 + \frac{a_{4,\mathrm{I}} + a_{4,\mathrm{Q}}}{8}\left[\tilde{y}^*_{\mathrm{RF}}(t)\right]^3. \end{aligned} \tag{2.43}$$

This equation needs to be further expanded with (2.39) to get the complete representation of the BB distortion including mixer I/Q imbalance ($k_1 \neq 1, k_2 \neq 0$) and further imbalance between the I and Q receiver chains ($a_{3,\mathrm{I}} \neq a_{3,\mathrm{Q}}$ and $a_{4,\mathrm{I}} \neq a_{4,\mathrm{Q}}$). The full form is given by

$$\begin{aligned} y'_{\mathrm{BB}}(t) = {} & \left\{\frac{a_{3,\mathrm{I}} + a_{3,\mathrm{Q}}}{2}k_1 + \frac{a_{3,\mathrm{I}} - a_{3,\mathrm{Q}}}{2}{k_2}^*\right\} y_{\mathrm{RF}}(t) \\ & + \left\{\frac{a_{3,\mathrm{I}} + a_{3,\mathrm{Q}}}{2}k_2 + \frac{a_{3,\mathrm{I}} - a_{3,\mathrm{Q}}}{2}{k_1}^*\right\} y^*_{\mathrm{RF}}(t) \\ & + \left\{\frac{a_{4,\mathrm{I}} - a_{4,\mathrm{Q}}}{8}\left[{k_1}^3 + 3k_1({k_2}^*)^2\right] + \frac{a_{4,\mathrm{I}} + a_{4,\mathrm{Q}}}{8}\left[({k_2}^*)^3 + 3{k_1}^2{k_2}^*\right]\right\} y^3_{\mathrm{RF}}(t) \\ & + \left\{\frac{a_{4,\mathrm{I}} - a_{4,\mathrm{Q}}}{8}\left[{k_2}^3 + 3({k_1}^*)^2k_2\right] + \frac{a_{4,\mathrm{I}} + a_{4,\mathrm{Q}}}{8}\left[({k_1}^*)^3 + 3{k_1}^*{k_2}^2\right]\right\} \left[y^*_{\mathrm{RF}}(t)\right]^3 \\ & + \left\{\frac{a_{4,\mathrm{I}} - a_{4,\mathrm{Q}}}{8}3\left[{k_1}^2k_2 + |k_2|^2{k_2}^* + 2|k_1|^2{k_2}^*\right]\right. \\ & \quad \left. + \frac{a_{4,\mathrm{I}} + a_{4,\mathrm{Q}}}{8}3\left[{k_1}^*({k_2}^*)^2 + |k_1|^2k_1 + 2|k_2|^2k_1\right]\right\} y^2_{\mathrm{RF}}(t)y^*_{\mathrm{RF}}(t) \\ & + \left\{\frac{a_{4,\mathrm{I}} - a_{4,\mathrm{Q}}}{8}3\left[k_1{k_2}^2 + |k_1|^2{k_1}^* + 2|k_2|^2{k_1}^*\right]\right. \\ & \quad \left. + \frac{a_{4,\mathrm{I}} + a_{4,\mathrm{Q}}}{8}3\left[({k_1}^*)^2{k_2}^* + |k_2|^2k_2 + 2|k_1|^2k_2\right]\right\} y_{\mathrm{RF}}(t)[y^*_{\mathrm{RF}}(t)]^2. \end{aligned} \tag{2.44}$$

According to (2.22), $y_{\mathrm{RF}}(t)$ contains the original signal $x(t)$ and its RF distortion, hence, both components have also mirror components stemming from $y^*_{\mathrm{RF}}(t)$, $[y^*_{\mathrm{RF}}(t)]^2$, and $[y^*_{\mathrm{RF}}(t)]^3$. The other terms $y^2_{\mathrm{RF}}(t)y^*_{\mathrm{RF}}(t)$ and $y_{\mathrm{RF}}(t)[y^*_{\mathrm{RF}}(t)]^2$ cause further IMD around the already existing RF distortion. Note that (2.44) can be simplified to (2.33) and (2.34) assuming $k_1 = 1, k_2 = 0$

and separating in I and Q component. Assuming perfectly aligned I and Q paths ($a_{3,\mathrm{I}} = a_{3,\mathrm{Q}}$ and $a_{4,\mathrm{I}} = a_{4,\mathrm{Q}}$) reduces (2.44) to

$$y'_{\mathrm{BB}}(t) = a_3 y_{\mathrm{RF}}(t) + \frac{1}{4} a_4 \left[y^*_{\mathrm{RF}}(t)\right]^3 + \frac{3}{4} a_4 y^2_{\mathrm{RF}}(t) y^*_{\mathrm{RF}}(t). \tag{2.45}$$

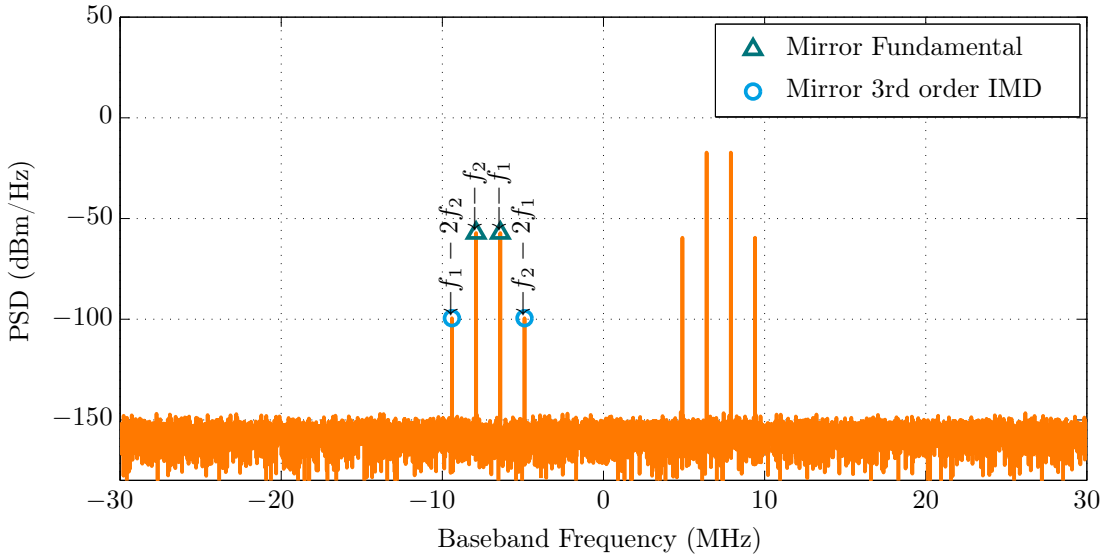
In the following, two-tone simulations are performed to demonstrate the impact of the imbalance effects and their interaction with the RF and BB non-linearity. The used parameters are summarised in Table 2.4. One important parameter is the IRR, the receiver's ability to suppress mirrors of the RF signal to be I/Q down-converted. The value is set to 40 dB, which is a realistic value for a typical DCR as will be seen later in measurements. Moreover, a small

Table 2.4: I/Q imbalance simulation parameters

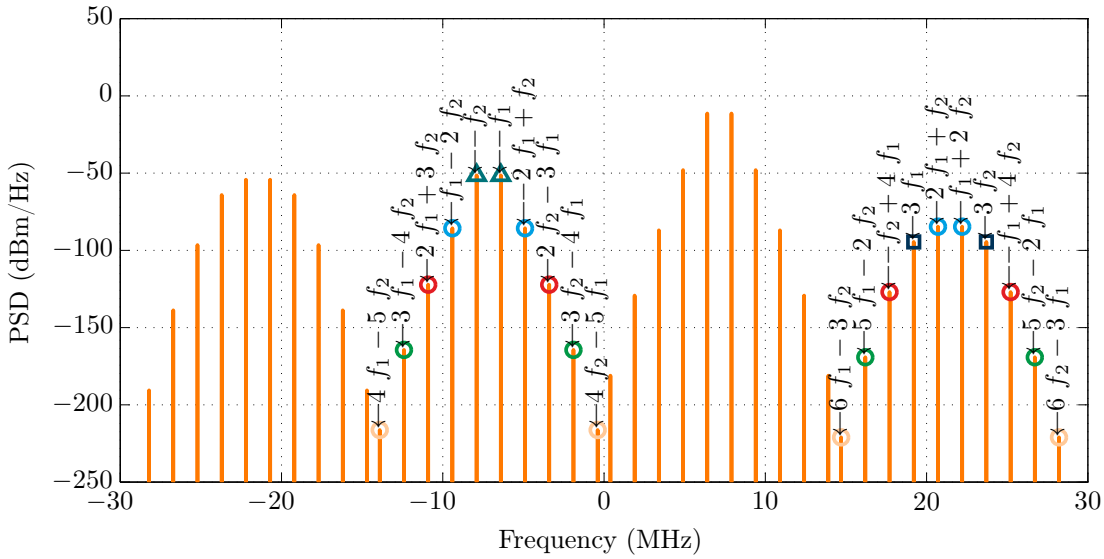
IRR	40 dB	$a_{3,\mathrm{I}}$	2
g_{m}	0.99	$a_{3,\mathrm{Q}}$	2.2
ϕ_{m}	$0.0174 \approx 1^\circ$	$a_{4,\mathrm{I}}$	-1000
k_1	$0.9949 - j0.0086$	$a_{4,\mathrm{Q}}$	-900
k_2	$0.0051 - j0.0086$		

phase mismatch of the mixer has been assumed that results in complex mismatch coefficients k_1 and k_2 according to (2.40) and (2.41). Slightly different third-order coefficients $a_{3,\mathrm{I/Q}}$ and $a_{4,\mathrm{I/Q}}$ take also I/Q imbalance at the two physically separated I/Q paths into account.

After $y_{\mathrm{RF}}(t)$ passes the mixer, mirror products of the fundamental tones and RF distortion appear, as shown in Figure 2.22(a). Compared to Figure 2.15(c), only additional components due to the imbalance effects are highlighted. Feeding $\tilde{y}_{\mathrm{RF}}(t)$ to the BB stages yields the total RF and BB distortions under I/Q imbalance conditions according to (2.44). In Figure 2.22(b), additional components compared to those in Figure 2.20 are highlighted. Due to the imbalance effects, basically same frequency components are generated as listed in Section 2.5.5, but all with negative sign (complex conjugates) and with reduced power levels dictated by the IRR. In other words, the spectral gaps in Figure 2.8 are filled due to the mirror effects. Note that also mixer non-linearity and further BB non-linearities can be combined in the non-linear coefficients $a_{3,\mathrm{I/Q}}$ and $a_{4,\mathrm{I/Q}}$. Different third-order coefficients between the I and Q path result in different power levels of distortions, but same frequency components are generated. It just denotes different scaling of the real and imaginary parts that sum up in the complex-valued representation.



(a) RF distortion after passing a non-ideal mixer manifesting I/Q imbalance.



(b) Further distortion at BB with I and Q paths incorporating different gain and third-order coefficients.

Figure 2.22: RF and BB non-linear distortion under I/Q imbalance conditions.

2.5.7 Modelling of Memory Effects

In case of wideband signals, the frequency-dependent behaviour of the front-end cannot be neglected any more [Sch09]. That is, the limitations of memoryless polynomials reported in Section 2.5.2 become problematic, if the signal bandwidth becomes comparable with the bandwidth of the non-linear front-end. Memory effects can be itself linear or non-linear, and short- or long-term, respectively [Sch09].

The most general model for describing both non-linearities and memory effects is the Volterra series [VR03; Ken00; Pin01]. It provides a complete and reliable description of the non-linear system's function by using multidimensional convolutions [Pin01]. A finite discrete-time Volterra series model is given by [MMK+06]

$$y(s) = \sum_{n=1}^{N} y_n(s) \tag{2.46}$$

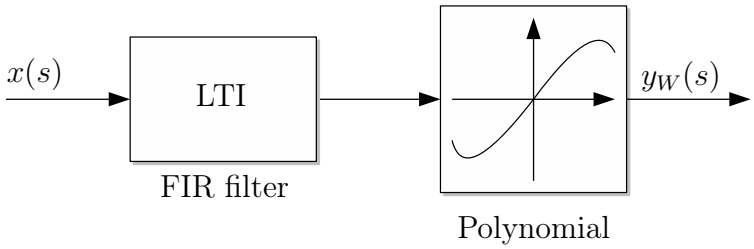
$$y_n(s) = \sum_{m_1=0}^{M-1} \cdots \sum_{m_n=0}^{M-1} h_n(m_1, \cdots, m_n) \prod_{l=1}^{n} x(s - m_l), \tag{2.47}$$

$y_n(s)$ being the n-dimensional convolution of the input with Volterra kernel h_n. It is obtained by combining the convolution sum of a linear time invariant (LTI) system, modelling the memory, and the power series according to (2.7), modelling a memoryless non-linearity. Thus, the set of Equations (2.46) and (2.47) is also called a power series with finite memory of length M [Rou09]. The Volterra kernels denote different transfer functions (impulse responses) for each order n. For $n = 1$, the kernel $h_1(m_1)$ coincides with a linear impulse response. Non-linearity is characterised by any higher-order impulse response ($n > 1$), e.g. $h_2(m_1, m_2)$ and $h_3(m_1, m_2, m_3)$ describe the second- and third-order non-linearity, respectively. However, the Volterra model also has some drawbacks. First, the parameters increase with respect to non-linear order N and memory depth M significantly, essentially complicating its practical implementation. Second, high-order Volterra kernels are difficult to measure and convergence of the Volterra series is also not guaranteed. Third, the Volterra model shows, due to the Taylor series approximation, only good approximation at the point where it has been extended, i.e. it suits for modelling weakly non-linear systems (mildly non-linear regimes) [Sch09].

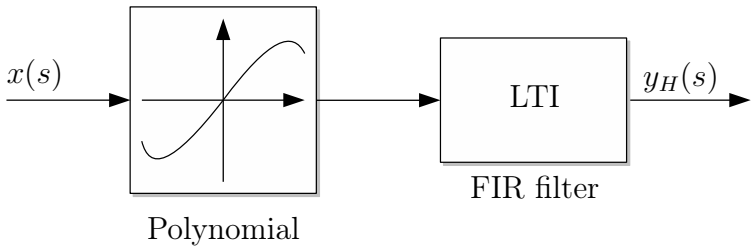
Next, some simpler memory structures are considered that model linear memory and constitute special cases of the general Volterra formulation.

Modelling of Linear Memory

In general, two-box, three-box, and parallel-cascade models are distinguished [Sch09]. Most common simplified two-box structures of the Volterra approach are the Wiener and Hammerstein model [Sch09]. These are cascades of a static non-linear system (memoryless polynomial) and a memory system, implemented by a linear filter (LTI subsystem). If the linear filter comes before the non-linearity, it is called Wiener model, otherwise Hammerstein. Both structures are depicted in Figure 2.23(a) and Figure 2.23(b), respectively. The Wiener model



(a) The Wiener model (linear – non-linear).



(b) The Hammerstein model (non-linear – linear).

Figure 2.23: Two-box models as simplifications of the Volterra series model [Sch09].

is equivalent to taking just

$$h_n(m_1, \cdots, m_n) = c_n h(m_1) \cdots h(m_n), \quad n = 1, 2, \ldots N \tag{2.48}$$

in (2.47), hence, the model can be written as [MMK+06]

$$y_W(s) = \sum_{n=1}^{N} c_n \left[\sum_{m=0}^{M-1} h(m) x(s-m) \right]^n. \tag{2.49}$$

On the contrary, the Hammerstein model, which is formed by a static non-linearity followed by a common linear filter $h(m)$, can be written as

$$y_H(s) = \sum_{m=0}^{M-1} h(m) \sum_{n=1}^{N} c_n x^n(s-m). \tag{2.50}$$

Here, the equivalent Volterra kernels are given by

$$h_n(m_1, \cdots, m_n) = \begin{cases} c_n h(m) & m_1 = m_2 = \cdots = m_n, \\ 0 & \text{otherwise}, \end{cases} \tag{2.51}$$

i.e. the kernels are only non-zero along their diagonals [Sch09].

The Hammerstein model acquires special significance since this structure is used later in the mitigation architecture (Chapter 4). Finally, all modelling of non-linearity is still based on memoryless polynomials, but memory effects are considered by a cascade with an adaptive filter (AF) manifesting the linear subsystem.

The model memory depth M determines the length of the finite impulse response (FIR) filter adopted in these models. It can be estimated from the first-order Volterra kernel. There are several approaches to obtain the linear and non-linear block coefficients, the most common being the least squares (LS) method from broadband time-domain measurements. Details on that and of further structures, such as the three-box Wiener-Hammerstein model or the parallel-cascade models, can be found in [Sch09].

Modelling of Non-Linear Memory

The aforementioned Wiener and Hammerstein models assume that the qualitative shape of the conversion does not change with the frequency. Hence, no interaction between the tones of a wideband signal is considered, even though such a behaviour occurs in real devices with finite memory [Sch09]. There are more accurate non-linear models that consider both linear and non-linear memory effects, such as the memory polynomial [MMK+06; Dup12; ZWC10] or the time-delay neural network (TDNN) [Sch09], among others.

The Hammerstein model can be generalized by choosing different filters for each non-linearity order n [MMK+06]

$$y_{GH}(s) = \sum_{n=1}^{N} \sum_{m=0}^{M-1} a_{n,m} x^n(s-m), \tag{2.52}$$

where $\{a_{n,m}\}$ denotes a combination of different filter and polynomial coefficients c_n into a 2D-array. By assuming that the signal's bandwidth is small compared to the centre frequency f_c, the real input signal can be expressed by its complex BB representation x and yields the memory polynomial, similar to (2.8),

$$y_{MP}(s) = \sum_{n=1}^{N} \sum_{m=0}^{M-1} a_{n,m} x(s-m) \left|x(s-m)\right|^{n-1}. \tag{2.53}$$

Note that (2.53) ends up in a memoryless polynomial model according to (2.8) for $M = 0$. The memory polynomial can be seen as a parallel structure of different Hammerstein subsystems. Equation (2.53) can be simplified further by taking only the odd-order terms $2(n-1)$ as the even-order terms fall at frequencies very far from f_c. A block scheme of a memory polynomial is shown in Figure 2.24. The unit delay tap is depicted by the symbol

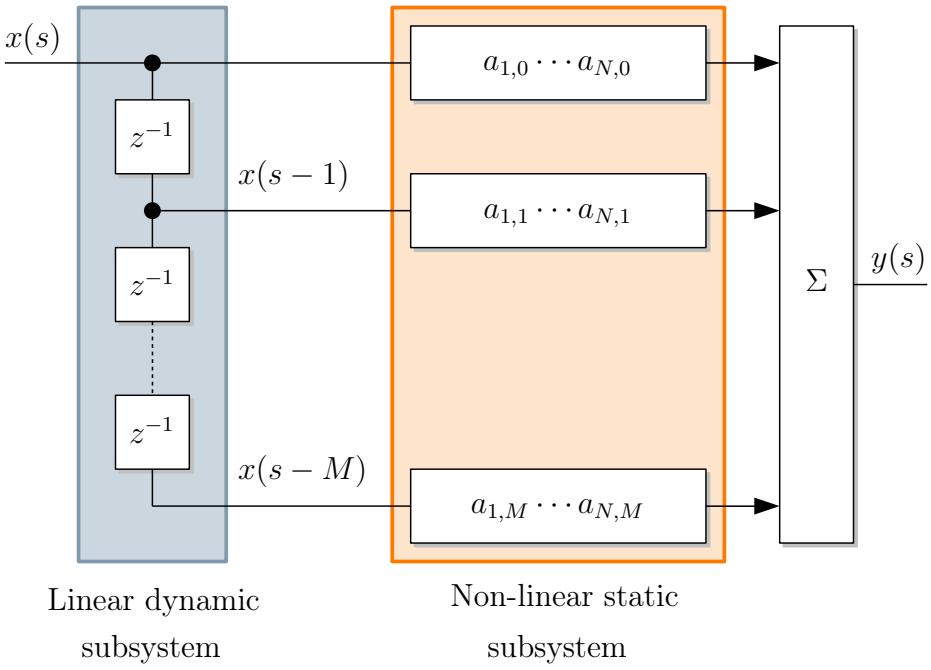


Figure 2.24: Block scheme of a memory polynomial model [Sch09].

z^{-1} and introduces a delay of one sampling interval m (unit time delay). The total number of parameters is $N \times M$, hence, easing the implementation of this model significantly compared to other memory models [Dup12].

The complex polynomial coefficients $a_{n,m}$ can be estimated by the LS method from L samples of the complex envelope BB signal. Assuming that input and output signal are collected in two vectors $\underline{x} = [x(0), x(1), \ldots, x(L-1)]^T$ and $\underline{y} = [y(0), y(1), \ldots, y(L-1)]^T$, the matrix $\mathbf{R}$ is given by

$$\begin{aligned} \mathbf{R} &= [\mathbf{R}_0, \mathbf{R}_1, \ldots, \mathbf{R}_m], \\ \mathbf{R}_m &= \left[\mathbf{x}_m, \mathbf{x}_m |\mathbf{x}_m|, \mathbf{x}_m |\mathbf{x}_m|^2, \ldots, \mathbf{x}_m |\mathbf{x}_m|^{N-1}\right], \\ \underline{x}_m &= [x(0-m), x(1-m), \ldots, x(L-1-m)]^T, \end{aligned}$$

with $m = 0 \ldots M-1$. Then, (2.53) can be written in vectorial notation as

$$\underline{y}_{MP} = \mathbf{R}\underline{a}, \tag{2.54}$$

where $\underline{a}$ is the vector with the complex polynomial coefficients $a_{n,m}$ arranged as $\underline{a} = [a_{1,0}, \ldots, a_{N,0}, a_{1,1}, \ldots, a_{N,1}, \ldots, a_{1,M} \ldots, a_{N,M}]^T$. The LS solution for (2.54) is

$$\underline{a} = \left(\mathbf{R}^H \mathbf{R}\right)^{-1} \mathbf{R}^H \mathbf{y}, \tag{2.55}$$

where $(.)^H$ denotes the Hermitian transpose of the matrix $\mathbf{R}$ [Sch09; Dup12]. A development of a memory polynomial model based on measured data is discussed in Section 2.7.3.

In the TDNN approach, the memoryless non-linear subsystem is modelled by an artificial neural network. However, as it is trained to a given data set, its usability for other data is questionable. The interested reader is referred to [Sch09].

2.6 Circuit-level Simulation with SimRF

Although mathematical behavioural modelling is on focus in this thesis, some abstract circuit-level simulations have been conducted for the purpose of comparisons. In the following, a modern RF circuit simulation approach is presented that can be easily integrated in a system-level simulation framework.

SimRF is an extension of MATrix LABoratory (MATLAB)'s system-level simulation tool Simulink and provides a dedicated library for modelling of RF systems including its impairments [Sim; Zuc13]. With traditional circuit models based on simulation program with integrated circuit emphasis (SPICE), that process true RF passband signals, it is fairly impossible to assess the complete RF chain. Instead, a smarter way is desirable to deal with the RF signals, i.e. to apply models at higher levels of abstraction and solvers with a larger time steps. Circuit-envelope simulation, being one simulation mode in SimRF beside traditional BB equivalent modelling, is a possible solution to reduce the simulation time while taking into account multiple RF carriers at the input. SimRF enables design and simulation of wireless transceivers with at higher level of abstraction by basic blocks, such as mixers, amplifiers, and S-parameter blocks. There are input and output ports that behave as gateways and allow to translate Simulink signals to voltages that have a certain power and are centred around a certain carrier frequency. Thereby, SimRF blocks can be easily connected to other Simulink blocks, such as signal generation or spectrum indication (fast Fourier transform (FFT) sink).

Figure 2.25 shows an example of a SimRF model for the USRP WBX front-end. It is based on the schematic [Wbx] and data sheets of its discrete components, for inputting realistic parameters into the blocks. The total model needs to be refined into a chain of components

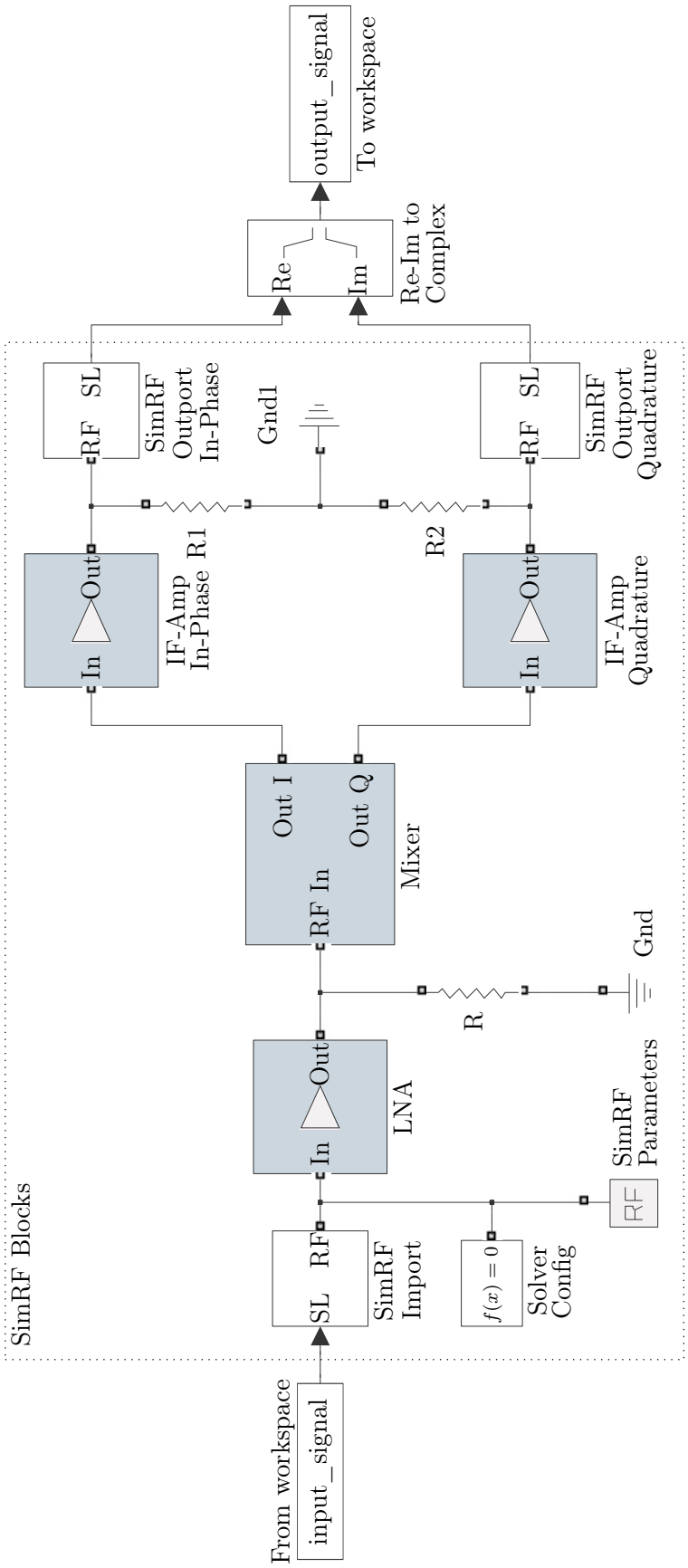


Figure 2.25: Top-level Simulink model of the WBX front-end.

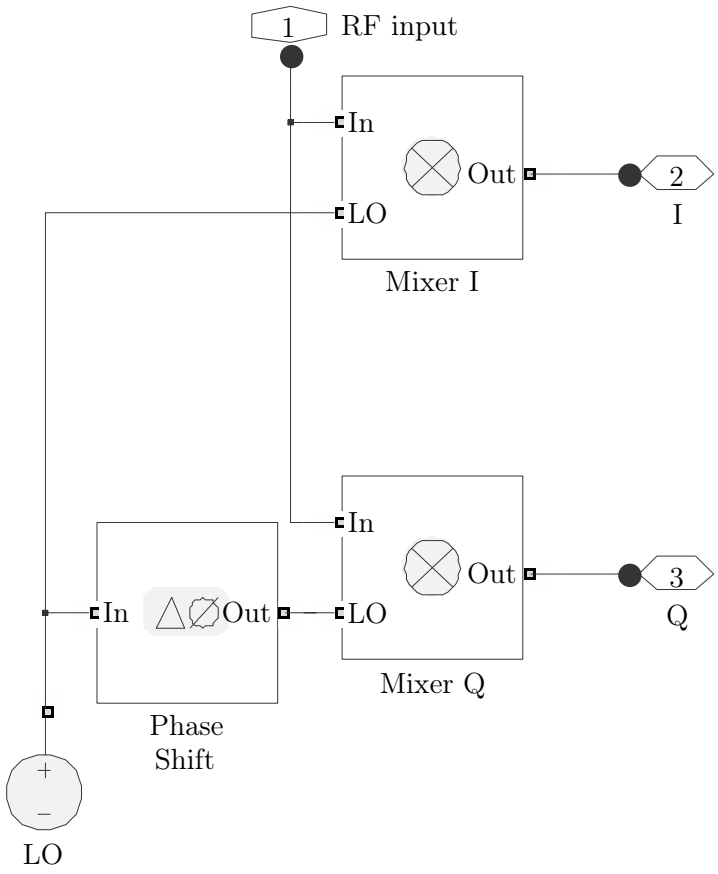


Figure 2.26: Simulink model of the mixer subsystem in Figure 2.25.

that actually build up the receiver. Essentially, the model represents the low-IF receiver architecture from the RF down to the BB amplifier similar to Figure 2.6. However, ADC and analogue BB filters are neglected in this simulation, as well as LO phase noise or mixer leakage effects. Figure 2.26 depicts the mixer subsystem that is part of the top level model Figure 2.25. It implements a quadrature I/Q mixer based on the mixer core blocks of SimRF.

The SimRF simulation provides a realistic representation of the HD and IMD product levels. I/Q imbalance effects as well as BB distortions can be easily tracked and match the output of the mathematical cascaded model perfectly, as depicted in Figure 2.27 for a two-tone excitation. Compared to the fully BB equivalent mathematical model, the circuit-level simulation has been simulated at $f_c = 920\,\text{MHz}$. Same input power, noise level, IRR and intercept points for RF and BB amplifier have been assumed in both simulations. However, memory effects have not been simulated and seem not to be included in the SimRF blocks itself. In [Zuc13] it is said that purely static ninth order polynomial expressions are behind the SimRF amplifier blocks, hence producing same distortion profiles. However, own blocks can

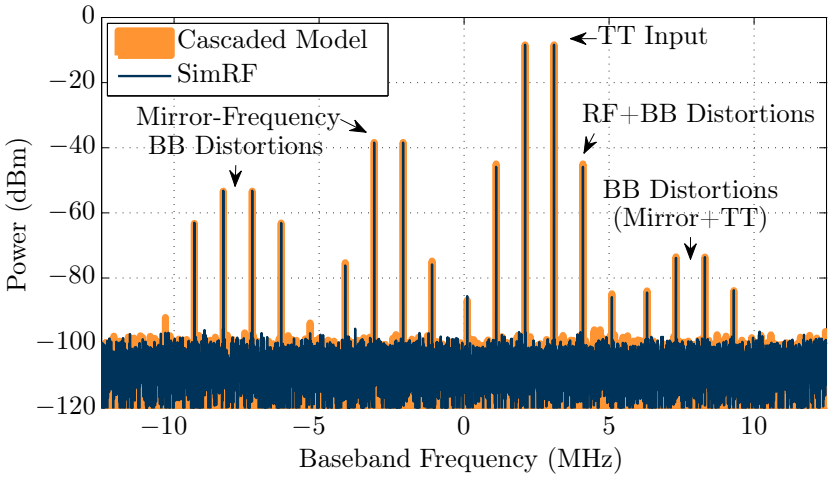


Figure 2.27: Comparison between behavioural and circuit model.

be built by using the SimScape language. On the other hand, much more different RF impairments can be included in the SimRF simulation, e.g. the LO mixer feed-through by a resistor between LO and RF input port of the mixer, or impedance mismatches between the several stages. Thereby, severe imperfections can be modelled that go beyond the simple BB equivalent mathematical model or are at least difficult to model at this level.

2.7 Measurement-based Analysis

2.7.1 Model Identification Techniques

Beside physical and behavioural modelling, real RF device measurements with realistic excitation signals can be conducted. There are two approaches to derive a mathematical model out of measured data, namely mathematical modelling and system identification [Sch09; Pin01]. In mathematical modelling, basic laws of physics are used to describe the system's dynamic behaviour. However, the few physical insights are unlikely to work even with simple single-stage receivers that are on focus in this thesis. Instead, the DCR should be modelled as a whole. The idea of system identification is to fit a model to the data extracted from empirical measurements, therefore, it is also called direct empirical modelling [Sch09]. It requires less physical insight and eases the construction, but is limited in validity.

System Identification Procedure

First, available a-priori knowledge concerning the RF chain and the desired model application is used to determine the type of excitation signal, the measurement setup including the measurement procedure and the operating range. Here, the main goal is to extract a simple model that describes the non-linear behaviour of the receiver to optimise the performance of the radio by RF impairment mitigation techniques. In general, the test signal should excite all possible states of the non-linear front-end (persistent excitation) [Sch09]. The numbers of coefficients should be chosen in a way that enough degrees of freedom are available that describe the full range of the front-end behaviour. After parameter extraction from the measured data, the model is typically validated by using other data than in the extraction process. There are some factors that may lead to a deficient model [Sch09]. The measurements may not provide sufficient information or are too noisy, e.g. if the ADC quantisation noise covers higher-order distortion products. Next, the model may be inappropriate to represent the front-end behaviour, e.g. if the order is too low or too high. Finally, the model identification algorithm itself may fail to extract the parameters correctly.

Excitation Signal

There are several different types of excitation signals that can be used for system identification techniques [Sch09]. These are single-tone, two-tone, multi-sine, band-limited noise, pseudo-noise, or digitally modulated signals. In addition, the excitation signal is typically swept over the input frequency and power range of the device under test (DUT) to capture its complete behaviour.

The excitation signal class plays a key role for extracting a valid model. For instance, the IMD caused by a two-tone interferer will differ from the IMD caused by a wideband digitally modulated signal. The distortions, and hence the frequency response of the DUT, depend on the amplitude distribution of the excitation signal [Pin01]. In order to get the best linear approximation, it is important to use the same kind of excitation in terms of power spectrum and amplitude distribution as will be applied later during operation of the radio [TGT+96]. Otherwise, the approximation can become invalid. Assuming a fully memoryless behaviour, the excitation signal does not need to match the application signal, however, swept-frequency response should be measured to establish the band limits of the memoryless property.

2.7.2 Two-tone Excitation

Performing a two-tone (TT) analysis to characterise the extent of circuit non-linearity is a classical approach especially in RF and microwave engineering. In general, measuring the generated spectra, i.e. HD and IMD, is easier than evaluating the variations of the fundamental signal, which is done for obtaining the AM/AM and AM/PM characteristics [VR03]. However, the analysis can become very complex if three or more tones are present at the input of the non-linear DUT. Therefore, it is a common practice to limit the analysis to two tones [Anr00]. A two-tone excitation signal is simple to create and shows a non-constant envelope with a PAPR = 6 dB. On the other hand, two tones are not representative for wideband communication signals in strict sense, as the amplitude distribution and power spectrum do not match common communication signals [TGT+96]. Although memory effects remain hidden due to the narrowband nature of the two-tone signal, bandwidth-dependent effects can be recognized if the phase of the IMD product differs at low and high tone spacing [VR03; Sch09].

Model Extraction

A memoryless polynomial model according to (2.7) can be directly extracted by the intercept points obtained by prior RF measurements. Intercept points provide a good indication of the degree of non-linearity under simple signal conditions [Ken00], whereas digitally modulated signals allow only for little insight in the operation of the analysed system. The points are typically found by extrapolation of the IMD power against different input powers. That is, fundamental and IMD product are plotted in a power level diagram and intercept points are found by graphical intersection of their extrapolated curves.

Measurement Setup

The two-tone measurement setup is sketched in Figure 2.28. The two-tone signal is generated by two individual signal generators, whose outputs are combined by a power splitter in reverse mode. A controllable attenuator is used to set the desired signal power and to provide enough isolation of the individual tones. Finally, the signal is fed to the USRP2+WBX, the non-linear DUT. Generators and attenuator are connected via general purpose interface bus (GPIB), a control bus for instrumentation and measurement that enables full-automated operation. Moreover, the setup enables coherent sampling by

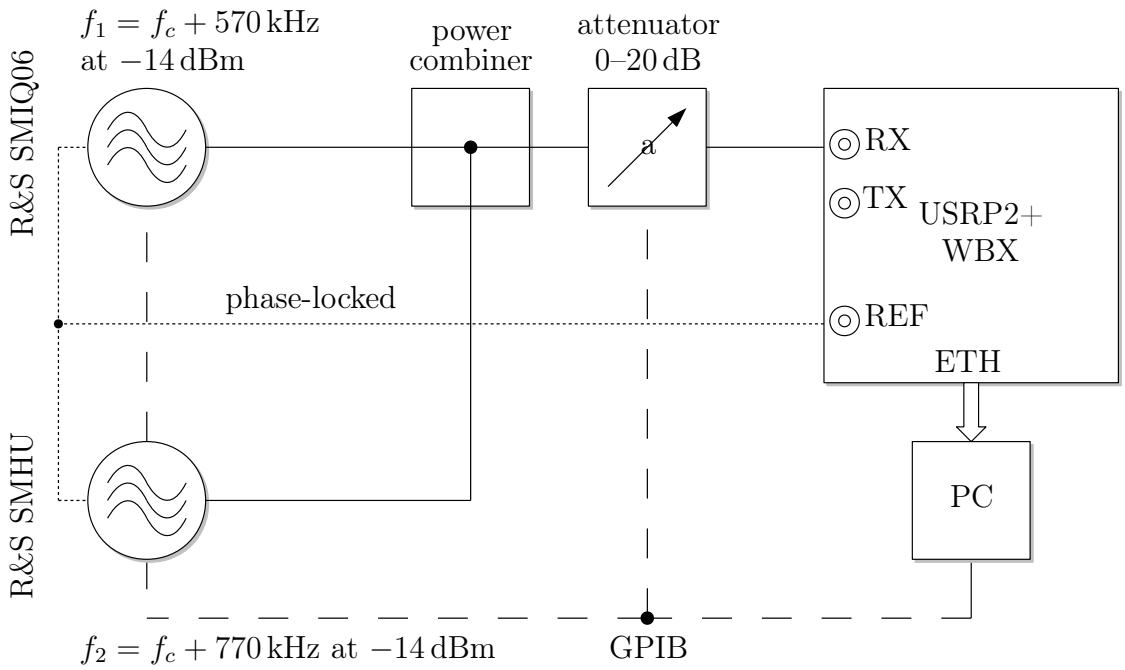


Figure 2.28: Sketch of the two-tone measurement setup.

phase synchronisation (10 MHz reference) between the generators and USRP. This is essential for avoiding spectral leakage and for obtaining absolute power levels. Note that the total RF front-end non-linearity is captured as a whole and determined by the FFT spectrum of the digitised BB signal. Thus, the sinusoidal frequencies are chosen as integer multiples of an FFT bin that represents the sampling grid in the frequency domain with

$$f_{\text{bin}} = \frac{f_s}{\text{NFFT}}, \tag{2.56}$$

where f_s and NFFT denote the sampling frequency and the FFT length, respectively. Indeed, the FFT can be interpreted as a filter bank with the bandwidth f_{bin} in each channel. The higher the FFT length, the lower the channel bandwidth of the FFT – a behaviour like the resolution bandwidth of a spectrum analyser. By choosing the tone frequencies within the FFT grid, its powers are concentrated on single FFT bins and allow for proper power detection. Moreover, the frequencies should be kept small so all distortions, especially those in the third harmonic zone, remain within the BB bandwidth. Finally, the distortion products at the sum- and difference frequencies of the two tones will also fall onto FFT bins. In order to avoid spectral leakage, integer number of signal periods should be obtained, in this case $N \cdot t_p = N \cdot 1/f_{\text{beat}} = 2N/(f_1 + f_2)$, where f_{beat} denotes the beat frequency of the two-tone complex envelope. The spectral leakage can be further reduced by adopting a

window function, such as the Hann window. However, when applying the aforementioned rules, almost no leakage will occur and windowing does not provide further improvement.

When aiming for characterizing the non-linearity of the DUT, all the measurement equipment should be spurious-free [Anr00]. That is, any residual IMD by signal generator should be avoided and appropriate filtering of the two-tone signal should be applied, if necessary. Next, signal combiners with a good input-to-input isolation are required, as well as attenuators for decoupling the generators. Before running measurements with the DUT, the spectral purity of the generated signal is to be checked, e.g. with a conventional spectrum analyser. In practice, the spurious-free dynamic range (SFDR) of the generators should exceed the receiver dynamic range.

A general drawback of this two-tone measurement setup is the inability to characterise AM/PM distortion. Extensions with vector network analysers and reference IMD generators, as well as broadband setups similar to modern communication systems have been reported in [Sch09].

For the actual RF measurements, the tone spacing was 200 kHz. A swept-tone measurement has been conducted at frequencies across the entire operating band of the analogue front-end WBX from 70 MHz to 2.2 GHz in steps of 67 MHz in order to check for potential frequency-dependent effects. The input power was solely set by the controllable attenuator. Gain losses of the DUT at higher frequencies were compensated by the attenuator.

Results

Figure 2.29 illustrates an example of the received BB spectrum at 472 MHz and −13 dBm input power. Beside second and third-order IMD in the fundamental zone, also BB distortions in the third harmonic zone are visible. In addition to that, there is a mirror frequency due to mixer I/Q imbalance, as well as a DC offset that results from the ADC offset and the LO-mixer feed-through (see Section 2.2). Due to the dominating RF distortion, third-order IMD products in the fundamental zone have been used for extracting the model coefficients. A gain loss of the RF front-end of 16 dB from the lower to the higher end of the frequency range has been observed and compensated by a reduced attenuation. A little imbalance between the upper and higher IMD3 products is visible, that indicate some memory effect.

The corresponding power level diagram at 472 MHz is illustrated in Figure 2.30. The measured fundamental and IMD powers are depicted in straight lines, whereas its extrapolated curves have dashed lines. The extrapolation is done by simple polynomial fitting with the assumed slopes of 1 : 1, 1 : 2, and 1 : 3 for the fundamental, IMD2, and IMD3, respectively

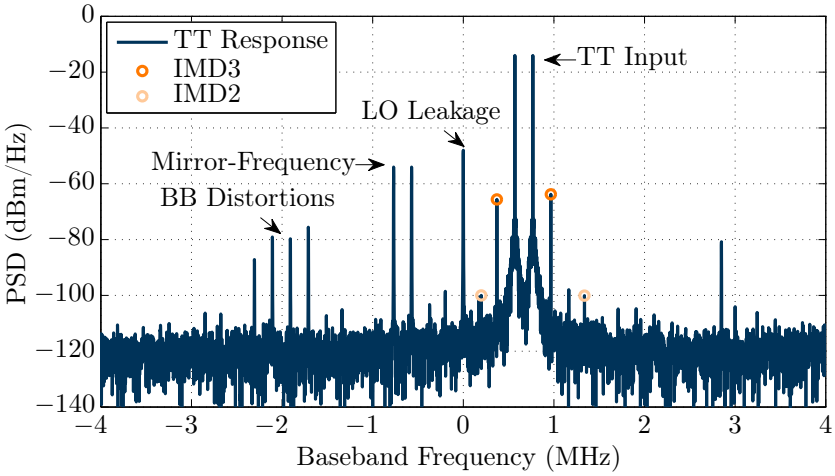


Figure 2.29: BB spectrum of a two-tone excitation measured at 472 MHz and −13 dBm with USRP2+WBX [Ett].

(see also Section 2.3.2). The measurement of the IMD power levels is challenging as the RF front-end is quite linear in general. Thus, clipping distortion at the ADC occur before the IMD products reach a measurable level. That is, the high distortions at around the P1dB are mainly induced by clipping. The resulting power level for P1dB, IIP2, and IIP3 in this case are: −11 dBm, 64 dBm, and 11 dBm, respectively.

The results obtained for the intercept points over the total receiver operating bandwidth are illustrated in Figure 2.31. Apart from the small notch for IIP2 at around 1.7 GHz, the non-linear behaviour is approximately frequency-independent and static over the input signal band. The average values for the intercept points are IIP2 ≈ 59.7 dBm and IIP3 ≈ 13.0 dBm. Assuming an RF receiver gain of $G = 10$ dB, the real coefficients c_n in Equation (2.7) can be computed by (2.15), (2.16), and (2.17) to the values in Table 2.5.

Table 2.5: Extracted coefficients of the memoryless polynomial.

Parameter	**Value**
G	10 dB
IIP2	59.7 dBm
IIP3	13.0 dBm
c_1	3.1623
c_2	0.0734
c_3	−106.0839

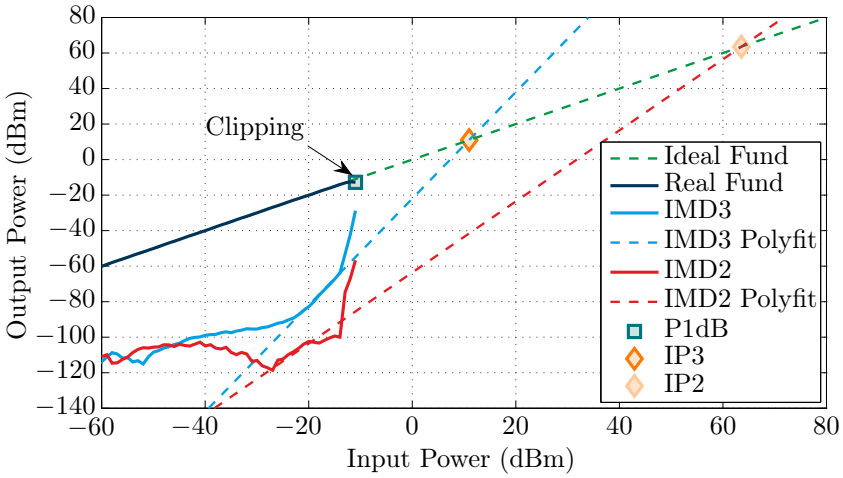


Figure 2.30: Power level diagram at 472 MHz highlighting the power of the IMD products including their extrapolation.

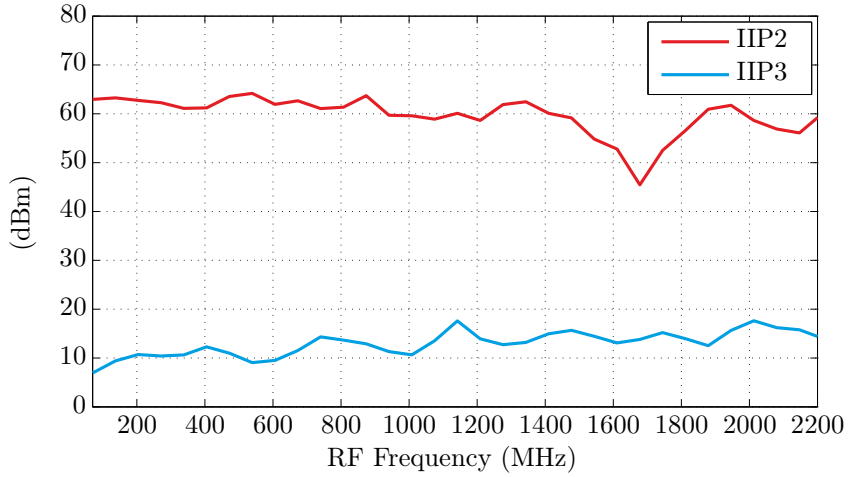
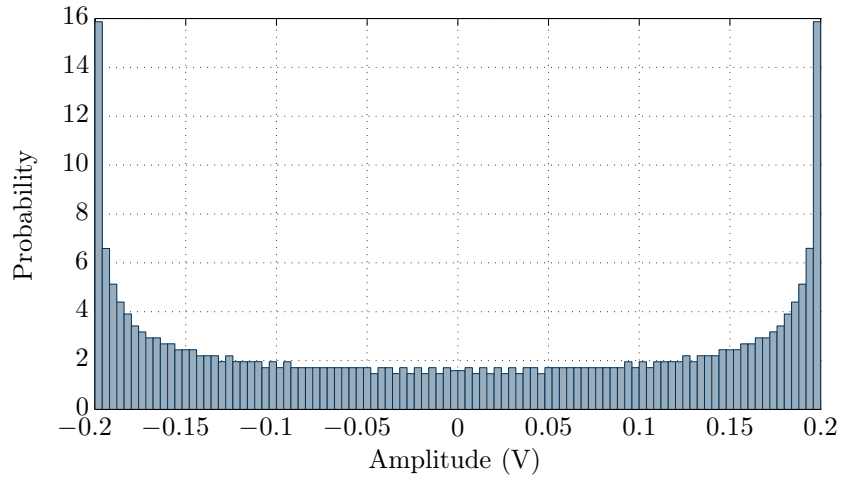


Figure 2.31: IIP2 and IIP3 measured with USRP2+WBX [Ett].

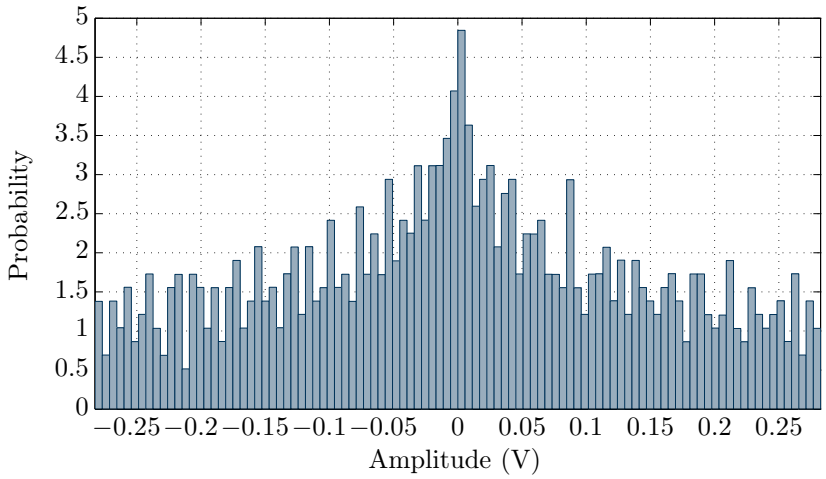
2.7.3 Wideband Test Signals

Multi-tone Signal

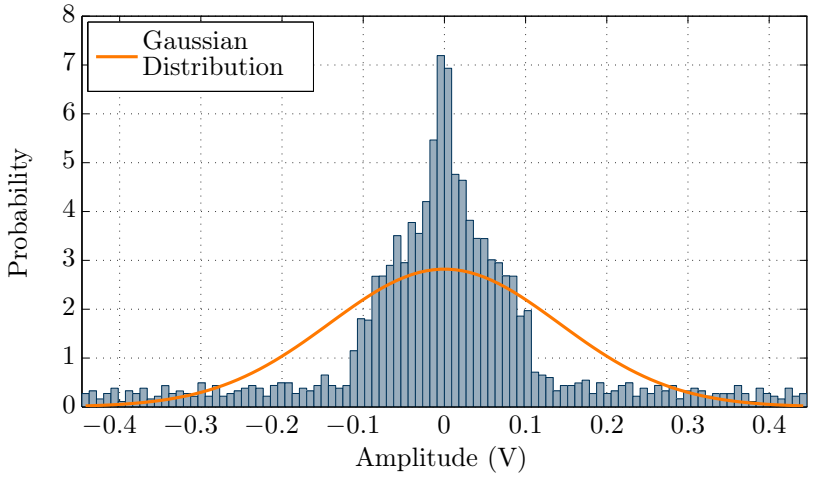
The previous two-tone signal is unrealistic in terms of its amplitude statistics and power spectrum, though it provides a rough description of the extent of receiver non-linearity and eases the extraction of a simple model. Figure 2.32 depicts the probability density functions (PDFs) of various CW signals. The single sinusoid has an U-shaped amplitude statistic, also known as bath tube curve, as illustrated in Figure 2.32(a). The peak envelope of the tone and the PAPR are 0.14 V and 3 dB, respectively. Combining two sinusoidal functions of different frequencies with 100 kHz spacing yields a superposition



(a) Single-tone signal (PAPR = 3 dB).



(b) Two-tone signal (PAPR = 6 dB).



(c) Multi-tone signal (5 tones, PAPR = 10 dB).

Figure 2.32: Amplitude statistics of various CW signals.

of two bath tube curves, as depicted in Figure 2.32(b). Note that the overall power is kept constant, by weighting the sum signal with its total power. The peak power adds up by factor 2 to 0.28 V, the resulting PAPR = 6 dB. Next, a multi-tone signal with 5 sinusoids is considered in Figure 2.32(c) to come closer to a realistic distribution of complex digitally modulated signals. Note that the phases of each tone are constant zero in all cases, hence, resulting in a high PAPR. The peak power and the PAPR are now 0.45 V and 10 dB, respectively. In addition, also the Gaussian normal PDF is plotted in Figure 2.32(c) that is typical for modern OFDM-like communication signals. According to the central limit theorem in probability theory, the sum of independent and identically distributed amplitude distributions of sine wine waves will tend to a Gaussian distribution for a large number of sine waves with random phases [TGT+96]. However, for this low number of tones, both the high and low voltages occur with higher probability than in the normal PDF. On the other hand, the signal's PAPR is significantly high to provoke severe distortion.

Periodic multi-tone signals are effective in reproducing the statistics of a desired realistic communication signal, as they produce similar levels of distortions [LHG+10; TGT+96]. In fact, they can be used to design a specific amplitude statistic by modifying the amplitude and phase of each tone. Compared to other wideband signals, such as pseudo random noise, it is limited in bandwidth. Details on the design of arbitrary signal statistics by multi-tone signals are given in [TGT+96].

Next, an RF measurement with the USRP N210+WBX is discussed [GSH+12c; GSH+12a]. A signal configuration of a strong 5-tone blocker at 5 MHz and a weak 5-tone desired signal at 2 MHz, having a higher tone spacing of 400 kHz, has been considered. The sampling rate and the centre frequency of the device were 16.67 MHz and 487 MHz, respectively. Both signals were generated by two arbitrary waveform generators (AWGs) and combined using a power splitter, similar to the setup in Figure 2.28. The distorted output BB spectrum as well as the ideal input signal are illustrated in Figure 2.33. The blocker signal causes severe IMD distortion in the fundamental and the mirror zone. As the distortions at the third harmonic fall beyond the sampling range, they appear in the DC zone. Some distortion products are even hitting the desired multi-tone signal, which can be distinguished by the different tone spacings adopted for the two multi-tone signals.

WCDMA Signal

In [Dup12], a memory polynomial model has been developed for the USRP with WBX front-end. Here, the explicit goal was to check the DUT for memory effects by inputting a wideband test signal with fast varying amplitude (high PAPR). For that purpose, a

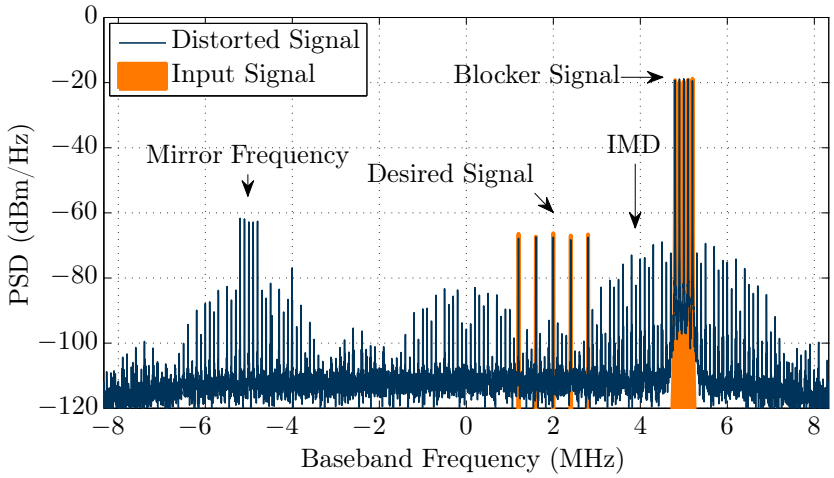


Figure 2.33: Distorted BB spectrum with multi-tone input, measured with USRP N210+WBX [Ett].

wideband code division multiple access (WCDMA) signal has been used that is adopted in the Universal Mobile Telecommunications System (UMTS) standard, better known as the third generation mobile cellular system.

The measurement setup is sketched in Figure 2.34. The WCDMA signal is generated in MATLAB by modulating random binary data with an offset-quadrature phase-shift keying and subsequent raised-cosine filtering. In WCDMA, the user data at relatively low rate is spread over a much wider bandwidth of 5 MHz, by using a sequence of pseudo-random units called chips (3.84 Mcps).

Next, the generated signal is uploaded to the signal generator R&S SMU200A that has an integrated AWG for replaying and up-converting the provided I/Q sample data stream. Signal generator and the non-linear DUT, USRP+WBX, are synchronized by a 10 MHz

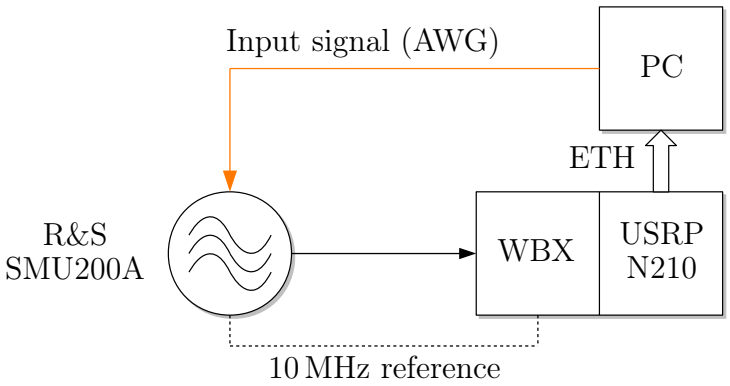


Figure 2.34: WCDMA Measurement setup.

reference clock. As the actual receiver input signal cannot be measured by this setup, it is assumed that all non-linear distortion is generated in the receiver itself. The SFDR of the signal generator is sufficiently high and supports this assumption. A generator power of −16 dBm has been used to stay within the mildly non-linear regime. The centre frequency and the sampling rate of the DUT were 120 MHz and 25 MHz, respectively. A complete WCDMA frame length of 10 ms has been captured (250 000 samples in total). This high number of samples has been used for reasons of accuracy. After reception, the output data is aligned according to the input data by correlation. The amplitude distribution (histogram) is depicted in Figure 2.35. As just the modulation (waveform) is generated without taking

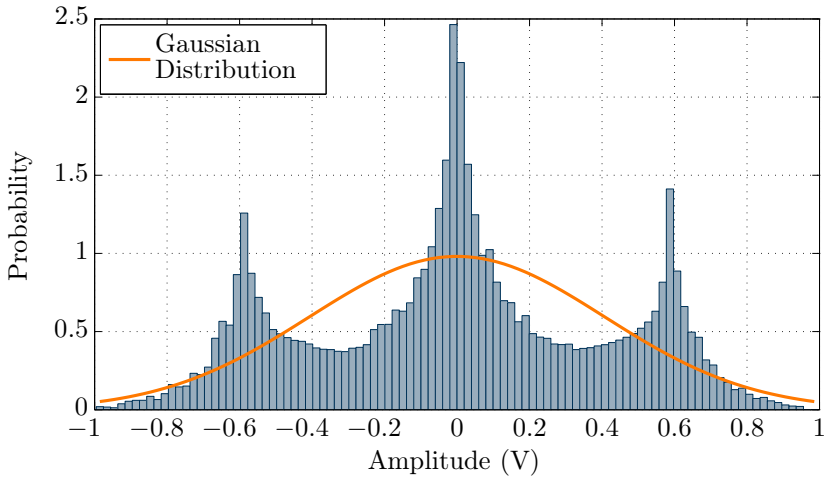


Figure 2.35: Amplitude distribution of the measured WCDMA signal (PAPR = 5.1 dB).

into account the specific WCDMA frame structure, the output signal is not Gaussian. It has a moderate PAPR of 5.1 dB in order to avoid CD at the receiver. Figure 2.36 illustrates the ideal input and measured output WCDMA spectrum. It clearly shows the generated IMD in the fundamental zone.

Next, a memory polynomial (2.53) is extracted out of this WCDMA measurement. First, the memory depth M and the polynomial order N need to be estimated [FB11; WLR+06]. This is done by the method of false nearest neighbours [Dup12]. That is, the normalised mean square error (NMSE) between the measured output $\mathbf{y}$ and the model estimation $\hat{\mathbf{y}}_{MP}$ is evaluated until a preset criterion is reached. The NMSE is defined as

$$\mathrm{NMSE}|_{\mathrm{dB}} = 10 \log_{10} \left(\frac{\|\mathbf{e}\|_2^2}{\|\mathbf{y}\|_2^2} \right), \tag{2.57}$$

where $\|\mathbf{e}\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}_{MP}\|^2$. Figure 2.37 depicts the error obtained for different polynomial orders and memory depths. With a memory order $M = 1$, an improvement of approx.

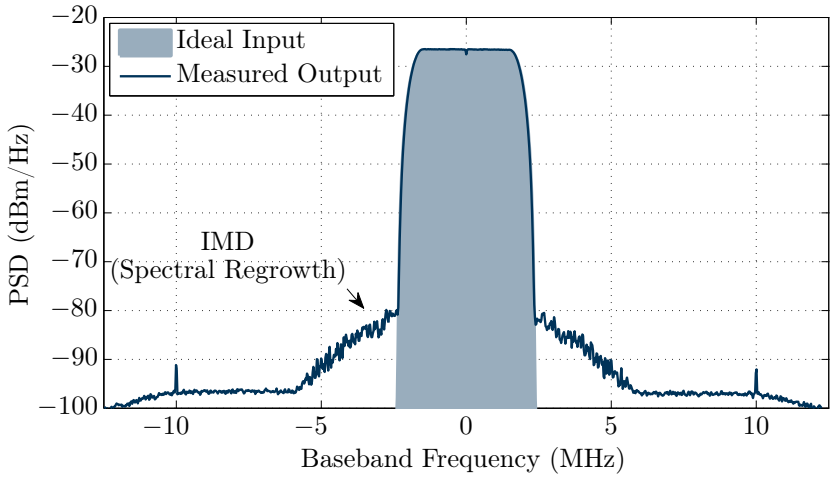


Figure 2.36: BB spectrum of the input and output WCDMA signal at 120 MHz and −16 dBm.

25 dB is achieved. The NMSE decreases only slightly further for $M = 2$ and onwards. In order to compare the results with the memoryless polynomial extracted in Section 2.7.2, $N = 3$ and $M = 1$ are chosen. The complex coefficients are computed by the LS method according to (2.55). The only critical point is the matrix inversion in (2.55), whose accuracy is expressed in terms of the condition number. Orthogonal memory polynomials have been proposed in [Dup12] that have lower conditions numbers. However, the NMSE performance of conventional and orthogonal memory polynomial are almost identical for the obtained data. For the aforementioned WCDMA data set, the computed coefficients of the memory polynomial are

$$\mathbf{a}_{n,m} = \begin{pmatrix} 0.1031 + j0.0868 & 0.7569 + j0.4423 \\ -0.0045 - j0.0045 & -0.0015 - j0.0031 \\ 0.0022 + j0.0028 & -0.0100 + j0.0046 \end{pmatrix}. \tag{2.58}$$

Note that the second-order coefficients are very small, thereby validating the low significance of the even-order distortion studied in earlier measurements. In Figure 2.38, both memory and memoryless polynomial models are visually compared based on the captured BB spectrum. The NMSE of the memoryless and the memory model are −4.5 dB and −26.2 dB, respectively, indicating a significant improvement of the memory model for wideband excitation, even though the magnitude response looks identical. There is only a significant small difference outside the triple bandwidth of the signal. However, this is due to the real-valued coefficients of the memoryless polynomial that cannot mimic the phase distortion of the receiver. A validation with the time domain signal yields that there is a significant phase mismatch of the waveform predicted by the memoryless polynomial, compared to the

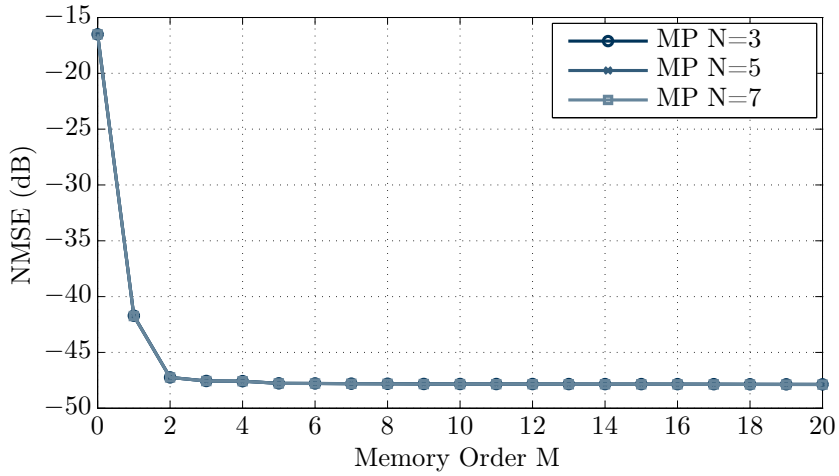


Figure 2.37: Error of memory polynomial modelling for different polynomial orders N and increasing memory order M obtained with a measured WCDMA signal.

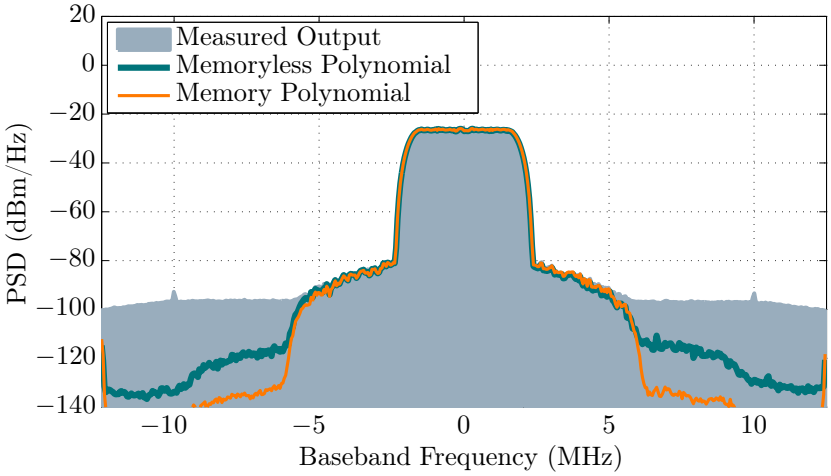


Figure 2.38: Comparison between memory and memoryless polynomial modelling.

memory model which follows almost ideally the measured curve. That is, validating the model predictions based on the power spectrum may lead to wrong conclusions. Finally, the memory polynomial output better estimates the distortion caused by wideband input.

This exposes that memory effects are present in the DUT and should not be neglected. From the system identification point of view, the memory model performs better than the memoryless polynomial in case of wideband excitation. However, for the purpose of distortion mitigation, memoryless polynomials are adopted due to the ease of implementation and sufficient level of accuracy. In Chapter 4, it will be discussed how memory effects can be considered in distortion mitigation by still employing a static and memoryless polynomial as a reference non-linearity.

2.8 Summary about Receiver Non-linearity

This chapter has shown that odd-order IMD is of major concern in wideband receivers with limited pre-selection at RF. Sources, effects, and metrics of receiver non-linearity have been discussed in detail. Crucial scenarios occur, if weak desired and strong unwanted (blocking) signals are simultaneously present at the RF antenna input. In such cases, strong blocking signals can enter the front-end amplification and mixing stages and may cause non-linear distortions. For the purpose of distortion mitigation, it has been demonstrated that non-linearities at RF and/or BB can be efficiently modelled by static memoryless polynomials. Thereby, a novel cascaded model is developed that considers the total receiver line-up of a DCR. Finally, a measurement-based characterisation of a typical SDR under test (USRP) has been conducted to derive simple polynomial models with and without memory. To this end, two-tone and wideband WCDMA signals have been used as test signals.

3 Mitigation Techniques

This chapter presents various techniques to handle front-end non-linearity and its effects. Thereby, linearisation techniques for *transmitter* non-linearity are also briefly discussed, to justify that they are inadequate for alleviating receiver-induced distortion. Next, suitable approaches to mitigate *receiver* non-linearity are classified and evaluated regarding their efficiency and limitations. Afterwards, state-of-the-art "Dirty RF" techniques, that denote system-level techniques to alleviate RF impairments, are presented briefly. In general, this chapter studies two main techniques to handle receiver non-linearity in detail. These are preventing non-linear distortions by cancelling the blocker signal and mitigating the distortion products. Although the digital feedforward mitigation is on focus in the thesis and detailed in Chapter 4, analogue cancellation and spatial filtering of strong blocker signals are discussed as alternative solutions.

3.1 Overview

3.1.1 Transmitter Non-linearity

Mitigating PA non-linearity at the transmitter side has been tackled for decades. Although it does not match to the research problem of mitigating receiver non-linearity, it is worth studying the principles that are used therein. Finally, mitigation at transmitter and receiver side can be clearly distinguished.

There is a huge number of circuit-level techniques for RF PA linearisation that can be classified into three main categories [Ken00]

- Feedback techniques,
- Feedforward techniques, and
- Predistortion.

The following paragraphs briefly discuss these techniques from an abstract level. For implementation details, the interested reader is referred to [Ken00].

Feedback systems intuitively cancel the error by feeding back a reference signal formed upon the output, as illustrated in Figure 3.1. The distortion $d(t)$ is mitigated by the fraction

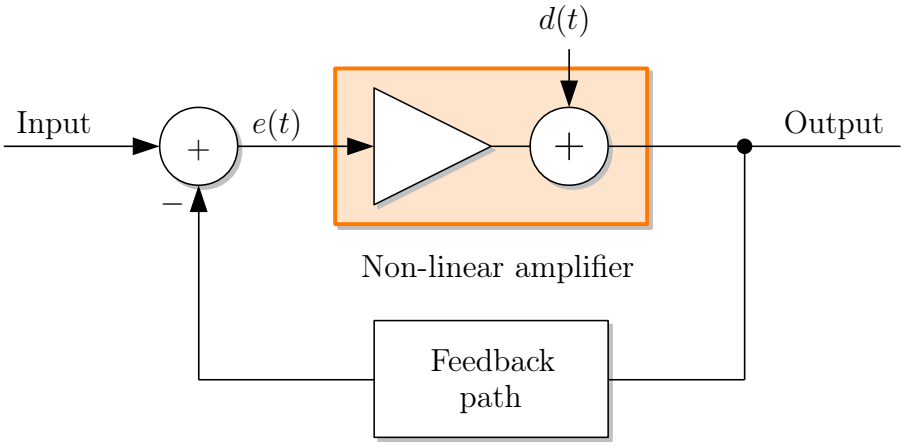


Figure 3.1: Feedback applied around a non-linear PA generating distortion [Ken00].

K/G, the ratio between the voltage division factor of the feedback path K and the gain G of the original amplifier. However, linearity is achieved at the expense of overall gain of the linear signal that is reduced to the factor K, constituting an inherent drawback of feedback systems [Ken00]. Moreover, high G is typically required to allow using high K. High gain is fairly difficult to achieve at RF frequencies and justifies the wide application of feedback systems for audio amplifiers, where gain is easily achieved at low frequencies. Instead, RF amplification is much more demanding, especially due to the higher frequencies, higher bandwidth, and more tight linearity requirements.

There are several variants of feedback linearisation techniques, such as passive feedback, envelope (modulation) feedback, polar feedback, and cartesian feedback [Ken00]. Passive feedback denotes feedback around the transistor circuit of the amplifier and is typically applied for gain levelling, as it improves only the linear distortion. Thus, it does not improve any IMD or HD. There are also approaches feeding back only new generated frequency components (IMD and HD) that in turn does not improve linear distortion. Envelope or modulation feedback a predistorted version of the BB input before actual up-conversion. The polar feedback adds a phase correction loop and is able to compensate for amplitude and phase distortion. The cartesian loop, being the most common feedback technique, is similar to the polar feedback, but the BB signal is now processed in I and Q. As the cancellation or error signal $e(t)$ is generated at BB, it requires up- and down-conversion.

Feedforward linearisation, as depicted in Figure 3.2, is less addressed in literature, most likely due to its high complexity and tight amplitude and phase matching requirements for

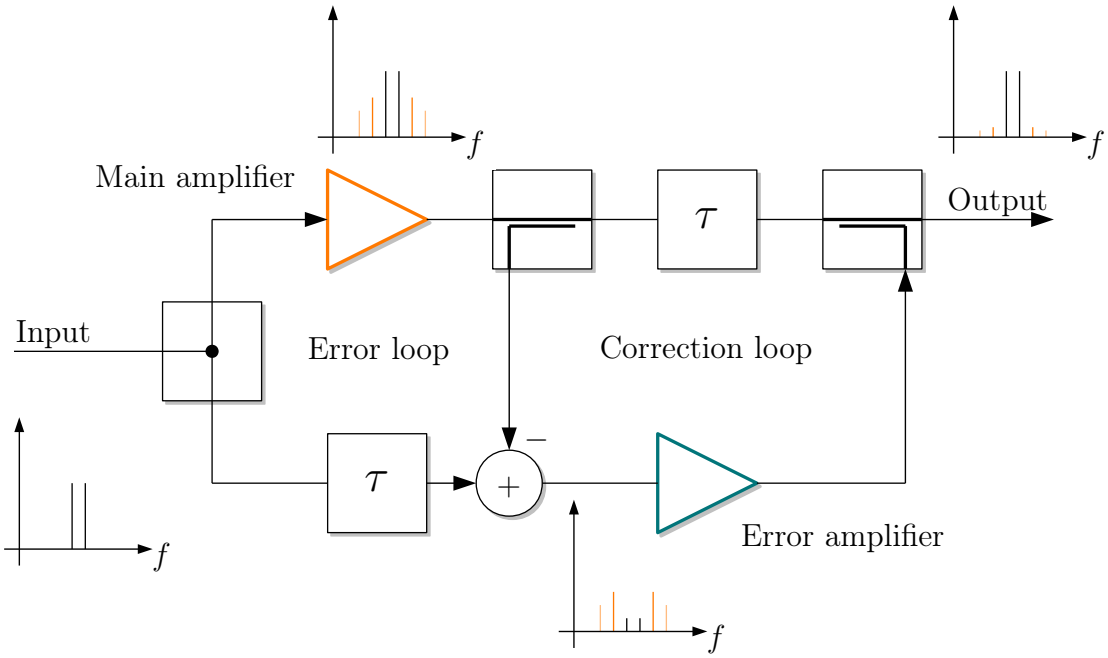


Figure 3.2: Configuration of a basic feedforward PA, sketching the signal flow by two-tone input [Ken00].

all its components. However, it is discussed here as its digital counterpart is implemented in this thesis (see Chapter 4). The input signal is split into two paths, whereas the top path constitutes the main amplifier that generates HD and IMD. Next, a part of the non-linear output is taken by a directional coupler and subtracted from a time-delayed version of the input. The signal flow is sketched with two-tone signals in Figure 3.2, but only IMD of up to 5^{th} order is illustrated here due to reasons of simplicity. As a result, an error signal remains that contains only the distortion in the ideal case. Finally, the error signal is amplified by the error amplifier and fed to the output coupler to cancel the distortion in the main path. Both main and error amplifier introduce some delay that need to be compensated by delay elements, denoted by τ, in the mutual paths. Beside exact amplitude and phase matching at RF frequency and full signal bandwidth, the error amplifier needs to be distortion-free. Otherwise, new distortion by the error amplifier is generated at the output. These issues have a strong impact on the overall mitigation performance. Another drawback of the feedforward system is its inability to measure its own performance due to the lack of intrinsic feedback.

Predistortion techniques constitute the simplest form of linearisation of an RF PA [Ken00]. Here, the key idea is to introduce a purposeful function prior the PA that is complementary to the PA non-linearity, so called non-linear inverse [FB11]. When cascading both transfer functions, the output will have little or no distortion and approaches the linear

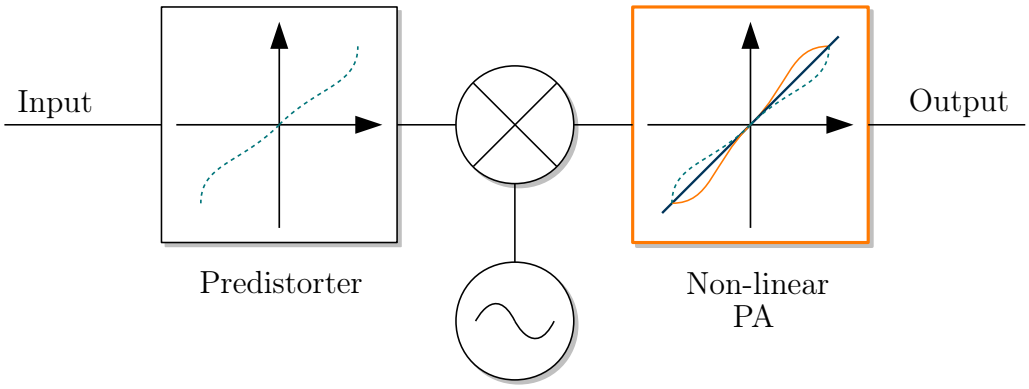


Figure 3.3: Principle of BB predistortion, indicating that expansive and compressive curve result in a linear characteristic.

characteristic. Variants of predistortion (RF, IF, or BB predistortion) are classified according to the placement of the predistortion block in the transmitter chain. BB predistortion or digital predistortion (DPD) is the most common approach and is illustrated in Figure 3.3 [FB11; BBC88; BCG+10; ZWC10]. On the contrary to the RF or IF predistortion, it relaxes the burden on the non-linear inverse block, because the BB signal is manipulated before up-conversion. Furthermore, feedback can be adopted to provide updating of this block. In general, predistortion has the ability to linearise the entire bandwidth, however, finding a circuit with expansive (third-order inverse) characteristic may be difficult. But, shifting the model of a non-linear inverse to BB is much more beneficial and modern radio transceivers anyway employ some form of DSP in their BB processing. Following this approach also enables adaptive predistortion to handle updating and accuracy problems that are in common for all the linearisation techniques. Successful DPD applications have been reported, e.g. in [BCG+10; ZWC10]. In [BCG+10], a Hammerstein model with static polynomial coefficients has been developed, whose coefficients stem from a co-simulation between RF and BB-DSP. An adaptive DPD approach has been followed in [ZWC10], that employs adaptive filtering for finding the coefficients.

A main advantage of mitigating transmitter non-linearity is that the actual input and respective wanted output is always perfectly known. This holds for all aforementioned techniques. Attempts of postdistortion in the receiver fail or perform worse due to the lack of knowledge of the original transmitted signal. Therefore, mitigating distortion in the receiver is much more challenging, as the original input is to be estimated first. That might be one reason why the receiver side is much less considered in the state of the art. In [Ken00], the possibility of postdistortion in the receiver is also briefly discussed, but worse performance has been reported due to the aforementioned reason. Also, the postdistortion model need to be trained according to some known signal configurations, such as a vacant channel that is

filled by IMD of two adjacent channels. Finally, postdistortion at the receiver can only assist prior predistortion at the transmitter and relax the performance of the mutual linearisation techniques. Another problem is that only a single non-linearity is taken into account here, namely the PA non-linearity. Modelling frequency-conversion effects and cascaded non-linear systems with higher-order distortion terms, such as contained in a typical DCR front-end line-up, seems to be impossible with these techniques. Beside adaptive BB distortion, most of the approaches are not taking into account temperature drifts, component ageing, or any change in the non-linear characteristic. Due to these reasons, an online-adaptive approach is highly desirable. Finally, most of the approaches have constraints on the signal bandwidth or modulation the feedback or feedforward path can handle.

3.1.2 Receiver Non-linearity

There is a number of analogue and digital approaches for mitigating the specific receiver non-linearity that can be classified into three main categories (see also Chapter 1). These are discussed in the following paragraphs.

Design Optimisation includes all circuit-level techniques to generally improve the linearity of the front-end. This is achieved, e.g. by utilising highly linear components in the RF front-end chain. Thus, the level of HD and IMD is reduced and the front-end can tolerate higher input power. But, there are strict limitations due to ever increasing demands for lower voltages, scaling, etc. [Sch08]. In addition, using highly linear amplifiers can be problematic for battery-powered terminals, as such amplifiers are typically very energy-inefficient [DF12]. There is always a trade-off between linearity and power consumption. For mobile terminals, low cost and low power consumption is typically desirable.

Another example for improving the linearity of receiver front-end by circuit design is the use of 3 dB hybrids and a balanced amplifier, as illustrated in Figure 3.4 [Sha]. Here, the input power is split into two paths that are 90° out of phase, amplified by two separated identical amplifiers, and finally combined again. By applying this technique, the intercept point and the P1dB are bootstrapped by 3 dB. The 3 dB hybrids can be also replaced by conventional power splitters. The advantage of 90° hybrids is that reflections by the amplifiers due to impedance mismatch are passed on to the matched input/output of the hybrids. Thus, perfect matching is achieved outside the balanced amplifier, that can be beneficial for the following receiver components. However, 90° hybrids are typically frequency-selective and not suitable for all applications.

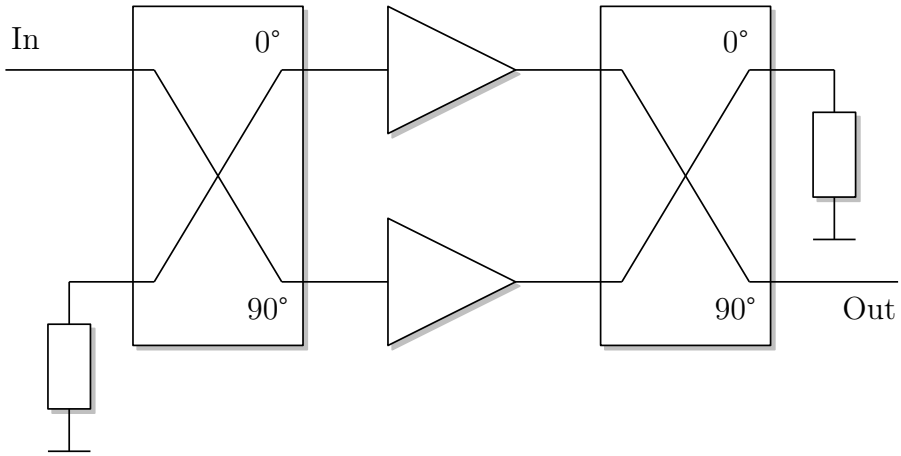


Figure 3.4: Principle of a balanced amplifier according to [Sha].

Preventing Non-linear Distortions can be also achieved by cancelling or modifying the actual signals that caused distortions. The most intuitive approach is to filter these signals, either by analogue filters, filter banks [Mar09], or tunable bandpass filters. The latter one, the tunable preselection bandpass between antenna and LNA, is very difficult to implement and not available as an off-the-shelf component [ZMS09]. Although placing the filter before the LNA will result in a higher noise figure (NF), strong blocker signals need to be suppressed before they are fed to any non-linear component in the RF chain. Otherwise, distortions are generated that may already fall into the desired band and that cannot be simply filtered out afterwards. Also, spatial filtering by an antenna array or analogue feedback cancellation with a replica is thinkable [WSW+10]. However, the most common practice is to attenuate the input so the operating point is always within a certain back-off to the saturation point (P1dB) and to be far away from the heavily non-linear region. Hence, the signals experience the linear part of the transfer function of the components [Sch08]. A significant drawback of this simple approach is that the sensitivity for weak signals is simultaneously reduced. An interesting idea to solve that problem is proposed in [OAKS+10], by cross-correlating the output of two duplicated front-ends. Adopting this approach and assuming uncorrelated noise in both sources can enable the detection of such weak signals while adopting a higher back-off. Finally, variation of the signal envelope can be reduced (PAPR reduction techniques) or by using constant-envelope modulation. For instance, constant-envelope is adopted in Global System for Mobile Communications (GSM) to be resistant to any non-linear distortion in the BTS or mobile station (MS) as far as possible.

Mitigating Non-linear Distortions can be realised by pre- or postcorrecting techniques, although postcorrection by means of DSP is on focus here. Examples of such techniques are

just mentioned here – selected approaches are discussed in the next section in more detail. The received signal can be equalized at symbol level with, e.g., embedded training or pilot symbols [DS07; Chi11; Sch08], but also blind estimation may work [DKF10; DF11; DF12]. The performance may be also improved by an iterative detection at symbol level using the Turbo equalisation principle [DSA09]. Also, digital postdistortion with a non-linear inverse, i.e. a DPD counterpart, can mitigate distortion, but seems to be less reasonable with an inaccurate model. The approach followed in this thesis depicts a feedforward architecture and regenerates distortion products with the help of a reference model and subsequently subtracts them from the received signal [VSHGA+06; KH08b; AMV10b; GSH+12b; GSH+12a; GAM+14].

Efficiency and Limitations Not all of the aforementioned techniques are reasonable and really effective in practice. First of all, any modification of the front-end circuit is typically not desirable due to reasons of cost, complexity, and flexibility. Indeed, it is also not in line with the SDR concept. Although analogue approaches have a great potential for avoiding blocking conditions, circuit non-linearity is more efficiently improved by system-level techniques [Chi11]. In addition, the technological development in the digital domain still rises according to Moore's law, hence, more and more signal processing tasks are shifted into the digital domain. However, analogue circuitry is still much faster than digital circuitry [KH08a]. Moreover, implementation of mitigation techniques by digital circuitry has the advantage of insensitivity to process variations [KH08a].

Simple filtering of out-of-band blocking signals may work as long as the distortion products are not in the band of interest. In that case, filtering should be employed directly at the RF input prior any non-linear component. This becomes difficult for multi-carrier or multi-channel receivers, where HD and IMD can hit the target band. If the distortion products have fallen onto the band of interest, more sophisticated signal processing becomes necessary rather than ordinary linear filtering [VSHGA+06]. Furthermore, tunable selective analogue filters are difficult to implement.

The proposed digital feedforward technique in its purely digital implementation also has some fundamental limits (see also Section 4.5). It requires that all distortion-producing carriers are available in the digital domain. Accordingly, the ADC needs to pass the full spectrum of potential problematic blockers within a frequency range much larger than the original signal bandwidth [KH08b; KH09b]. Hence, multiple RF chains or wideband digitisers are required. Although most of the approaches are blind mitigation techniques that do not require a training sequence or detailed knowledge of the interfering signal, they will fail under overload

conditions. Moreover, the distortion regeneration is affected by a non-ideal reference signal that is subjected to in-band distortion. Nevertheless, these techniques are most attractive as no exact knowledge of the non-linear distortion profile, the input waveforms, or the desired signal is needed.

Finally, only mixed-signal solutions will complement each other and provide the best mitigation performance. For instance, the purely digital feedforward correction can be assisted by some relaxed filtering at the input in order to avoid overload conditions.

3.2 Dirty RF Techniques State of the Art

The general idea of "Dirty RF", a term first coined in [FLP+05], is to tolerate a certain amount of RF imperfections and to mitigate them mainly by DSP at another part of the signal chain [Sch08]. Improving the system linearity by system-level DSP is a popular alternative to analogue design optimisation with the potential for power-saving and improved scalability [Chi11]. In the following paragraphs, some selected approaches from the state of the art for mitigating receiver non-linearity are briefly presented and evaluated according to digital feedforward technique, that is chosen in this thesis (see Section 3.3).

Blind Estimation of transmitted symbols over non-linear channels based on maximum likelihood (ML) method has been proposed in [DKF10; DF11; DF12]. Above those, an ML estimator can produce all possible transmit symbols, compute the distance to the actual received symbols, and finally choose the symbol with the minimal distance. The idea followed in aforementioned references is to estimate the parameters of the receiver amplifier model by ML estimation. The underlying model reproduces only AM/AM distortion and is found using the PDF after non-linearity. It is finally used in a decision feedback detector depicted in Figure 3.5, that remodulates the detected signals $\hat{s}$, feed that through the estimated non-linearity model, isolates the distortion $\hat{d}$ and subtracts it from the input y. The architecture is based on Bussgang's theorem [Sch08],

$$\hat{x} = \alpha\hat{s} + \hat{d}, \tag{3.1}$$

which states that non-linearity causes a scaling of the wanted signal and introduces an additive error term that is typically treated like additional noise. Different variants for calculating the convolution of the PDFs are discussed with regard to its computational complexity, aiming for a real-time implementation in LabView based on USRP [DF12].

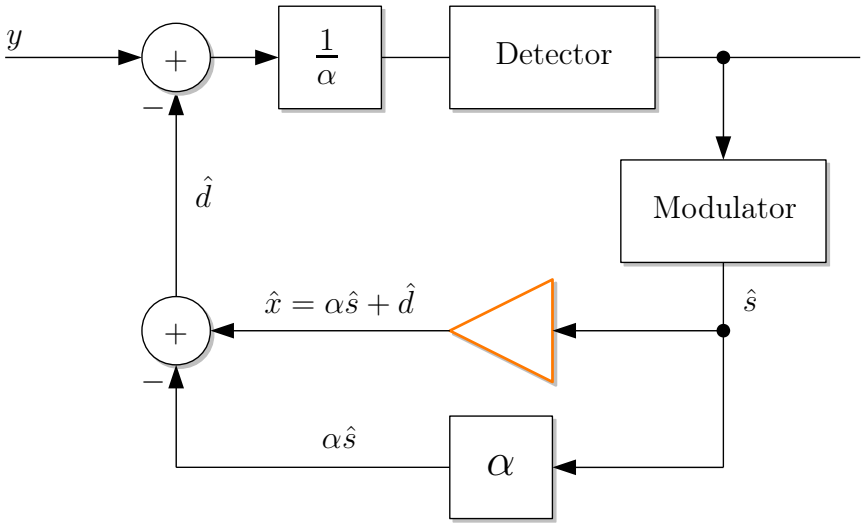


Figure 3.5: Decision feedback detector with estimated receiver non-linearity [DF11].

A good mitigation performance in terms of SNR and bit error rate (BER) is reported therein, especially in case the real-world behaviour matches to the amplifier model considered in the detector. In real RF measurements, approx. 3–6 dB improvement in the signal-to-interference ratio including noise and distortion (SINAD) and up to two orders of magnitude in terms of the BER have been obtained. It is likely that the low SINAD improvement is due to the mismatch between the underlying amplifier model and the physical RF front-end architecture. This issue is exactly one of the gaps that is treated in this thesis.

Two-channel receiver is proposed in [ZMS09] to jointly estimate the channel response and non-linearity parameters by applying a-priori knowledge of the desired and blocker signal. Figure 3.6 illustrates the proposed architecture, where the RF is split after the LNA into two receiver chains. Alternatively, the signal may be split already at the antenna, to get amplified by two individual LNAs. Thereby, cross-modulation distortion (see Section 2.3.1) of a strong blocker on a weak desired carrier is explicitly addressed in context of SDR. The custom receiver with two RF chains is employed for separated reception and down-conversion of the blocker, however, only one blocker at a time can be considered. Finally, the distortion term can be subtracted via LS and minimum mean square error estimation with the knowledge of the blocker amplitude that can be tracked by a low speed and low resolution auxiliary ADC in the second RF chain. Although this approach relaxes the burden on a single wideband ADC to capture both weak and strong signals, it requires a separate RF path for each blocker and increases the front-end complexity. The BER of an OFDM system can be reduced by up to one order of magnitude, as shown by simulations in [ZMS09].

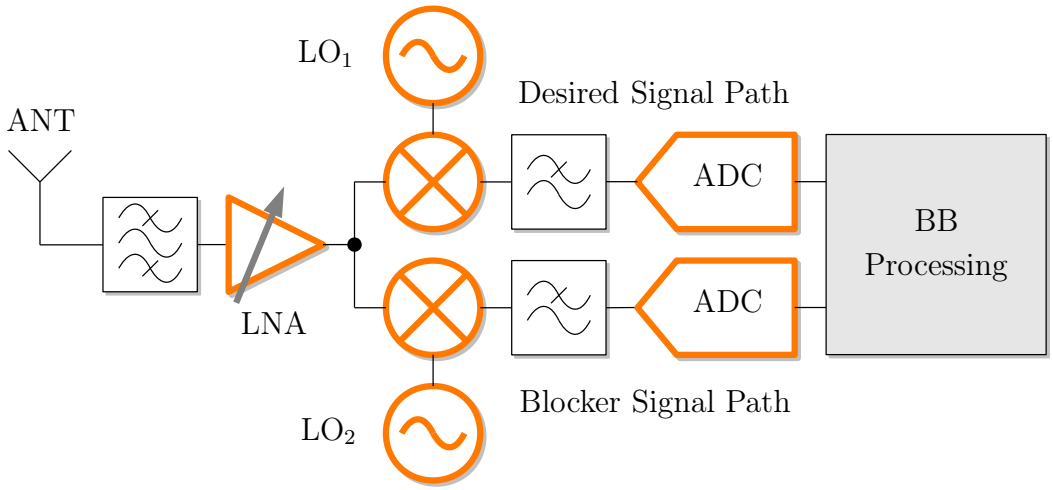


Figure 3.6: Two-channel wideband SDR receiver for acquiring wanted and blocker signal [ZMS09].

Mixed-Signal Feedforward Mitigation by using a custom-designed RF front-end is proposed in [KH08b] and related references [KH08a; KH09b; KH09a]. The proposed architecture is illustrated in Figure 3.7. Here, the main idea is to regenerate the third-order IMD products

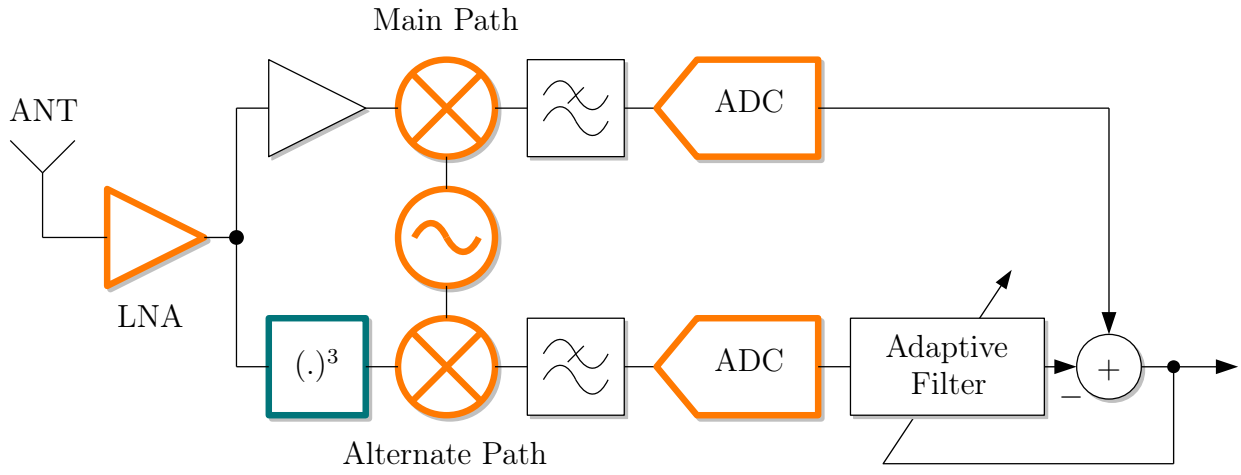


Figure 3.7: Mixed-signal feedforward mitigation with RF cubic-term generator highlighted [KH08b].

at the analogue RF domain in an alternate path and to pass only problematic IMD products down to BB in the same manner as the desired signal (using same LO). The processing and adaptive filtering takes place at digital BB. Thus, digital BB and ADC requirements are not greater than that of the main path. It generally solves the main limitations of purely digital feedforward mitigation followed in this thesis, but also requires a lot of additional circuitry to be implemented on chip. The main advantage of this approach is that all distortion products,

appearing at the full receiver RF bandwidth, are represented by the cubic-term generator in the alternate path [KH09a], and only those falling into the desired BB bandwidth are further processed in the digital domain. Following this approach, only RF distortion is considered. Further BB distortion by mixer or BB amplifiers are not considered herein, although it could be mitigated by adaptive filtering in the digital domain as well. It is also noteworthy that RF distortions are implicitly modelled by physically down-converting these distortion products to BB. In [KH08b], a custom RF front-end for UMTS application has been designed to meet the linearity specifications imposed by an out-of-band blocker test. The effective IIP3 has been improved by 12 dB, and an SNR gain of 20 dB and BERs of down to 10^{-4} for the UMTS scenario have been reported. These are very impressive results that are achieved at the expense of circuit complexity.

ADC Distortions are specifically addressed in [AMV10a; AMV10b; ALM+12; Chi11]. Here, the main goal is to mitigate both mild non-linear distortion and also clipping distortion, i.e. to remove any kind of distortion regardless of its source. Beside aforementioned distortion phenomena, ADC-specific non-idealities are DNL, INL, and clipping (see Section 2.2). In [AMV10a], an alternate path with an additional ADC is proposed to obtain a non-clipped reference signal. That auxiliary ADC can be a low cost and low resolution type as in the two-channel receiver approach proposed in [ZMS09]. Finally, clipping distortion is mitigated by recovering the clipped waveforms using interpolation. In [AMV10b], a mathematical model of zero-symmetric clipping based on Fourier analysis is developed for understanding and modelling the impact of the ADC and essential IMD distortion on top of weak signals. This model is used for post-correcting the ADC non-linearity by using only one RF path. Clipping-induced IMD is regenerated and subtracted from the received signal using adaptive filtering.

Digital Feedforward Mitigation has been first proposed in [VSHGA+06; gha11] and is detailed in Chapter 4. This algorithm, illustrated in Figure 4.1, is working purely in the digital domain. The main idea is to regenerate distortion products by specific strong distortion-producing signals and to subtract them from the desired signal observation in an adaptive manner. Since the error signal and reference model are never precisely known in practice, an adaptive filter is used to subtract the reference signal from the main receiver path. There is a gap in the state of the art with regard to the mitigation architecture. The specifics in modelling of a complete RF chain of a DCR have not been considered so far. Up to now, the non-linearity of the total RF front-end is modelled by a single polynomial. This approach does not distinguish different types of RF and BB distortion and cannot capture

all distortion products generated by such front-ends. Therefore, the model is extended in a way that it matches to the physical front-end architecture.

3.3 Reasons for Selecting Digital Feedforward Mitigation

There are several advantages why the digital feedforward mitigation has been chosen for this work, among other analogue or digital techniques. First, the algorithm is suitable for universal application and can be simply integrated, especially due to the fully DSP-based approach. Thereby, a very flexible implementation of the algorithm on an FPGA or DSP can be achieved without employing any additional hardware. It is efficiently correcting the I/Q signals in a way that the amount of non-linear distortion is reduced. As it is solely processing the distortion-producing carriers, it is independent of the desired signal and its specific modulation.

Second, non-constant characteristics, such as production tolerances in the front-end's components, thermal influences, or device ageing can be easily taken into account by the adaptive filtering stage. This is a clear advantage over all analogue linearisation techniques discussed in Section 3.1.1 in context of transmitter non-linearity mitigation.

Last, it is noteworthy that no detailed knowledge of the non-linearity profile and no exact polynomial coefficients are necessary [VSHGA+06], only basic knowledge about the physical receiver front-end architecture is required.

Due to the aforementioned advantages, the approach has been followed in this thesis for alleviating non-linear distortions in the receiver. Beside general limits and requirements of this technique, it perfectly fits to the presented research problem and represents an effective solution for mitigating receiver non-linearity at system level.

3.4 Cancellation of Distortion-Producing Signals

In the following, blocker or interference cancellation techniques are discussed as an alternative to alleviate non-linear distortion at the receiver. Although leaving the scope of "Dirty RF" digital post-correction, such techniques constitute a promising solution to prevent the generation of non-linear distortion at all.

3.4.1 Analogue Cancellation

An active feedback interference cancellation technique has been proposed in [WSW+10] that bases on an analogue cancellation filter core, as illustrated in Figure 3.8. Here, the

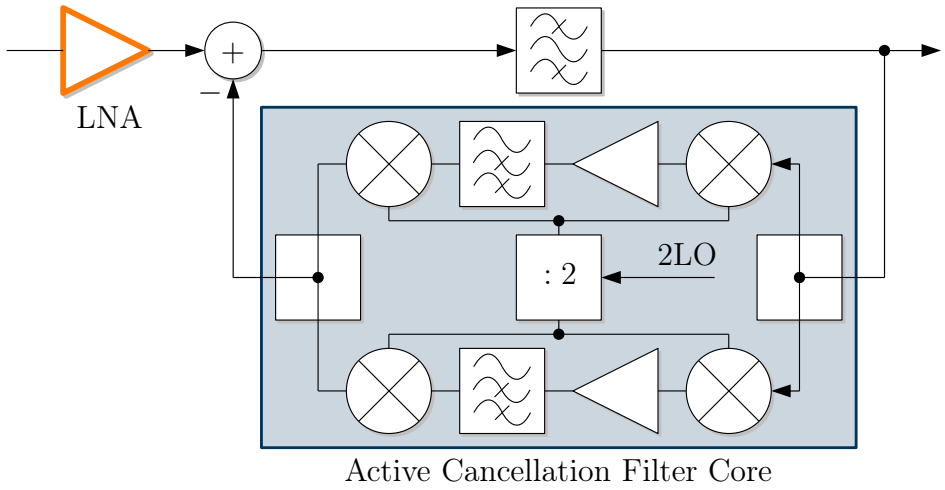


Figure 3.8: Block diagram of the active feedback interference cancellation proposed in [WSW+10].

received signal including the wanted and blocker signal are down-converted by the receiver LO signal. Then the wanted signal is eliminated by highpass filtering, thus, the remaining signal constitutes a blocker replica. Finally, the blocker replica is again up-converted to RF and subtracted from the received RF signal, resulting in a partial cancellation of the blocker.

This intuitive approach has some major drawbacks. First, the highpass corner frequencies are determined by the bandwidth of the actual wanted signal at DC, assuming a zero-IF down-conversion. The remainder of the signal after highpass filtering is considered as interference. It may comprise multiple strong blockers, but also other useful signals, that are finally fed to the subtraction node. That is, the filter core is fix in its cut-off frequency and not selective to specific blocker signals. Second, as the blocker subtraction is conducted after the non-linear LNA, distortion can be already on top of the wanted carrier and cannot be simply filtered out by this approach. Ideally, the subtraction should be performed prior the LNA, even though the overall NF is affected. Filter core noise may be added in the main path also in the proposed configuration, especially due to the LO phase noise. Third, there are also some implementation challenges reported in [WSW+10]. For instance, the filter core IF bandwidth determines the loop stability. The filter implementation is also challenging when the blocker comes close to the wanted signal. In addition, I/Q imbalance in the up-

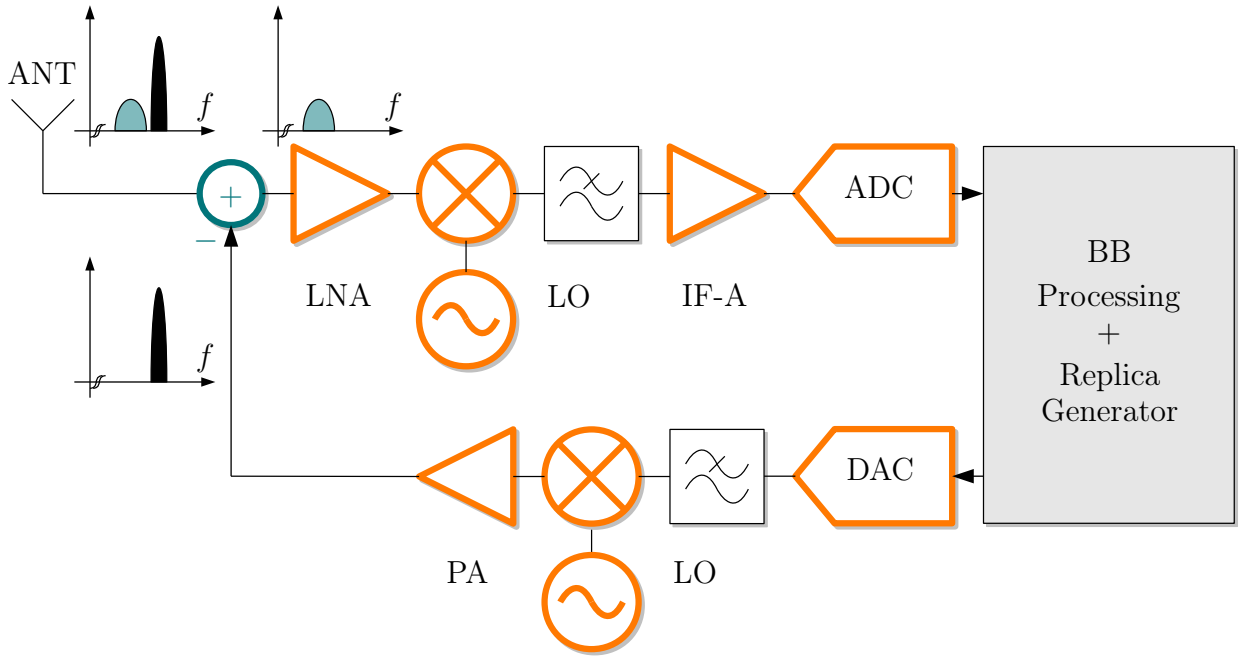


Figure 3.9: Block diagram of the analogue blocker interference cancellation.

and down-converter paths may cause blocker images in the main LNA path. Although the problem of amplifier non-linearity is not mentioned in [WSW+10] and references therein, the proposed cancellation technique is an interesting solution to avoid non-linear distortion at the receiver by simply cancelling crucial signals. However, the solution is not very flexible and rather unsuitable for cleaning the whole BB signal from distortions, as it is needed for application in CR. It can rather provide a performance improvement for a single wanted signal of interest.

A similar approach has been followed at the beginning of this work, but has not been pursued further due to major drawbacks like those mentioned in the aforementioned article. The proposed architecture is depicted in Figure 3.9. The key idea is to cancel the blocker signal at RF before entering the LNA or any other non-linear component in the receiver front-end. Thereby, non-linear distortion provoked by a specific blocker signal is prevented. As the blocker replica is generated at digital BB, it is much more flexible than the previous approach. First, the blocker needs to be located by using a coarse sensing algorithm such as energy detection (see Section 7.1). Next, a replica of the blocker is generated using the complete transmitter chain of the SDR architecture. Finally, the blocker is subtracted from the input at RF level.

This approach has several advantages when assuming ideal operating conditions. It is likely that the blocker can be successfully cancelled if the replica has a perfect match in amplitude

and phase. It requires no demodulation of the blocker, although a remodulation based on the detected symbols might be a possible solution. A major advantage is that no model of the non-linear receiver is required at all, hence, the approach is fully independent from the specific RF front-end design. Even if the blocker is only partially cancelled due to imperfect match of the replica, the linearity requirements of the front-end are reduced significantly. Except than the subtraction node, that can be a differential LNA in practice, no hardware modification is necessary, as the transmitter stage is typically available in an SDR.

However, there are some practical challenges and drawbacks that make this kind of blocker cancellation not feasible in practice. There is an unavoidable delay of the replica generation and the feedback loop as the complete transmitter is included in the feedback loop. That is, it will only work for deterministic or at least slow-changing blocker signals that can be predicted. In theory, the RF signal should be delayed to provide sufficient time for generating the replica. In order to establish an exact amplitude and phase match, a closed loop is required. Also, in case of blocker modulation and channel fading, amplitude and phase of the blocker will change over time. Beside these conceptual challenges, there are also some implementation issues that need to be looked at closely. For instance, significant noise may be introduced prior the LNA due to the feedback path. Also, the PA at the transmitter may add new undesired components. However, the blocker replica needs to be only amplified to relatively small power levels at the order of approx. $-80\,\mathrm{dBm}$. Thus, leakage effects through the receiver antenna should be significantly small.

Proof-of-Principle Experiment

Nevertheless, some real RF measurements have been carried out in order to pursue the idea and to discover the challenges of this technique in more detail. The setup is depicted in Figure 3.10. For reasons of simplicity, only one blocker has been assumed that is static over time. The actual blocker and its replica are generated by two individual USRP SDR platforms, equipped with WBX daughterboard [Ett]. Random static phase offsets at USRP generating the blocker replica have been adapted by a delay line. These offsets are caused due to the FPGA-based processing on the USRP. A desired signal is generated by an AWG and is coupled into the main receiver path. In this setup, the blocker's amplitude and phase are assumed to be perfectly known and both transmitting USRPs are fully synchronised by a 10 MHz reference clock and a pulse per second (PPS) signal. Finally, desired signal, blocker, and replica are fed to a third USRP that acts as the actual receiver. In the final application, the receiving USRP should be able to create the blocker replica by itself.

Figure 3.11 illustrates the measured BB spectrum before and after cancellation. Desired

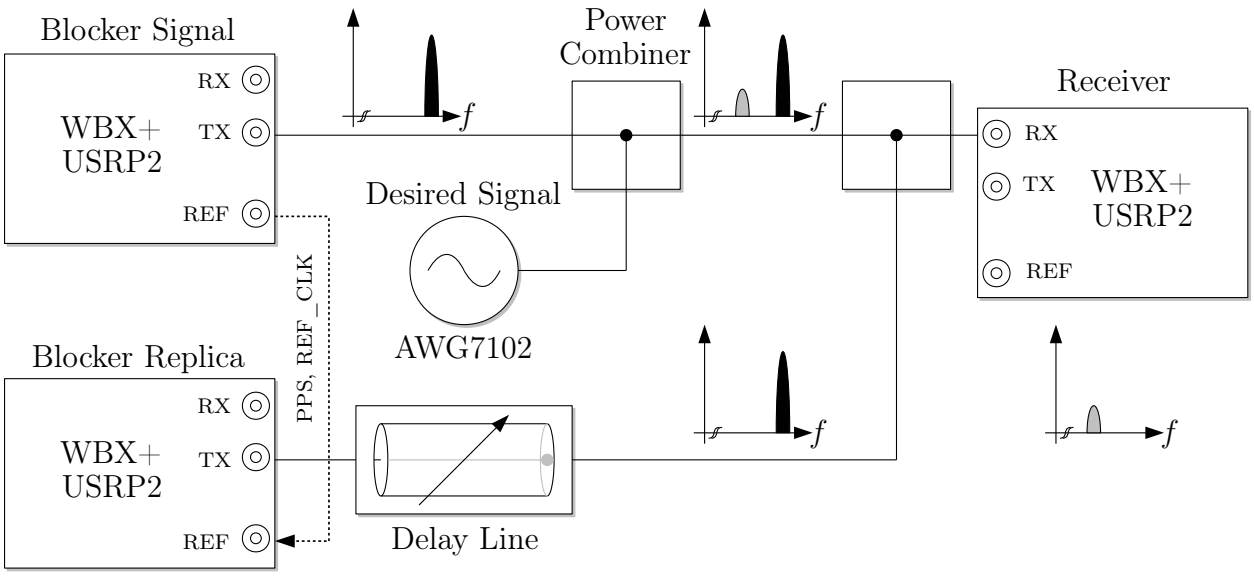


Figure 3.10: Experimental setup for the analogue blocker cancellation.

and blocker signal are CW 5-tone sinusoidal signals with different tone spacing and carrier frequencies. Up to 40 dB blocker suppression have been achieved, depending on the resolution of the delay line. It is also visible that the IMD distortion caused by the 5-tone blocker almost vanished. As the LO leakage is different in the blocker and replica transmitter, it could not be compensated. To sum up, the general concept of blocker cancellation has been proven to be functional, however, many of the aforementioned drawbacks make it not practical. For these reasons, digital feedforward mitigation has been improved further instead of pursuing the analogue blocker cancellation.

3.4.2 Spatial Filtering using Antenna Arrays

Another interesting approach to increase the dynamic range and to avoid blocking conditions is to adopt the filtering of strong blockers in the spatial domain by an antenna array instead of the frequency domain [CB05]. A SDR with a large number of receiver antennas is illustrated in Figure 3.12 [Sch08]. Basically, desired and blocker signals may impinge from different directions. Therefore, the main idea is to achieve a selective reception and suppression of specific impinging waves with antenna arrays and beam-forming techniques, that combine different antenna signals before digitisation. Here, closely placed antenna elements are fed in a way that a desired radiation pattern is generated. This is accomplished by controlling the power levels and phases of each path, indicated by antenna array coefficients a_i in Figure 3.12. These coefficients are computed in the digital domain. Finally, nulls in the antenna diagram

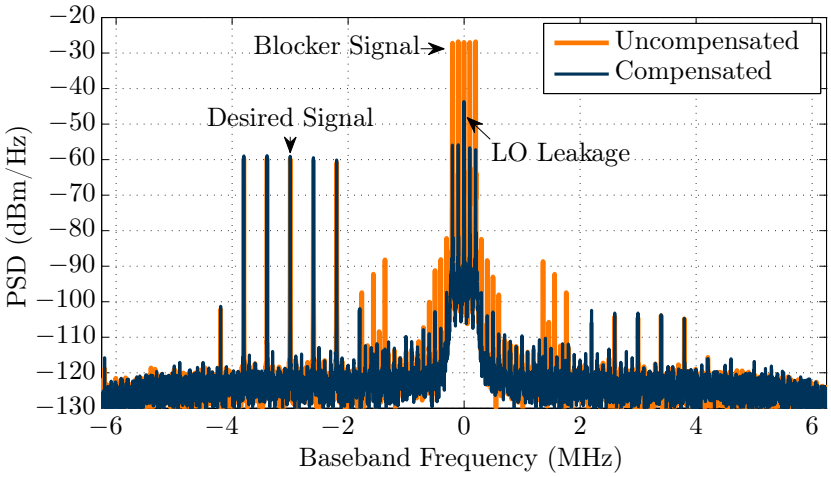


Figure 3.11: BB spectrum of the received signal before and after blocker cancellation.

in the direction of strong blockers can be generated after prior measurement of their angle of arrival by sweeping the coefficients. Simultaneously, weak signals can be amplified by imposing some higher gain in their angle of arrival.

Beside beamforming techniques, also a simple opportunistic choice of the best antenna is thinkable. That is, all possible antenna diagrams are scanned and the optimal diagram with the most suitable signal configuration of desired and blocker signal is just switched. In any case, spatial filtering will support other digital mitigation algorithms by avoiding overload conditions and relaxing the burden on the A/D interface.

Such interferences scenarios can be emulated using the setup for over-the-air testing in a virtual electromagnetic environment (OTA*in*VEE) build at Ilmenau University of Technology [KGM+11; MKG+11; GKM+11a]. This testbed allows for emulating a realistic radio environment with reproducible signal scenarios for a single centred radio node, surrounded by an antenna array in a circular arrangement. For details on this test approach, the interested reader is referred to the aforementioned references.

Though, spatial filtering is likely to be effective, the approach is out of the scope of this thesis and has not been pursued further.

3.5 Summary of Possible Mitigation Techniques

This chapter has presented a broad range of analogue and digital techniques to alleviate circuit non-linearity. Receiver non-linearity can be handled by design optimisation, cancelling distortion-producing signals, or by pre- and post-correction techniques. The latter

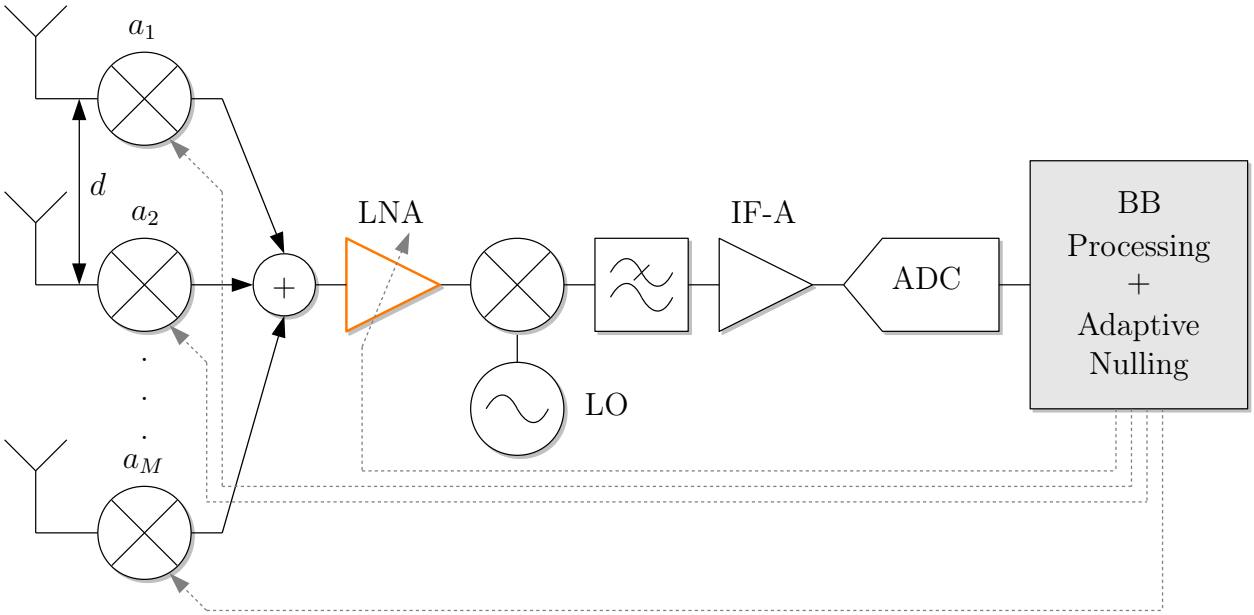


Figure 3.12: Wideband SDR front-end with antenna array for spatial filtering [CB05].

one includes "Dirty RF" signal processing, that represents a trendy system-level technique to alleviate non-linear distortions. Beside receiver non-linearity, techniques for mitigating *transmitter* non-linearity have been also discussed, as some principles are in common. Finally, a digital feedforward mitigation algorithm has been selected, especially due to its simple integration and efficient post-correction of distorted BB signals in the receiver.

4 Digital Feedforward Mitigation

This chapter discusses the selection mitigation technique in detail. First, the main architecture and its individual components are explained. The actual mitigation architectures are developed based on the appropriate BB models discussed in Chapter 2. Then, the effectiveness of the algorithm is demonstrated by extensive simulations and RF measurements. Finally, fundamental limits are discussed that need to be considered for employing this technique into real-world devices.

4.1 Principle

The key idea of the feedforward mitigation algorithm is taken from [VSHGA+06]. Figure 4.1 illustrates the algorithm in its general form. The signal flow is hereby illustrated by sketched

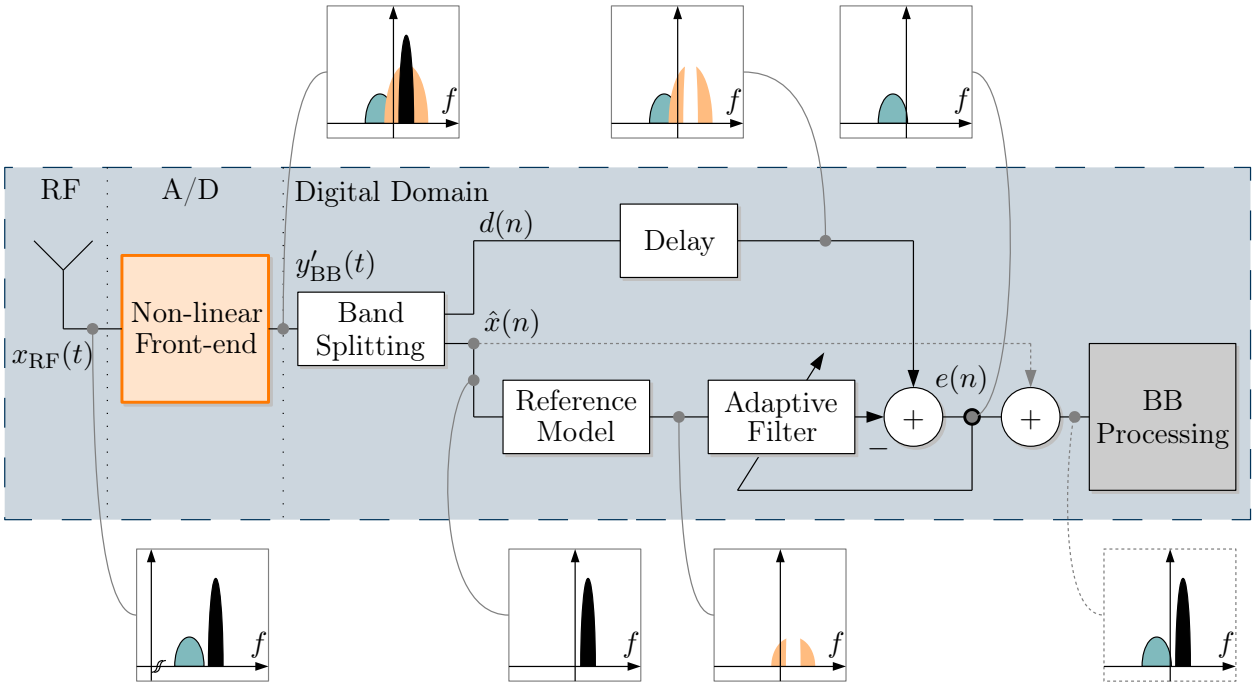


Figure 4.1: Block scheme of the generalized digital feedforward mitigation algorithm.

BB equivalent spectra at different points along the signal processing chain.

The RF antenna signal $x_{\mathrm{RF}}(t)$ passes the non-linear front-end where the signal gets amplified, down-converted, and digitised. This analogue processing introduces noise and distortions in the signal, resulting in the distorted version $y'_{\mathrm{BB}}(t)$. In most of the applications, the input signal consists of a weak desired signal (green-coloured signal in the sketches) and a close strong blocker (black-coloured). At BB, additional distortions around the blocker (orange-coloured area) show up and can fall on top of desired signals or onto free frequency ranges.

Now, the main idea is to mitigate these distortions by introducing some signal post-processing in the digital domain that is placed in between the A/D-conversion and the system-specific BB processing for the desired signal or spectrum sensing. That is, the algorithm is working at waveform level with the I/Q signal. It consists of the following components:

- Band splitting including interference detection,
- Reference non-linearity model, and
- Adaptive filter.

The key idea is to extract crucial interferers contained in the BB and to reproduce their distortion products. First, a filter pair with inverse characteristics, e.g. bandpass/bandstop or lowpass/highpass configurations, splits the distorted signal $y'_{\mathrm{BB}}(t)$ into a desired signal $d(n)$ and the reference signal $\hat{x}(n)$. The upper path signal, $d(n)$, consists of the signal of interest and distortions products caused by the blocker, while the reference path signal $\hat{x}(n)$ constitutes an estimate of the originally transmitted blocker. Second, the reference signal in the lower branch is fed into a reference model of the non-linear front-end reproducing RF and BB non-linearity as well as I/Q imbalance, to regenerate the distortion (HD, IMD) that was generated during non-linear reception. The reference model output at this stage does not match the amplitude and phase of the distortions in the upper branch (illustrated by lower amplitude of the orange-coloured area) due to model inaccuracy or incorrect estimate of the interferer signal. Therefore, the reference distortions are further processed by an AF to adjust them to exact values. Thus, by subtracting the AF output from $d(n)$, the effect of non-linearly induced interference is diminished. The cleansed signal is the common error signal $e(n)$ of the AF.

In general, two different use cases need to be distinguished. These are referred to as

1. Single signal of interest (SOI), and
2. Sensing (or multi-carrier BTS).

Examples for these use cases are given in Chapter 7. In the first case, a certain narrow band is to be cleaned from distortions that have been generated by strong neighbouring blockers.

This small band typically holds a weak carrier and does not occupy the whole BB bandwidth. Then, in the upper branch, this single SOI is isolated using a bandpass. In contrast, in the lower branch, the desired signal is filtered out and the remainder of the BB signal is used for modelling the distortion falling into the desired band. In the last sketched spectrum, only the desired signal remains (upper right spectrum in Figure 4.1).

On the other hand, the whole BB bandwidth is to be cleaned from distortions in the second use case, e.g. to enhance the reliability of spectrum sensing techniques or for wideband BTS receivers that simultaneously demodulate all down-converted carriers. A bandstop filter in the upper branch filters out the strong blocker to provide a good desired signal for the AF. Conversely, a bandpass filter in the lower branch isolates the strong blocker that is used for distortion modelling. For applications like spectrum sensing, it is desirable to have the strong blockers present after mitigation, because they can be also interpreted as useful signals. These scenarios aim for an undistorted BB representation of $x_{\mathrm{RF}}(t)$ at the antenna input. In such a case, the blocker estimate $\hat{x}(n)$ needs to be added back after mitigation to the error signal $e(n)$, as it is indicated in Figure 4.1 by a dashed arrow. Finally, a cleaned version of $y'_{\mathrm{BB}}(t)$ is obtained, i.e. without the orange-coloured distortion components.

XMD between strong blockers and weak desired signals, as discussed in Section 2.3.1, are not taken into account by this architecture. However, this specific kind of distortion is typically weak compared to the IMD caused by dominating blockers [VSHGA+06]. Furthermore, it is sufficient to isolate only the strongest blockers as they cause the most harmful distortion. The amount of generated distortion always depends on the peak signal power present at the input. By selecting specific distortion-producing signals, the distortion estimates will be also more accurate, as no additional distortions are generated that are actually not present in the received signal. Feeding a large band of signals through the reference path likely causes unwanted distortions, depending on the actual power levels and frequency separations between the carriers.

The mitigation processing in the lower branch, especially the filtering, introduces some delay. To compensate for that effect, the same delay can be easily added in the upper branch. This is one of the major advantages of performing this processing digitally compared to the main drawback of the analogue blocker cancellation techniques discussed in Section 3.4.

For the simulation testbed, the three main components have been implemented in MATLAB using an object-oriented design. This simulation framework is referred to as non-linearly induced interference mitigation (NONLIM) algorithm and is detailed in Section C.1. In the subsequent sections, all components are discussed from the system-theoretic point of view in detail.

4.2 Components

4.2.1 Bandsplit Filtering

The derivation of the desired and reference signal plays an important role for the effectiveness of the adaptive mitigation processing, as their stopband attenuation dictate the spectral purity of the input for the AF. That is, the filters should have a high slope to clearly separate the blocker from the IMD and the remainder of the BB signal. The exact filter characteristics particularly depend on the RF signal configuration and the underlying receiver front-end at hand. If the scenario is known in advance, fixed filter settings can be used. Otherwise, the filters need to be tuned accordingly by changing their coefficients during runtime. The first case is valid for applications with one SOI, while the latter one should be considered for highly dynamic scenarios like for spectrum sensing applications in CR.

For the spectrum sensing application, strong blockers near to desired signals or bands for secondary transmissions can be identified in the first step by a coarse energy detector (Section 7.1). A threshold of 20 dB below the maximum input signal level has been found through experiments and is assumed to be reasonable to detect all strong blockers. Thereby, the filter pair may consider a single blocker at a time or even multiple blockers by a multi-band filter.

By filtering the blocker out of the input signal, an estimate $\hat{x}(n)$ of the originally transmitted signal $x(t)$ causing the distortions is obtained. However, the strong signal (or blocker) itself suffers from non-linear distortion, so called in-band distortion. Therefore, the reference distortions provided by the following reference model will differ from those in the upper branch. This problem is discussed in detail in section 4.5.

The computational complexity of the filtering should not be underestimated. The filter pair should have linear phase (or equivalently constant group delay), high stopband attenuation, and unity amplitude response in the passband in order to not distort the signal further. These characteristics are required for phase-sensitive communications and are fulfilled by FIRs with symmetric coefficient sequences. Because FIR filters do not require any feedback, they are inherently stable compared to their counterparts, the infinite impulse response filters.

Following Figure 4.1, a pair of a complex bandpass and bandstop filter is typically required, especially due to the mirror-frequency interference occurring in any practical receiver. The classical way of designing such filters is to frequency-shift a real-valued lowpass and highpass FIR prototype, respectively. The following section will give a brief overview on FIRs. More details can be found in literature, e.g. in [MB04; Hay02; Spt].

FIR Filters are the most common LTI filters that act on the input through linear convolution [MB04]. Hence, the output is given by

$$y(n) = \underline{w}^H(n)\underline{u}(n) = \sum_{k=0}^{M} w_k^* u(n-k), \tag{4.1}$$

where M is referred to as filter order with the filter coefficients w_k and corresponding input $u(n)$. The FIR filter architecture is depicted in Figure 4.2 in a transposed transversal form [MB04]. This structure is also known as "tapped delay line" and consists only of the three

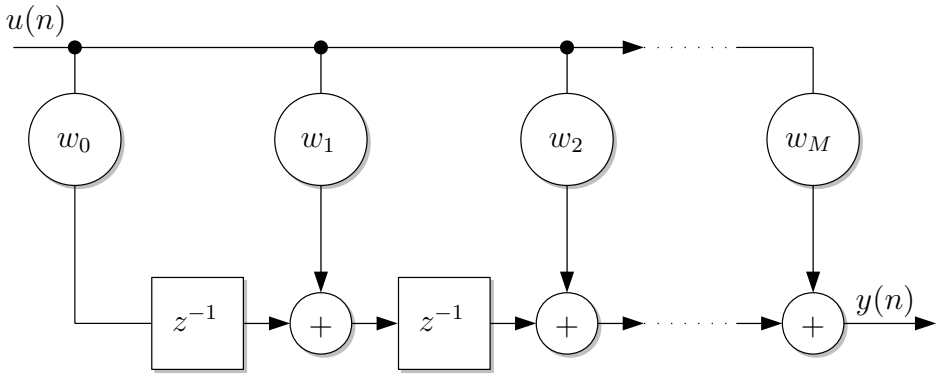


Figure 4.2: FIR filter in transposed structure [MB04].

elements delays, adders, and multipliers. The task of the multiplier is to multiply each tap input $u(n)$ by a filter coefficient w_k that is also referred to as tap weight [Hay02]. The transposed form is preferred here as no shift register is necessary for holding the input samples [MB04].

A main advantage of FIR filters is their inherent stability as they do not employ any feedback. Hence, they only show zeros in the pole-zero plot. The number of the delays represents the finite duration of its impulse response. Linear-phase FIRs have a constant group delay

$$\tau(\omega) = \frac{\mathrm{d}\phi}{\mathrm{d}\omega} = \text{const.} \tag{4.2}$$

over a range of frequencies, if the filter is even- or odd-symmetric. Symmetric FIRs also require only half the number of multipliers, constituting an important advantage for efficient FPGA implementation.

FIRs are typically designed using direct computer specifications and algorithms, such as the MATLAB Signal Processing Toolbox [Spt]. For this purpose, different algorithms are available, such as equiripple, Kaiser window design, or LS FIR method. Throughout the thesis, the equiripple method following the Parks-McClellan optimal FIR filter design algorithm

[Spt] based on the Chebyshev approximation theory is adopted. This and other methods require the desired filter characteristics as input, i.e. the transfer function in the frequency domain represented by passband, transition band, and stopband characteristics. Then, a mathematical algorithm optimises the filter coefficients according to the desired specifications. There is also the possibility to design and perform the filtering in the frequency domain, by using the FFT [MRS91]. However, this approach just provides less computational complexity for high filter orders M or is preferred if subsequent signal processing is anyhow conducted in the frequency domain.

4.2.2 Reference Non-linearity Models

The mitigation algorithm requires a signal model of the front-end's imperfections to regenerate the distortions caused by strong signals (blockers) in the BB. The model output basically tells where distortions are located in the BB by distorting the input in the (ideally) same way like the actual physical front-end. The reference model has to reproduce the behaviour of the original front-end as accurate as possible (behavioural modelling). That is, the structure of the reference model should match the physical origins of the distortion in the front-end. Depending on the receiver architecture, there are different types of models or subsets of the full cascaded model can be used. Three types of reference models are distinguished, single RF non-linearity, single BB non-linearity, and combined/cascaded non-linearity. The total cascaded model is adopted first to derive a complete mitigation architecture. Then, it can be decided from the use case and the underlying receiver which parts of the mitigation architecture are finally employed for the algorithm. This novel kind of modelling has been published in [GAM+14].

Total Cascaded Non-linearity Model

The theoretical architecture for joint mitigation of RF and BB non-linearity as well as mixer and BB I/Q imbalance is depicted in Figure 4.3. The complex bandpass and bandstop filters are denoted by $h^{\mathrm{C}}_{\mathrm{BP}}$ and $h^{\mathrm{C}}_{\mathrm{BS}}$, respectively. New symbolic blocks have been used for the digital bandpass and bandstop filters to differentiate them from their analogue counterparts used in the receiver block diagrams. The most intuitive way to design the reference model is to simply cascade RF, mixer, and BB model that have been developed in Section 2.5. Hence, the total model decomposes into several blocks. However, this structure will only work if the exact coefficients are known a-priori. In fact, an online adaptation of all the model coefficients is practically required, hence, AFs are included and depicted by gray-coloured arrows and the

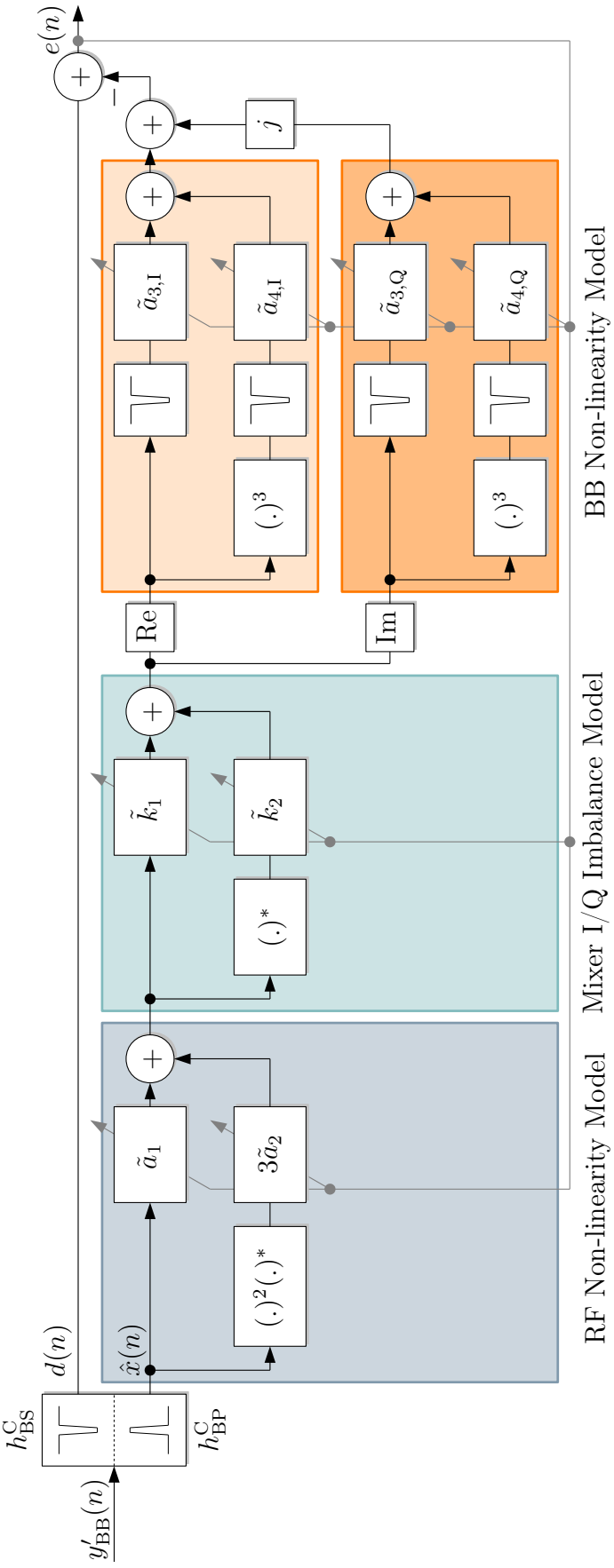


Figure 4.3: Theoretical architecture for mitigating cascaded non-linearity [GAM+14].

tilde coefficients in Figure 4.3, indicating estimates of the exact coefficients. The challenging issue is that RF and BB distortion generate same spectral content, especially the IMD in the fundamental zone, and thereby cause severe difficulties for the convergence behaviour of the individual AFs. By using the sequential structure, the AF adapting the coefficients of the RF non-linearity model will try to eliminate the IMD in the fundamental zone at once. If this distortion is already vanished, adapting the BB non-linearity coefficients becomes difficult, as some of the BB distortion components are not any more in the signal. The same happens if the BB distortion is removed before the RF distortion. In other words, there is a certain time-domain correlation between the distortion components that originate from the same blocker signal. Difficulties of the adaptation process with correlated input has also been discussed in [Hay02]. These convergence problems are relaxed if the AF could be assisted by telling the basic relations between the distortion terms. For instance, if the RF non-linearity coefficient a_3 is exactly known a-priori, the BB distortion term can be easily adapted. Since a_3 is raised up to the third power while passing the BB non-linearity, the accuracy of a_3 must be high to avoid severe deviation. In order to solve these problems in practice, the coefficients should be updated simultaneously and not sequentially [GAM+14].

Next, the terms produced by the cascaded model developed in Section 2.5.5 and Section 2.5.6 are analysed and then a simplified parallel mitigation architecture is derived that enables simultaneous adaptation of the distortion components. The I/Q imbalance is neglected for extracting the distortion terms, but is considered again in the final architecture. The distortion terms produced by the total cascaded model can be extracted from the complex representation (2.45) and are summarised in Table 4.1. The terms are listed with regards

Table 4.1: Distortion terms produced by the cascaded non-linearity model without I/Q imbalance.

No.	**Coefficient**	**Term**	**I/Q Representation**
1	$a_3 a_1 + \frac{27}{2} a_4 {a_2}^2 a_1 \lvert z \rvert^2$	x	$x_\mathrm{I} + j x_\mathrm{Q}$
2	$3 a_3 a_2 + \frac{3}{4} a_4 {a_1}^3 + \frac{18}{4} a_4 a_2 {a_1}^2 \lvert x \rvert^2$	z	$x_\mathrm{I}^3 + j x_\mathrm{Q}^3$
3	$\frac{1}{4} a_4 {a_1}^3$	$(x^*)^3$	
4	$\frac{9}{4} a_4 a_2 {a_1}^2$	$x^2 z^*$	$x_\mathrm{I}^2 z_\mathrm{I} + j x_\mathrm{Q}^2 z_\mathrm{Q}$
5		$(x^*)^2 z^*$	
6	$\frac{27}{4} a_4 {a_2}^2 a_1$	$x^* z^2$	$x_\mathrm{I} z_\mathrm{I}^2 + j x_\mathrm{Q} z_\mathrm{Q}^2$
7		$x^* (z^*)^2$	
8	$\frac{81}{4} a_4 {a_2}^3$	$z^2 z^*$	$z_\mathrm{I}^3 + j z_\mathrm{Q}^3$
9	$\frac{27}{4} a_4 {a_2}^3$	$(z^*)^3$	

to their power contribution in descending order, while the time variable t is omitted to enhance readability. The last column in Table 4.1 provides the I/Q branch counterparts to the complex terms, indicating that each term in I/Q form corresponds to two complex terms. Doing the same analysis with I/Q imbalance included results in another nine terms that are the complex conjugates of the terms listed in Table 4.1. In general, all distortion terms can be considered if all of them have separate *parallel* branches. However, not all terms have a significant contribution in practice, hence, the architecture can be simplified further.

Only the most significant distortion terms plus their complex conjugates should be considered in order to reduce the computational complexity and to overcome difficulties in the adaptation process for weak terms having contribution below the noise power level. In addition, the terms should be carefully selected so they do not have much spectral overlap. The final terms that have been selected for the parallel mitigation architecture are listed in Table 4.2. Hats ($\hat{}$) and the discrete variable n are used in the term notation to indicate that an

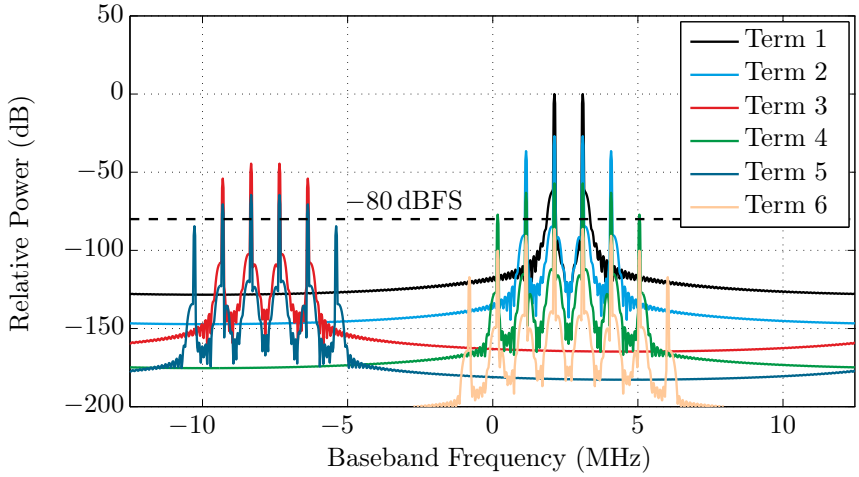
Table 4.2: Selected terms for parallel mitigation architecture [GAM+14].

No.	**Term**	**Filtering**
1	$\hat{x}^*(n)$	SL
2	$\hat{z}(n)$	WL
3	$\hat{x}^3(n)$	WL-RC

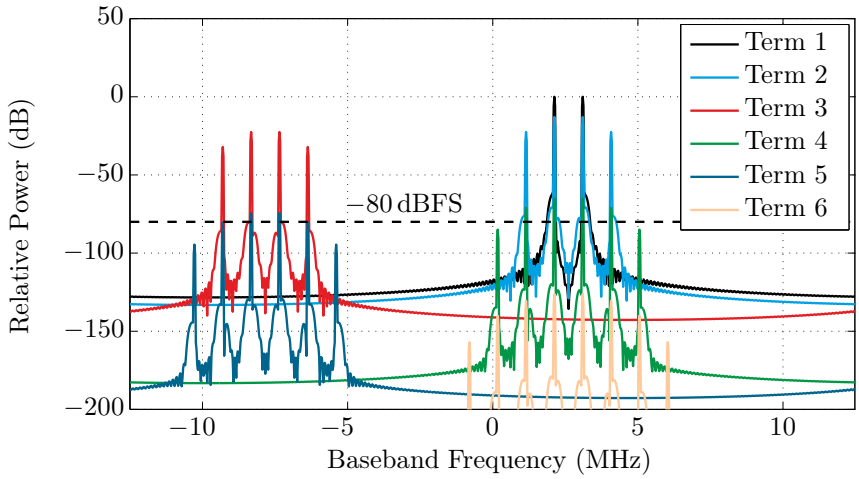
estimate of the distortion-producing blocker signal $x(t)$ has been obtained by filtering in the digital domain. In fact, the blocker estimate $\hat{x}(n)$ also contains in-band distortion, a problem that is further detailed in Section 4.5. The third column indicates if the terms should have strictly-linear (SL), widely-linear (WL), or reduced-complexity widely-linear (WL-RC) filtering [GAM+14; ALA09]. This is related to the adaptive filtering implementation and is discussed later in this section. They mainly differ in the number of complex terms to be processed in the AF stage.

Term 1 holds the mirror of the blocker and has typically the strongest contribution that can be easily assessed from the IRR of the receiver. Term 2 has contributions from RF and BB non-linearity, and Term 3 is the BB distortion in the fundamental and third harmonic zone. It is likely that the remaining terms are negligible due to 5^{th}, 7^{th}, and 9^{th} order terms. The previous terms have a maximum order of three and therefore represent the most significant distortion that finally need to be subtracted from the desired signal with a least-mean square (LMS) AF.

In order to demonstrate the relevance of the distortion terms, power levels of the distortion Terms 1-6 of Table 4.1 are studied by means of a two-tone simulation. Noise and I/Q imbalance (mirror-frequency interference) are not considered in this simulation for the sake of simplicity. Figure 4.4(a) depicts the distortion estimates for dominating RF non-linearity, assuming a vector $\mathbf{a} = [a_1, a_2, a_3, a_4] = [5.62, -(84351 + j74391), 3.16, -1588.7]$, i.e. having $|a_2| > |a_4|$. These parameters correspond to the IIP3 and gain values listed in Table 4.3. It



(a) Dominating RF non-linearity.



(b) Dominating BB non-linearity.

Figure 4.4: Power levels of distortion estimates Terms 1-6 of Table 4.1 [GAM+14].

is apparent that the Terms 2 and 3 are the most significant values in this case. Furthermore, it is sufficient to consider the common spectral content of multiple terms by only one term, as the AF adapts them to the total distortion at these frequencies. That is, only distortions of higher-order terms need to be considered that have not been covered already by the lower-order terms.

Table 4.3: Parameters of the Two-Tone Simulation

Parameter	**Value**
$\mathrm{IIP3_{RF}}$	$-10\,\mathrm{dBm}$
G_{RF}	$15\,\mathrm{dB}$
$\mathrm{IIP3_{BB}}$	$+6\,\mathrm{dBm}$
G_{BB}	$10\,\mathrm{dB}$
P_{input}	$-30\,\mathrm{dBm}$

In addition, SNR of the ADC is an important limiting factor for resolving weak distortion terms. The SNR of a typical ADC is approx. 60–80 dB depending on its resolution. Thus, all terms below $-80\,\mathrm{dBFS}$ are covered by the quantisation noise. Consequently, Terms 4-6 in Figure 4.4 are likely to be masked by noise in practice. Typically, the ADC quantisation noise is stronger than noise introduced by the analogue RF front-end.

Figure 4.4(b) illustrates the case for dominating BB non-linearity with $\mathbf{a} = [3.16, -1588.7, 5.62, -(84351 + j74391)]$, i.e. having $|a_2| < |a_4|$. These values result from exchanging the IIP3s and gains in Table 4.3 of RF and BB stage to make the BB stage the dominating non-linearity. Again, Terms 2 and 3 are the most significant ones and IMD in the fundamental and third harmonic zone are almost equally strong, indicating that only little RF distortion is present. Indeed, the order of the meaningful distortion terms can slightly vary.

Finally, it is assumed that the selected distortion terms of Table 4.2 are sufficient for any practical values of RF and BB non-linearities. However, the mitigation architecture can be easily extended for front-ends manifesting a higher sensitivity, severe non-linearity, or both that make the inclusion of higher-order terms necessary.

Another limiting factor for the practical non-linearity modelling is the ADC sample rate f_s used for the mitigation processing. The blocker should have a frequency less than $f_s/6$ to model third-order non-linearity without aliasing. The reason is that any third-order non-linearity occupies three times the bandwidth of the distortion-producing blocker band.

As a conclusion of the aforementioned findings, a practical mitigation architecture is proposed that is a carefully found compromise between the computational complexity and the achievable mitigation performance [GAM+14]. This architecture is proposed in Figure 4.5. Here, Terms 1 to 3 are moved in parallel and manifest separate branches that are adapted simultaneously by multitap AFs to consider also frequency-selective nature of the non-linearities (memory effects). Some branches employ a real bandstop, depicted by $h_{\mathrm{BS}}^{\mathrm{R}}$ in Figure 4.5,

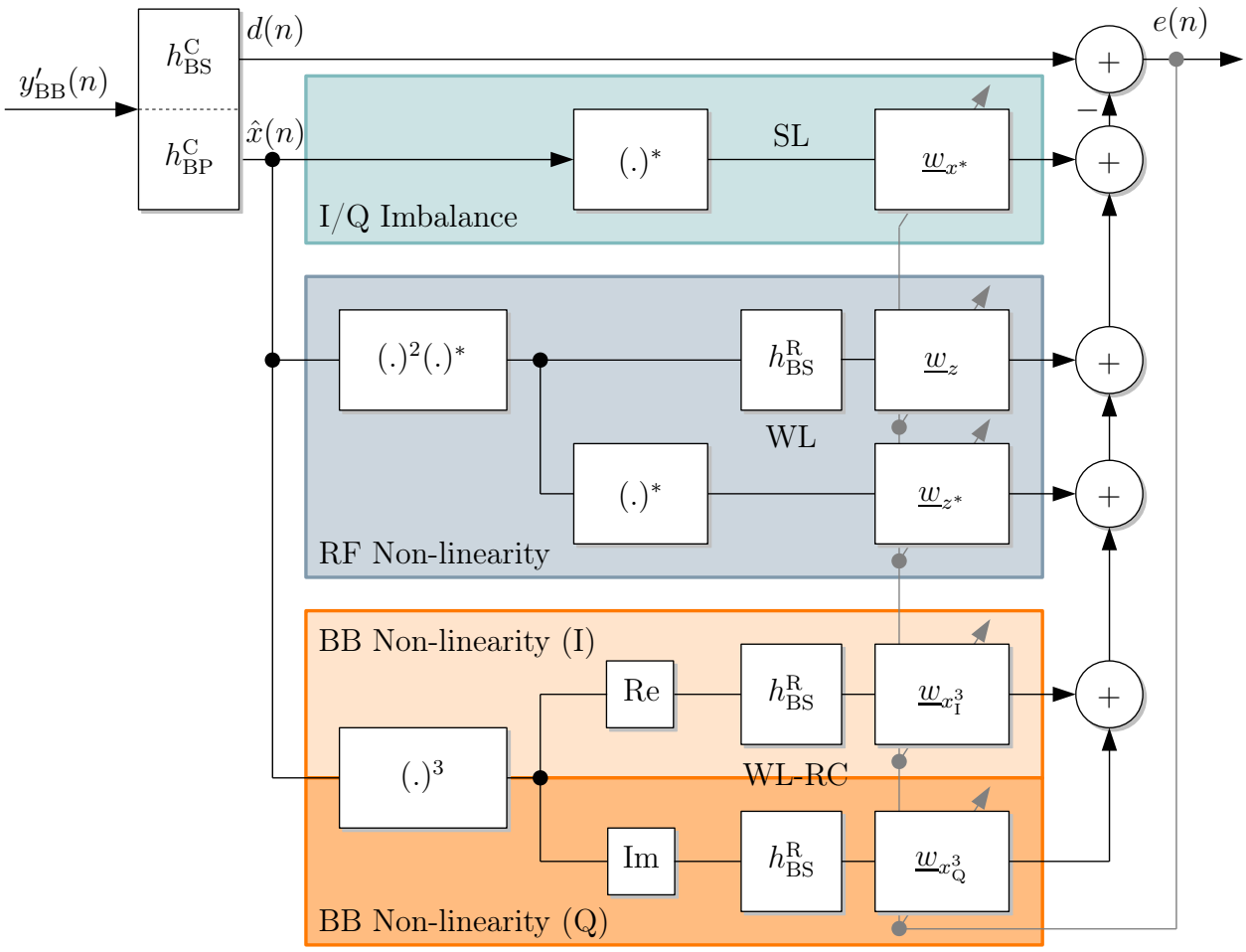


Figure 4.5: Parallel mitigation architecture for cascaded non-linearity [GAM+14].

that filters out the spectral content from the original blocker band. Otherwise, there is a risk of misadjustment by the AF stages as the desired path $d(n)$ does not hold any content from the original blocker band. Different types of filtering have been implemented to find a good trade-off between computational complexity and the additional complex conjugate terms. For Term 1, $\hat{x}^*(n)$, SL filtering is deployed which means that one set of complex AF coefficients $\underline{w}_{x^*}$ is found. For Term 2, $\hat{z}(n)$, also the complex conjugate need to be included to consider mixer I/Q imbalance effects. Therefore, WL is used to find two separate complex sets of AF coefficients $\underline{w}_z$ and $\underline{w}_{z^*}$. In addition $\hat{z}(n)$ need to be filtered to cancel the blocker band content, whereas no filtering is necessary (or at least not harmful) for $\hat{z}^*(n)$ as it contains no signal content at the original blocker band. WL filtering is also used for Term 3, i.e. $\hat{x}^3(n)$ and $[\hat{x}^*(n)]^3$. As both terms need to be filtered by $h^{\mathrm{R}}_{\mathrm{BS}}$, the WL-RC approach can be exploited, which means that finding two complex sets of coefficients for the complex signal and its conjugate is equivalent to finding two complex set of coefficients for the real and

imaginary part of that signal [ALA09; NNS10; GAM+14]. Hence, the complexity of WL is reduced to that of SL. In other words, finding AFs $\underline{w}_{x^3}$ and $\underline{w}_{x_c^3}$ for

$$\underline{w}_{x^3}(n) * \hat{x}^3(n) + \underline{w}_{x_c^3}(n) * [\hat{x}^*(n)]^3 \tag{4.3}$$

is equivalent to finding filters $\underline{w}_{x_I^3}(n)$ and $\underline{w}_{x_Q^3}(n)$ for real and imaginary parts of $\hat{x}^3(n)$, i.e.,

$$\underline{w}_{x_I^3}(n) * \mathrm{Re}[\hat{x}^3(n)] + \underline{w}_{x_Q^3}(n) * \mathrm{Im}[\hat{x}^3(n)]. \tag{4.4}$$

Altogether, the computational complexity is reduced by factor two, because complex AFs are applied for real-valued signals in (4.4) instead of complex-valued signals as in (4.3). Due to the selected filtering approaches, there are in total five distortion branches in the mitigation architecture of Figure 4.5. Note that due to the complex implementation of the AFs also AM/PM distortion at the RF non-linearity, amplitude and phase mismatch at the mixer, and I/Q imbalance in the physical I and Q BB branches are taken into account.

Single RF or BB Non-linearity

If the non-linearity of either the RF amplifier (LNA) or the BB amplifier is dominating, the single type can be considered by just using a subset of the branches of the proposed parallel mitigation architecture in Figure 4.5. Adopting just the BB non-linearity model, i.e. the separated I/Q processing of the blocker raised to its third power, is exactly the procedure followed in the prior work, such as in [VSHGA+06; gha11; GSH+12a]. RF-only modelling has been adopted in [GSH+13; SGT13], assuming severe LNA non-linearity. In fact, the BB-only model sufficiently mitigates the non-linear distortion if the receiver manifests only very mild RF distortion. In Section 4.4, it will be shown that the proposed cascaded model clearly outperforms the RF-only and BB-only model mitigation architectures.

4.2.3 Adaptive Filter

The reference distortions are further processed by an AF to adjust amplitude and phase to the actual distortions in the desired signal branch so that they can be finally subtracted (mitigated). As such parameter learning algorithms are considered as advanced signal processing techniques, a short survey on AF theory is given in the following paragraphs. At the end of the section, a compact notation of the AF is given for mitigating cascaded non-linearity.

AFs have been successfully applied in many different technologies, such as communications, radar, sonar, seismology, and biomedical engineering [Hay02]. They are used in all cases, where a system should respond to a changing input during runtime. That is, an AF adjusts its free parameters (coefficients) in accordance with the input data. Four main classes of applications are distinguished: identification, inverse modelling, prediction, and interference cancellation. The latter one is of interest for mitigating non-linear distortions in a received signal.

Figure 4.6 illustrates the basic architecture of an AF applied for interference cancellation. A

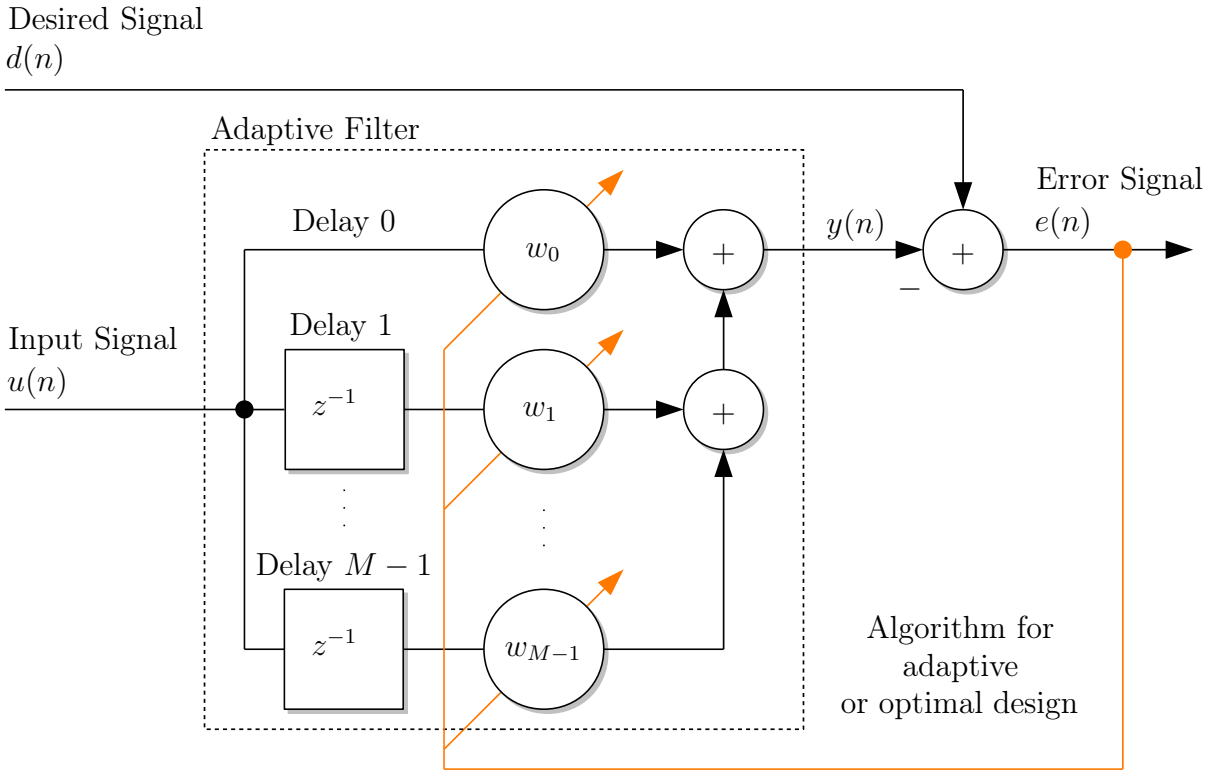


Figure 4.6: Basic AF architecture for interference cancellation

classical application of this AF configuration is the removal of the power-line hum of 50 Hz from an information-bearing desired signal. In the distortion mitigation application, the AF is used to cancel unknown interference contained in the primary (desired) signal. A reference signal, containing an estimate of the distortions generated by the non-linear front-end, is employed as an input to the filter. For this application, the reference signal is first to be derived by passing the distortion-producing blocker(s) through the reference non-linearity model. Then, the main task of the AF is to adjust these estimates so they match those in the desired signal. Finally, a cleaned desired signal is obtained by subtraction of the adjusted

reference distortions. In the following, it is discussed how the algorithms for an optimal or adaptive design of the filter parameters w_k look like.

Optimal Design: Wiener Filter

First, the optimal design, so called Wiener filter, is considered. In the following discussion, FIRs are considered due to reasons of stability and practical considerations. The output signal of the filter is given by (4.1)

$$y(n) = \sum_{k=0}^{\infty} w_k^*(n)u(n-k),\ n = 0, 1, 2, \ldots. \tag{4.5}$$

The algorithm should find the parameters $\{w_0, w_1, w_2, \ldots\}$ such as to minimize the mean-square error (MSE) defined as

$$J = E\left\{e(n)e^*(n)\right\} = E\left\{|e(n)|^2\right\}, \tag{4.6}$$

i.e. the expectation value $E\{.\}$ of the square of the error signal with

$$e(n) = d(n) - y(n). \tag{4.7}$$

The criterion for the cost function J to attain its minimum is that the kth entry of its gradient must be identically zero, that is

$$\begin{aligned}\nabla_k J &= \frac{\partial J}{\partial a_k} + j\frac{\partial J}{\partial b_k} \\ &= E\left\{\frac{\partial e(n)}{\partial a_k}e^*(n) + \frac{\partial e^*(n)}{\partial a_k}e(n) + j\left[\frac{\partial e(n)}{\partial b_k}e^*(n) + \frac{\partial e^*(n)}{\partial b_k}e(n)\right]\right\} \\ &= -2E\left\{u(n-k)e^*(n)\right\} \equiv 0,\end{aligned} \tag{4.8}$$

where the kth filter parameter w_k is denoted in terms of its real and imaginary part $w_k = a_k + jb_k$. Equation (4.8) yields the fundamental principle of orthogonality, which states that the error signal $e_{\text{opt}}(n)$ of the filter in optimum condition is orthogonal to the samples $u(n)$ of the AF input:

$$E\left\{u(n-k)e_{\text{opt}}^*(n)\right\} = 0,\ k = 0, 1, 2, \ldots. \tag{4.9}$$

It can be also used to test if a given filter is operating in its optimum condition. As a corollary of the principle of orthogonality, it can be shown that also the filter output $y_{\text{opt}}(n)$

and the error $e_{\mathrm{opt}}(n)$ are orthogonal to each other. Expanding equation (4.9) with equations 4.5 and 4.7 yields

$$E\left\{u(n-k)\left(d^*(n)-\sum_{i=0}^{\infty} w_{\mathrm{opt}_i} u^*(n-i)\right)\right\}=0$$

$$\sum_{i=0}^{\infty} w_{\mathrm{opt}_i} E\left\{u(n-k)u^*(n-i)\right\}=E\left\{u(n-k)d^*(n)\right\} \tag{4.10}$$

In fact, the left side of equation (4.10) contains the autocorrelation function $r(i-k)$ of the input at lag $i-k$, and the right the cross-correlation $p(-k)$ between the filter input $u(n-k)$ and the desired signal $d(n)$. Replacing the expectations with the correlation functions yields the famous Wiener-Hopf equations in its general form

$$\sum_{i=0}^{\infty} w_{\mathrm{opt}_i} r(i-k)=p(-k). \tag{4.11}$$

Assuming now an FIR with M parameters (length of its impulse response) reduces equation (4.11) to a system of M equations. Then, matrix representation of the Wiener-Hopf equations, now in complex notation, can be formulated with

$$\begin{aligned}
\underline{u} &= \left[u(n)\, u(n-1)\, \ldots\, u(n-M+1)\right]^T \\
\mathbf{R} &= E\left\{\underline{u}(n)\underline{u}^H(n)\right\} = \begin{bmatrix} r(0) & r(1) & \cdots & r(M-1) \\ r^*(1) & r(0) & \cdots & r(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r^*(M-1) & r^*(M-2) & \cdots & r(0) \end{bmatrix} \\
\underline{p} &= E\left\{\underline{u}(n)\underline{d}^*(n)\right\} = \left[p(0)\, p(-1)\, \ldots\, p(1-M)\right]^T \\
\underline{w}_{\mathrm{opt}} &= \left[w_{\mathrm{opt}_0}\, w_{\mathrm{opt}_1}\, \ldots\, w_{\mathrm{opt}_{M-1}}\right]^T,
\end{aligned}$$

where $\mathbf{R}$ is the autocorrelation matrix of the AF input in Toeplitz structure, i.e. it can be built from the first row and first column vector only (symmetry along the diagonal). The Wiener-Hopf equations in its compact form are then given by

$$\mathbf{R}\underline{w}_{\mathrm{opt}}=\underline{p}. \tag{4.12}$$

Assuming that the inverse $\mathbf{R}^{-1}$ exists, the optimum tap-weight vector of the FIR is

$$\underline{w}_{\mathrm{opt}}=\mathbf{R}^{-1}\underline{p}. \tag{4.13}$$

This solution can be obtained if the underlying statistics of the stationary process, meaning the autocorrelation of the input and the cross-correlation of the input with the desired signal, are known. The cost function in equation (4.6) can be written in matrix-vector representation as

$$J_{\underline{w}} = \sigma_d^2 - 2\underline{p}^H\underline{w} + \underline{w}^H\mathbf{R}\underline{w}, \tag{4.14}$$

thus, yielding the minimum MSE with the obtained Wiener-Hopf equations

$$J_{\underline{w}_{\text{opt}}} = \sigma_d^2 - \underline{p}^H\mathbf{R}^{-1}\underline{p}. \tag{4.15}$$

Adaptive Design: LMS Algorithm

Second, the adaptive design of the filter parameters is considered, which approaches iteratively to the Wiener solution. The adaptation of the filter parameters can be achieved by evaluating the gradient of the performance criterion and determining its steepest descent. The gradient of $J(n)$ (equation (4.14)) with respect to the parameters $\underline{w}(n)$ is

$$\nabla_{\underline{w}(n)}J(n) = -2\underline{p} + 2\mathbf{R}\underline{w}(n) = -2E\left\{\underline{u}(n)e^*(n)\right\}. \tag{4.16}$$

From this equation, the adaptation mechanism can be formulated. The idea is to find a new set of filter parameters by taking the old ones and adding something. By moving a small step μ to the opposite site of the gradient vector ($-$ in front of ∇), the parameters will approach to those of the optimal Wiener filter, no matter where the iterations are initialized [Hay02]. Hence, the steepest descent algorithm can be written as

$$\begin{aligned}\underline{w}(n+1) &= \underline{w} - \frac{1}{2}\mu\nabla_{\underline{w}(n)}J(n)\\ &= \underline{w}(n) + \mu E\left\{\underline{u}(n)e^*(n)\right\}.\end{aligned} \tag{4.17}$$

This algorithm will track the time variations of the signal's statistics without solving the Wiener-Hopf equations every time the statistics change. Applying the steepest descent method to the Wiener filter yields

$$\underline{w}(n+1) = \underline{w}(n) + \mu\left[\underline{p} - \mathbf{R}\underline{w}(n)\right]. \tag{4.18}$$

Since the terms $E\left\{\underline{u}(n)e^*(n)\right\}$ and $\underline{p} - \mathbf{R}\underline{w}(n)$ are equivalent, an error-driven adaptation process can be formulated. This is the basis for the common LMS algorithm, named by its originators Widrow and Hoff (1960, [Hay02]). To simplify the algorithm further, the

expectation operator in equation (4.16) can be neglected, that is, an immediately available approximation is used to get an estimate of the gradient

$$\hat{\nabla}_{\underline{w}(n)} J(n) = -2\underline{u}(n)e^*(n). \tag{4.19}$$

Hence, the tap-weight adaptation in the LMS algorithm is done with

$$\underline{w}(n+1) = \underline{w}(n) + \mu\underline{u}(n)e^*(n). \tag{4.20}$$

In practice, exact measurements of the gradient vector are not possible, because this would require prior knowledge of the correlation matrix of the inputs and the cross-correlation between the input and the desired response.

Note that due to the efficient method of the steepest decent, actually no a-priori knowledge of the non-linearity coefficients is necessary. Even if intentionally wrong values are used as initial parameters, the LMS algorithm will converge to the true values very fast. That is, the effort for detailed characterisations of the non-linearity is not necessary for this application. It is more important to choose the core of the model in a way that it matches the physical architecture of the front-end, as already mentioned.

Applying this simplification, computing correlation functions and matrix inversion are not required any more. Algorithms based on 4.20 are simple to design and highly effective in performance. As a consequence of equation (4.19), the adaptation suffers from gradient noise, that is, there will be a random motion around the minimum MSE $J_{\min}$. A general result and limit for LMS-AFs using only first order information is the slow rate of convergence, however, they can react immediately to changes in the input signal statistics.

With properly set step size μ the parameters $\underline{w}$ and the MSE will converge to the Wiener optimal solution, that is

$$\begin{aligned} \underline{w}(n) &\rightarrow \underline{w}_{\text{opt}} \\ J(\underline{w}(n)) &\rightarrow J(\underline{w}_{\text{opt}}). \end{aligned}$$

From the analysis of the convergence of the average parameter vector $E\{\underline{w}(n)\}$, a stability condition for the LMS algorithm can be obtained

$$0 < \mu < \frac{2}{\lambda_{\max}}, \tag{4.21}$$

where $\lambda_{\max}$ is the largest eigenvalue of the autocorrelation matrix $\mathbf{R} = E\left[\underline{u}(n)\underline{u}^H(n)\right]$. Indeed, the eigenvalues are not known in practice. If μ obeys this condition, the LMS converges in mean to the Wiener solution. Another criterion for μ can be derived from the convergence analysis of the MSE which yields

$$\mu < \frac{2}{\text{Power of the input}} \quad \text{for} \quad J_\infty \to J_{\text{opt}}, \tag{4.22}$$

where J_∞ is the steady-state value of the MSE. This relationship for μ can be used in practical implementations as the power of the input can be easily determined. The choice of a small μ is accompanied with a trade-off between convergence in speed and MSE. Using a small μ, the adaptation progresses slowly while achieving a small misadjustment between the steady-state MSE J_∞ and the minimum MSE J_{opt}. The reciprocal of μ can be viewed as the memory of the LMS filter. Low μ causes "long memory", that is, a large number of samples need to be processed to reach a significant small MSE. Choosing a large μ results in fast adaptation, but with increased misadjustment for the MSE.

There are several variants of the LMS algorithm, among others, the normalised LMS algorithm with the updating equation

$$\begin{aligned} \underline{w}(n+1) &= \underline{w}(n) + \mu(n)\underline{u}(n)e^*(n) \quad \text{with} \\ \mu(n) &= \frac{\tilde{\mu}}{\alpha + \sum_{i=0}^{M-1} u(n-i)^2} = \frac{\tilde{\mu}}{\alpha + \|\underline{u}(n)\|^2} \approx \frac{\tilde{\mu}}{\alpha + ME\left\{u(n)^2\right\}}, \end{aligned} \tag{4.23}$$

and a small constant $\alpha > 0$ (overcome numerical difficulties when $\|\underline{u}(n)\|$ is close to zero) as well as the adaptation constant $\tilde{\mu}$. In the normalised least-mean square (NLMS) algorithm, the step size changes every adaptation while taking the power of the input into account, that is, the adaptation becomes data-dependent. The power is computed as the sum of squares of the past M samples. Thereby, large fluctuations in the power levels and the length N of the input vector by reducing the step size are considered. The stability condition for NLMS reduced to the simple relationship

$$0 < \tilde{\mu} < 2. \tag{4.24}$$

In fact, the small constant a also depends on the input signal power, that is, the NLMS algorithm has 2 degrees of freedom, the numerical constant a and the adaptation constant $\tilde{\mu}$. As a result of a direct comparison between LMS and NLMS, the normalised version achieves a similar steady-state MSE but with faster convergence. Assuming similar convergence speed, NLMS provides lower steady-state MSE. In other words, the NLMS offers better trade-

offs between convergence speed and stationary MSE than conventional LMS at the cost of slightly higher computational complexity. In practice, (4.24) eases the choice of the step size dramatically compared to the LMS. This is especially important for mitigating cascaded non-linearity with multiple distortion estimates, that can have very different powers varying over several orders of magnitudes.

Alternative algorithms try to monotonically reduce the step size in such a way that a large μ is used at the beginning for fast adaptation and a very small μ is applied at the end to achieve a small MSE. However, these algorithms cannot react to changes of a non-stationary input any more. Sign algorithms achieve a very fast adaptation by modifying the estimate of the gradient $\hat{\nabla} J$ through replacing the multiplications with the sign of the error signal, the input, or both. Beside a fast convergence, the steady-state error will be higher. Further approaches base on smoothing the gradient estimate by filtering, double updating by computing the filter output with the old and new set of parameters, or the extension with quadratic combinations of the input for filtering (LMS Volterra algorithm). For details on these algorithms, the reader is referred to [Hay02].

Linear Least Squares Filtering

The method of the LS can be used as an alternative approach to develop a linear AF to adjust the distortion estimates. It provides a prediction for the optimal weights for the distortion estimates so they fit the desired signal. The advantage of the LS approach is the existence of closed-form expressions for the estimate as well as for its properties. The result depends on the number of samples N used for the computation. That is, the LS solution is computed block-wise. The LS can be seen as an alternative to the Wiener-Hopf solution in equation (4.13). While the Wiener solution is derived from ensemble averages (set of multiple realisations) and the parameters are the optimal ones for all realisations of the underlying physical process, the LS finds the optimum parameters for a single realisation, i.e. for the given data. In other words, the Wiener solution is the average of the block-wise LS solution and requires the full statistics of the signal. Although both are fully theoretical approaches and cannot be implemented in practice, they provide a kind of upper bound for the best achievable mitigation performance.

Assuming a set of distortion estimates samples $\underline{u_i}$ and the desired signal vector $\underline{d}$, the task is to find the parameters w_{k_i} of the linear filter with the output

$$y_i(n) = \sum_{k=0}^{M-1} w_{k_i}^* u_i(n-k) = \underline{w}_i^H(n)\underline{u}_i(n). \tag{4.25}$$

The estimation error is defined as the difference between the desired response and the filter output [Hay02]

$$e(n) = d(n) - y(n). \tag{4.26}$$

In the LS method, the goal is to minimize the cost function

$$\mathcal{E}(w_0, w_1, \ldots, w_{M-1}) = \sum_{i=i_1}^{i_2} |e(i)|^2, \tag{4.27}$$

that is, the sum of error squares, where i_1 and i_2 are the index limits where the minimisation is done. While the error criterion of the Wiener filter in equation (4.6) used expectations, this cost function is now closer to the actual data. To find the filter parameters $\underline{\hat{w}}$ for the minimal error $e_{\min}$ when the cost function is minimized, a so called normal system of equations is used [Hay02]. It can be derived by the gradient vector $\nabla\mathcal{E}$ or the principle of orthogonality. The matrix formulation of these equations is

$$\mathbf{\Phi}\underline{\hat{w}} = \underline{\psi}, \tag{4.28}$$

where $\mathbf{\Phi}$ and $\underline{\psi}$ are the correlation matrix of the input $\underline{u}$ and the cross-correlation vector between the input and the desired signal, respectively. The tap-weights of the LS filter are then

$$\underline{\hat{w}} = [\mathbf{\Phi}]^{-1}\,\underline{\psi}, \tag{4.29}$$

assuming that the inverse of $\mathbf{\Phi}$ exists. Equation (4.29) is the counterpart to the solution of the Wiener-Hopf equations in equation (4.13). A compact form of equation (4.28) with the actual data matrices is derived with the help of the following redefinitions:

$$\mathbf{\Phi} = \mathbf{A}^H\mathbf{A}$$
$$\underline{\psi} = \mathbf{A}^H\underline{d},$$

where $\mathbf{A}$ is a rectangular matrix constructed from the input samples by

$$\mathbf{A} = \begin{bmatrix} u(i_1) & u(i_1-1) & u(i_1-2) & \cdots & u(i_1-M+1) \\ u(i_1+1) & u(i_1) & u(i_1-1) & \cdots & u(i_1-M+2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u(i_2) & u(i_2-1) & u(i_2-2) & \cdots & u(i_2-M+1) \end{bmatrix} = \begin{bmatrix} \underline{u}(i_1)^H \\ \underline{u}(i_1+1)^H \\ \vdots \\ \underline{u}(i_2)^H \end{bmatrix}, \tag{4.30}$$

where the limits i_1 and i_2 belong to a used data window. Equation (4.29) can then be rewritten as

$$\underline{\hat{w}} = \left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H\underline{d}, \tag{4.31}$$

containing only the known input and desired signal. This solution is equal to LS approximation of the general overdetermined system of equations in linear algebra [Str09]

$$\mathbf{A}\underline{x} = \underline{b}, \tag{4.32}$$

where $\mathbf{A}$ is a $K \times M$ matrix and $\underline{b}$ a column vector of length K ($K \times 1$). The matrix $\mathbf{A}$ has more rows than columns ($K > M$), which is the common case in this application. There are many samples of input data, but only M unknown coefficients. Then there is a best $\underline{\hat{x}}$ which minimizes the sum of squares of the errors in the K equations

$$\|\mathbf{A}\underline{x} - \underline{b}\|^2 \tag{4.33}$$

with

$$\underline{\hat{x}} = \left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H\underline{b} = \mathbf{A}^+\underline{b}. \tag{4.34}$$

$\underline{\hat{x}}$ yields the combination of the normal equations with the smallest MSE. In other words, the estimate can be computed by the pseudoinverse (or Moore-Penrose generalized inverse) $\mathbf{A}^+ = \left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H$. Equation (4.31) and (4.34) are equivalent. In other words, $\underline{\hat{x}}$ fits the distortion estimates to the sample vector in the LS sense, without having an AF taken into consideration. The filter coefficients became here the unknown coefficients of a model, which fits to the data.

Finally, the adjusted estimates can be subtracted from the desired signal, to get the cleaned signal

$$\hat{y} = \underline{b} - \mathbf{A}\underline{\hat{x}}. \tag{4.35}$$

Attention has to be paid on the construction of the data matrix $\mathbf{A}$. In the simplest case, it contains the distortion estimate sample vectors in the columns. It is important to mention that the condition number of $\mathbf{A}$, indicating the accuracy for the matrix inversion, determines the quality of the results for the found coefficients. A condition number of 1 indicates a well-conditioned matrix. If it raises significantly, e.g. by adding more correlated distortion terms in the columns, the LS solution differs a lot from the LMS solution. High condition numbers basically mean that there are multiple solutions for solving the data fitting problem. Therefore, also different coefficients by LMS and LS will be found.

Finally, the LMS algorithm is chosen instead of the LS method. The LS filtering requires a long number of samples to provide useful results. For practical implementation, it is difficult to store and process a high amount of data, and to subsequently correct the cached desired signal with the found coefficients. Instead, the LMS algorithm with a continuous adaptation at every sample instant is much easier to implement. However, the LS is computed in simulations to compare it with the LMS solution.

LMS Notation for Cascaded Non-linearity

Finally, a vector notation for mitigating cascaded non-linearity based on LMS-based learning algorithm is given [GAM+14]. The special aspect for the non-linearity mitigation is that multiple reference signals, the distortion estimates, need to be subtracted from the desired signal at once. Furthermore, multi-tap AFs with a filter length $M > 1$ are typically required to compensate also for memory effects in the non-linear receiver front-end.

Following Figure 4.5, there are in total five distortion estimate signals that can be combined in one vector $\underline{s}(n)$ of size $5M \times 1$, namely

$$\underline{s}(n) = \begin{bmatrix} \underline{s}_{x^*} \\ \underline{s}_{z} \\ \underline{s}_{z^*} \\ \underline{s}_{x_\mathrm{I}^3} \\ \underline{s}_{x_\mathrm{Q}^3} \end{bmatrix}, \tag{4.36}$$

where the subscripts indicate the distortion branches. For instance, Term 1 vector is

$$\underline{s}_{x^*}(n) = \left[\hat{x}^*(n), \hat{x}^*(n-1), \ldots, \hat{x}^*(n-M+1)\right]^T. \tag{4.37}$$

Note that the vectors $\underline{s}_z(n)$, $\underline{s}_{x_\mathrm{I}^3}(n)$, and $\underline{s}_{x_\mathrm{Q}^3}(n)$ also include filtering with h_BS^R to remove the original blocker band content from the distortion estimate. Then, $\underline{s}_z(n)$, e.g., look like

$$\underline{s}_z(n) = \left[\hat{z}_\mathrm{filt}(n), \hat{z}_\mathrm{filt}(n-1), \ldots, \hat{z}_\mathrm{filt}(n-M+1)\right]^T, \tag{4.38}$$

where $\hat{z}_\mathrm{filt}(n)$ denotes the bandstop-filtered version of $\hat{z}(n)$. Next, also the AFs are combined in one vector $\underline{w}(n)$ of size $5M \times 1$ given by

$$\underline{w}(n) = \begin{bmatrix} \underline{w}_{x^*} \\ \underline{w}_z \\ \underline{w}_{z^*} \\ \underline{w}_{x_\mathrm{I}^3} \\ \underline{w}_{x_\mathrm{Q}^3} \end{bmatrix}, \tag{4.39}$$

where the subscript indicates the corresponding distortion branch. Note that each individual AF has a length M. Also different AF lengths for different distortion estimates can be employed, if e.g. RF and BB non-linearities manifest different levels of frequency-selectivity.

The combined AF output is then computed by

$$\tilde{x}(n) = \underline{w}^H(n)\underline{s}(n), \tag{4.40}$$

and the resulting error signal

$$e(n) = d(n) - \tilde{x}(n). \tag{4.41}$$

Finally, the tap-weight update equation according to (4.20) is given by

$$\underline{w}(n+1) = \underline{w}(n) + \mathrm{diag}(\underline{\mu})\underline{s}(n)\tilde{x}(n), \tag{4.42}$$

where diag(.) denotes conversion of a vector to a diagonal matrix. Thereby, every distortion estimate is processed by a different step size. They can be collected in a step size vector

$$\underline{\mu} = \left[\mu_{x^*}, \mu_z, \mu_{z^*}, \mu_{x_\mathrm{I}^3}, \mu_{x_\mathrm{Q}^3}\right], \tag{4.43}$$

where the subscripts indicate the corresponding distortion branches. Following the NLMS algorithm according to (4.23) yields step-size vector

$$\underline{\mu}_\mathrm{NLMS} = \left[\frac{\mu_{x^*}}{\alpha_{x^*} + \left\|\underline{s}_{x^*}(n)\right\|^2}, \ldots, \frac{\mu_{x_\mathrm{Q}^3}}{\alpha_{x_\mathrm{Q}^3} + \left\|\underline{s}_{x_\mathrm{Q}^3}(n)\right\|^2}\right], \tag{4.44}$$

with the additional selectable numerical constants $\underline{\alpha} = \left[\alpha_{x^*}, \alpha_z, \alpha_{z^*}, \alpha_{x_\mathrm{I}^3}, \alpha_{x_\mathrm{Q}^3}\right]$. NLMS is chosen in Section 4.4 due to its preferences in convergence and choosing the step size.

4.3 Computational Complexity

With the system-level implementation of all three main components (FIR filter, reference non-linearity, and AF), the complexity of the digital feedforward mitigation algorithm can be computed. The overall complexity is summarised in Table 4.4. The complexity for creating

Table 4.4: Computational complexity for mitigating cascaded non-linearity [GAM+14].

Operation	**# Ops.**	**Mutiplications**	**Summations**	**Divisions**
Ref. Modeling		16	8	-
Static Filtering	5	$12P$	$12P-8$	-
SL LMS	1	$8M+2$	$8M$	-
WL-RC LMS	1	$8M+2$	$8M$	-
WL LMS	1	$16M+2$	$16M$	-
NLMS Scaling	5	$10M+5$	$5M+5$	5
Overall		$12P+42M+27$	$12P+37M+5$	5

the distortion estimates is 16 real multiplications and 8 real summations according to the complex-valued operations in Figure 4.5.

Add to this, there is the cost of the five filters applied in the mitigation architecture, namely the two complex filters h_BP^C and h_BS^C and three instances of the real-valued filter h_BS^R. If P denotes the order of the direct-form FIR filters, P multiplications and $P-1$ summations are necessary for real-valued filters, processing a real-valued signal [MRS91]. Same orders are assumed for all filters, as same requirements in transition bandwidth and stopband attenuation hold for all of them. Two instances of h_BS^R process real-valued signals (BB non-linearity branches), whereas all other filters operate on complex-valued signals. Multiplying complex input with a complex tap weight requires $4P$ real multiplications and $2P+2(P-1)$ summations, where the $2P$ comes from the complex multiplications and $2(P-1)$ summations from the parallel filtering of I and Q signals. Hence, the computational load of the two complex filters h_BP^C and h_BS^C processing complex signals is twice that amount, i.e. $8P$ real multiplications and $8P-4$ summations. Then, there is one filter h_BS^R operating on complex signals, requiring $2P$ real multiplications and $2(P-1)$ summations for filtering the I and Q branches. The remaining two real filters on real signals need $2P$ multiplications and $2(P-1)$

summations. Overall, the complexity of the filtering sums up to $12P$ multiplications and $12P - 8$ summations.

Finally, there is the cost of the adaptation by the complex LMS. The complexity of the conventional LMS is $2M + 1$ complex multiplications and $2M$ complex summations per iteration [Hay02]. The SL and WL-RC cases require $8M + 2$ real multiplications, $4M$ each AF filtering and tap-weight-update plus two for the multiplication of real and imaginary part with the step size. The number of summations accounts for $4M - 1$ real summations for the tap-weight updating and the AF filtering, respectively, plus two summations in the tap-weight update equation and the error signal equation. Hence, the total amount of real summations for SL and WL-RC is $8M$. The complexity for WL LMS is twice the amount of SL LMS as two complex signals (strict and conjugate) need to be processed, introducing $16M + 2$ real multiplications and $16M$ summations. Including also the NLMS update introduces additional $2M$ real multiplications and M real summations for computing the euclidean norm plus 1 real division (Divs) and 1 real summation (with the numerical constant) for finally scaling the step size. Thus, the additional amount for NLMS scaling of five distortion estimates is $10M + 5$ real multiplications and $5(M + 1) = 5M + 5$ real summations.

The final numbers of multiplications, summations, and divisions are summed up in the last row of Table 4.4. The introduced processing delay of the digital feedforward mitigation is $P + M$ that is practically dictated by the filter order P, as $P \gg M$.

In particular, the computational load is important for real-time implementation, as discussed in Chapter 6, and for performance comparisons with other linearisation techniques.

4.4 Proof of Concept

This sections provides simulation and RF measurement results which evaluate the performance of the proposed digital feedforward mitigation. In particular, explicit comparisons between the novel joint processing of RF and BB distortion and the RF-only and BB-only approaches are conducted. These investigations base on two-tone and binary phase shift keying (BPSK) excitation signals that are used for both simulations and measurements. These results have been published in [GAM+14].

4.4.1 Performance Simulations

The basic simulation architecture is depicted in Figure 4.7. The complex BB signal passes an AWGN source, the RF non-linearity, the I/Q imbalance block, and the BB non-linearity. All the blocks within the dashed frame mimic a DCR front-end at equivalent BB. In the simulations, the SNR of the blocker is 61 dB, related to the entire BB bandwidth of 25 MHz. The IRR is set to 30 dB, a value that is realistic for an integrated DCR [Raz97]. Finally, the composite signal is processed by the digital feedforward mitigation algorithm presented in the previous sections.

Two-tone and BPSK signals have been generated as excitation signals to cover two signal classes, CW and modulated signals, for performance evaluation of the algorithm. The following investigations are evaluated from the sensing point of view (use case 2), i.e. there is no specific signal of interest. Instead, the processing aims for cleaning the whole BB spectrum from distortions meaning that only the strong input signal should be left after mitigation. For this purpose, the blocker is added back to the output signal $e(n)$ of the mitigation structure, as described in Section 4.1.

As there is a perfect match between the model mimicking the DCR in the simulator and the reference non-linearity model in the mitigation architecture, single-tap complex AFs are sufficient in simulations to accomplish full cancellation of the distortion. In addition, both models are memoryless. After mitigation, block-wise 1024-point FFT is applied to analyse and illustrate the remaining non-linear distortion. Always the last one of 29 FFT blocks is shown in which the convergence of the AF coefficients is guaranteed.

The mitigation gain is computed as a figure of merit in dB that describes the reduction of average signal power outside the original input (blocker) frequency band. The blocker mirror band has been also excluded from the mitigation gain calculations as the focus is purely on non-linear distortion mitigation. The mitigation gain computation is established

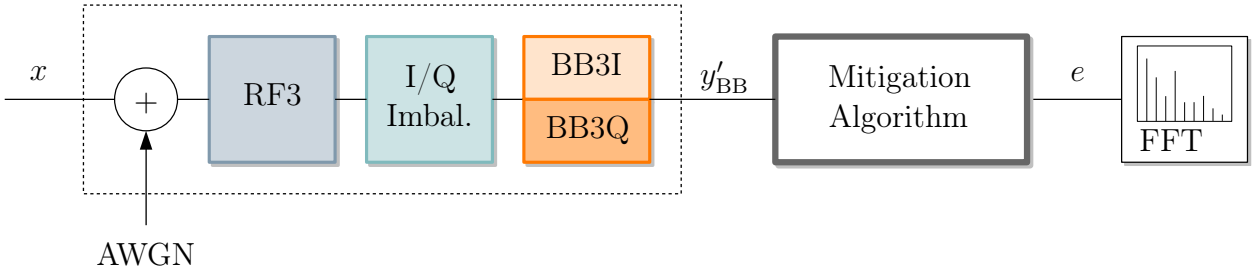


Figure 4.7: Simulation architecture for performance evaluation of the mitigation algorithm in MATLAB [GAM+14].

in the frequency domain by the following rule

$$\mathrm{MG} = 10\log_{10}\left(\mathrm{mean}\left\{\frac{\mathbf{Y}_{\mathrm{in}}(\overline{\mathrm{blocker} \vee \mathrm{mirror}})}{\mathbf{Y}_{\mathrm{mit}}(\overline{\mathrm{blocker} \vee \mathrm{mirror}})}\right\}\right), \tag{4.45}$$

where $\mathbf{Y}_{\mathrm{in}}$ and $\mathbf{Y}_{\mathrm{mit}}$ denote a matrix with the set of block-wise FFTs of the original input and the mitigated output, respectively. The logical indices contain all bins of the FFT block that do *not* belong to the blocker or mirror band. Compared to the BPSK case, definition of the mitigation gain for the two-tone case is a bit different. Here, the gain is computed explicitly at the actual frequency components containing the non-linear distortion, according to the rule

$$\mathrm{MG}_{\mathrm{TT}} = 10\log_{10}\left(\mathrm{mean}\left\{\frac{\mathbf{Y}_{\mathrm{in}}(\underline{\mathrm{spurs}}(s))}{\mathbf{Y}_{\mathrm{mit}}(\underline{\mathrm{spurs}}(s))}\right\}\right), \tag{4.46}$$

where $\underline{\mathrm{spurs}}(s)$ is a vector holding discrete indices of the spurious frequency components caused by the two-tone input. In addition, mirrors of the distortion products are considered in simulations due to exaggerated coefficients used for the models. However, in measurements, these components do not appear above the noise level. Consequently, the absolute mitigation gain obtained from simulations is a bit higher.

The parameters of the performance simulations are summarised in Table 4.5, extending the values of Table 4.3. It lists all coefficients of the simulator model including the non-linearity

Table 4.5: Parameters of the performance simulations for mitigating cascaded non-linearity.

Parameter	**Value**
$\mathrm{IIP3}_{\mathrm{RF}}$	$-10\,\mathrm{dBm}$
G_{RF}	$15\,\mathrm{dB}$
$\mathrm{IIP3}_{\mathrm{BB}}$	$+6\,\mathrm{dBm}$
G_{BB}	$10\,\mathrm{dB}$
P_{input}	$-30\,\mathrm{dBm}$
$\mathbf{a}$	$[5.62, -(84351 + j74391), 3.16, -1588.7]$
g_{m}	0.99
ϕ_{m}	$0.0628 \mathrel{\widehat{=}} 3.6°$ } $\mathrm{IRR} \approx 30\,\mathrm{dB}$
$\boldsymbol{\mu}$	$[1, 1, 0.01, 1, 1]$
$\boldsymbol{\alpha}$	$[10^{-9}, 10^{-8}, 10^{-4}, 10^{-9}, 10^{-8}]$

and I/Q imbalance coefficients. In addition, phase distortions created by the LNA and the

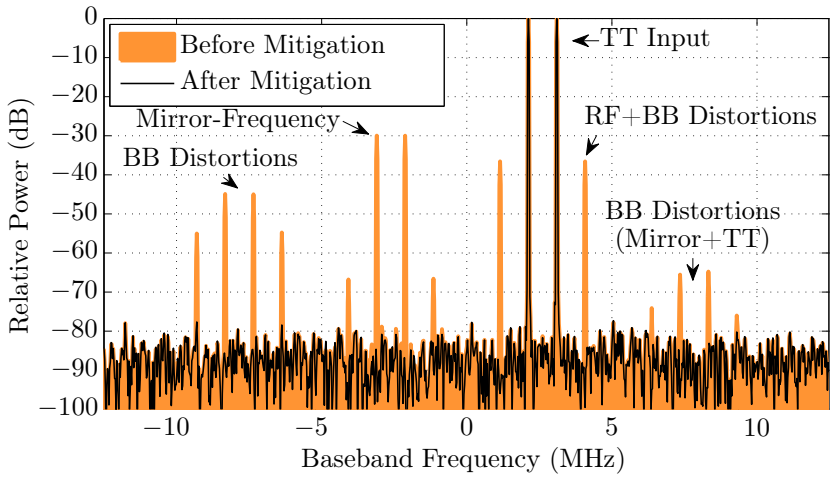
mixer are considered by complex coefficients a_2, k_1, and k_2. The coefficients $a_1 \dots a_4$ have been computed based on the listed gain and IIP3 values of the RF and BB amplifiers by utilising the conversion formulas (2.15), (2.16), and (2.17).

Two-tone Blocker Input

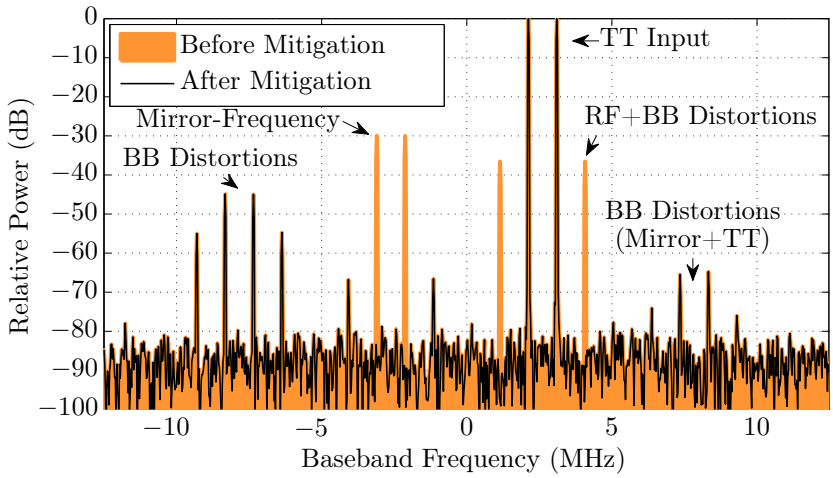
Figure 4.8(a) illustrates the BB spectrum with the two-tone input and after mitigation with the cascaded model. In total, 16 components are created due to the receiver non-linearities and I/Q imbalance. The mirror-frequency appears 30 dBc (below the original tones), whereas the strongest distortion in the fundamental zone is 36 dBc. First, the two-tone blocker is distorted by the RF amplifier, generating IMD in the fundamental zone at around +2.6 MHz (center of the two tones). Second, mirror-frequency interference of the blocker and the RF distortion appear due to down-converting mixer I/Q mismatches. Third, blocker, RF distortion and their respective mirror components are all further distorted by the BB I and Q non-linearities. Hence, distortion in the third harmonic zone at around −7.8 MHz and in the fundamental zone are added. In addition, the BB non-linearity causes HD and IMD originating from the mirror components which then appear as low-power tones around +7.8 MHz. Furthermore, the BB distortions can have respective mirror components, if there is I/Q imbalance in the BB branches, although these components are very likely to remain below the noise floor in practice. In fact, the mirror components of the main RF and BB distortions are mitigated with the WL filters in the proposed structure if they do appear. Furthermore, AM/PM distortion caused by the RF amplifier [Ken00] and phase mismatch introduced by the mixer are taken into account by the complex implementation of the AFs.

In Figure 4.8(a), the average suppression of non-linear distortion with the total cascaded model is 32.6 dB, indicating full cancellation of all distortion and mirror components. The convergence of the AF coefficients by using the NLMS algorithm is illustrated in Figure 4.9. The deployed NLMS parameters are also listed in Table 4.5. In Figure 4.8(a), magnitudes of the adapted complex coefficients are shown together with the corresponding LS solutions for the coefficients, depicted by dashed lines. In general, there is a good match between the steady-state solutions of the adapted coefficients and the LS solutions. About 3000 samples (iterations of the AF) are required to achieve convergence.

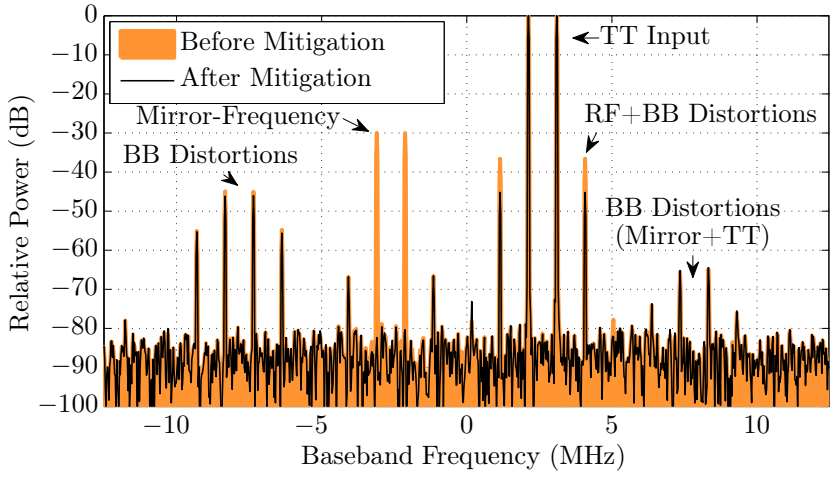
Beside the total cascaded model, simulations with the RF-only and BB-only model are also shown in Figure 4.8. However, the same simulator model mimicking the DCR front-end has been used for all cases. For the RF-only results in Figure 4.8(b), only the blocker mirror and the RF distortion in the fundamental zone are mitigated (Terms $\hat{x}^*$ and $\hat{z}$). Since no distortion estimate for the components in the third harmonic zone are provided by the



(a) Total cascaded non-linearity model.



(b) RF-only model.



(c) BB-only model.

Figure 4.8: Mitigation results with two-tone excitation obtained with different reference models [GAM+14].

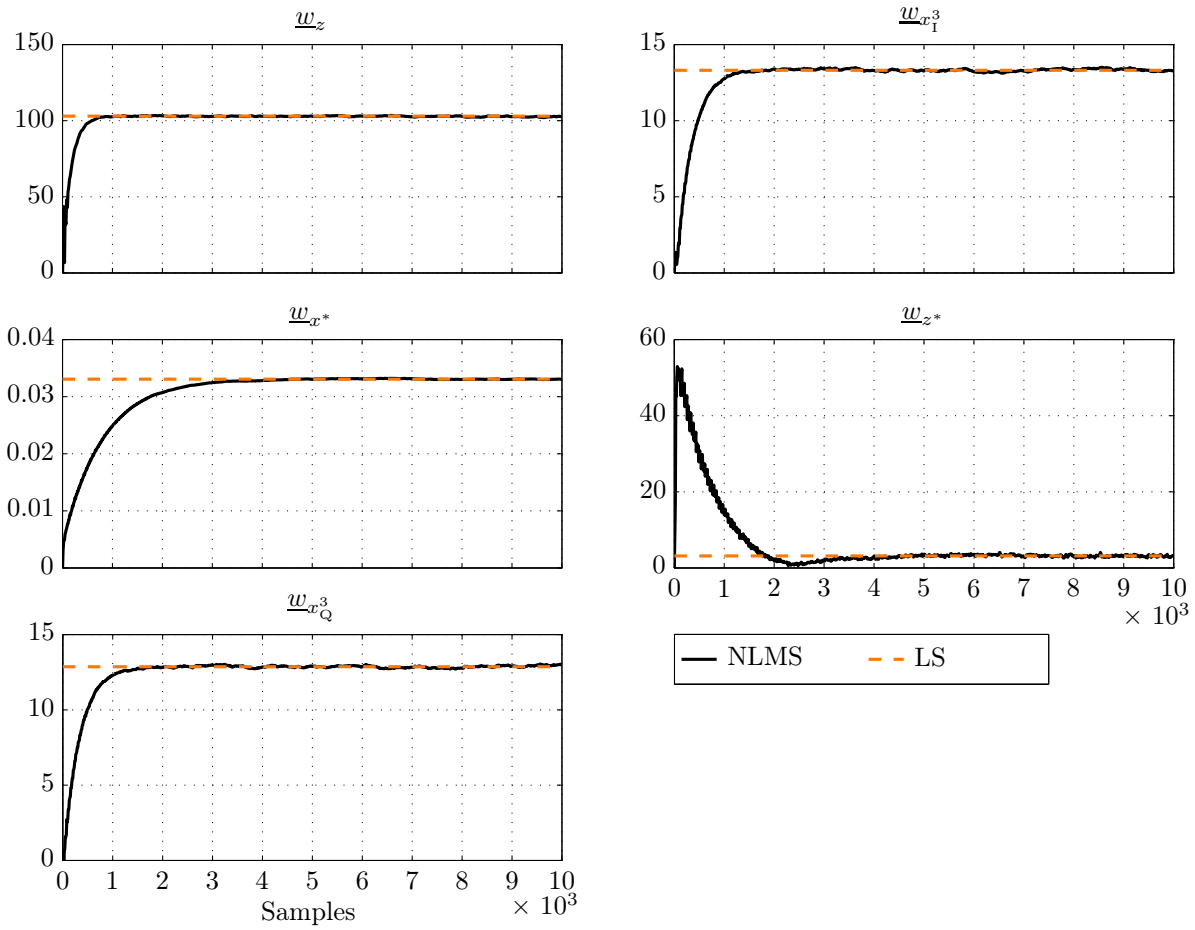


Figure 4.9: Adaptation of the AF coefficients with two-tone input (magnitude of complex coefficients) [GAM+14].

model, no distortion mitigation is achieved here. On the contrary, in Figure 4.8(c), only the blocker mirror and the third-order non-linearity at BB are considered (Terms $\hat{x}^*$, $\mathrm{Re}\,[\hat{x}^3]$, and $\mathrm{Im}\,[\hat{x}^3]$). There is a poor performance of the BB-only model as the RF distortion is not taken into account. Consequently, distortion products in the fundamental zone, but also HD and IMD of the mirror image cannot be fully cancelled. However, if the RF and mixer distortion is mild and does not manifest severe phase distortion (AM/PM), the BB-only model can totally remove all occuring BB distortion. Finally, the joint processing of RF and BB distortion clearly outperforms the RF-only and BB-only model considered in the state of the art.

Additional simulations have been conducted which analyse the mitigation gain vs. the power of the two-tone input obtained with the different models, to further proof the efficiency of the total cascaded model. The achieved gain of the three models as well as the ideal suppression, indicating how much spurious power has been added due to the receiver non-linearity, are

illustrated in Figure 4.10. With the deployed IIP3 values of Table 4.5, almost no IMD is

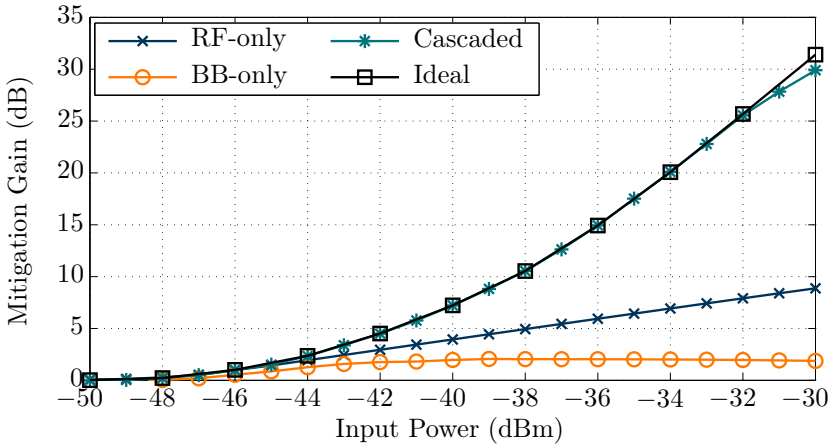


Figure 4.10: Average suppression of non-linear distortion components for different mitigation models obtained with simulated data [GAM+14].

generated up to the input power of −46 dBm. Here, only the blocker mirror x^* is dominating. As all models include the blocker mirror mitigation, the blocker mirror band is excluded from the mitigation gain calculations to enhance visibility of non-linear distortion mitigation. RF and BB distortions raise with increasing input power and different gains are achieved with different models. The BB-only model performs poorly due to the strong RF non-linearity, whereas the total cascaded model follows the ideal suppression. That is, all non-linear distortion products and mirror terms are essentially cancelled. In brief, the total cascaded model provides much better performance than the RF-only or BB-only non-linearity models due to their fundamental shortcomings. These results show that proper modelling of the underlying receiver architecture is essential to achieve improved mitigation performance in a broad variety of applications.

BPSK Blocker Input

Next, a BPSK-modulated blocker is assumed to show that the digital feedforward mitigation algorithm is totally independent from the actual blocker waveform. The BPSK signal is generated with a raised-cosine pulse-shaping filter, with a roll-off factor of 0.5, and a symbol rate of approx. 788 ksym/s. These values have been chosen arbitrarily to generate a simple modulated blocker that covers approx. 1 MHz bandwidth. Beside that, the parameters for the BPSK simulation are exactly the same as listed in Table 4.5. The RMS power has been chosen to be equal to that in the two-tone case. Due to the modulation, the PAPR of the BPSK is with $\mathrm{PAPR}_{\mathrm{BPSK}} = 4.1\,\mathrm{dB}$ higher than that in the two-tone case

($\mathrm{PAPR}_{\mathrm{TT}} = 3.01\,\mathrm{dB}$). This is important as the peak values of the time-domain signal determine the level of distortion.

A mitigation gain of 20.4 dB is achieved for the BPSK case, as indicated in the cleaned BB spectrum shown in Figure 4.11. The AF coefficients are converged after approx. 4000

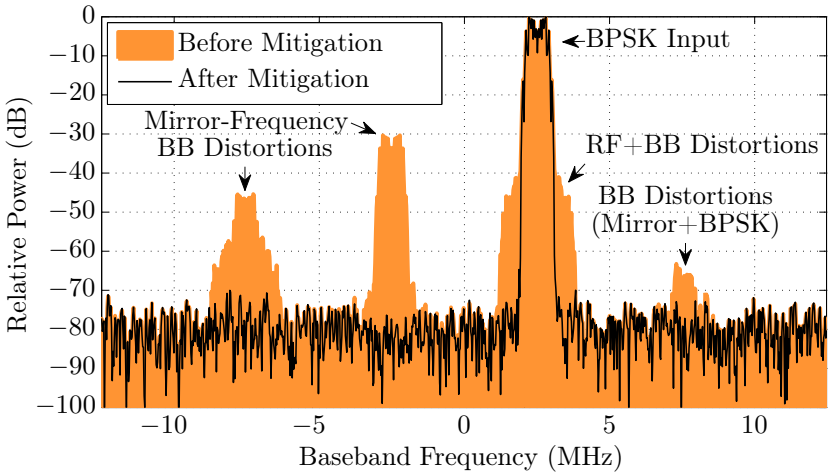


Figure 4.11: Mitigation results with BPSK excitation [GAM+14].

samples, by using the NLMS algorithm. The adaptation of the five distortion branches of the total cascaded model is illustrated in Figure 4.12. The results with BPSK input verify that the feedforward mitigation algorithm is able to mitigate distortions induced by modulated blocker waveforms.

4.4.2 RF Measurements

In order to verify the efficiency of the digital feedforward mitigation with true RF signals and components, real-world measurements with the USRP+WBX front-end have been conducted [Ett; GAM+14]. The automated setup for the experiments with the aforementioned signal scenarios, similar to that employed in Section 2.7.2, is illustrated in Figure 4.13. The signals were generated by the vector signal generator of type Rohde & Schwarz SMU200A. Before running the actual measurements, the spectral purity (SFDR) of the generator was checked with a conventional spectrum analyser, as they suffer from a limited linearity. The generator itself has a SFDR $> 80\,\mathrm{dB}$, i.e. it can be guaranteed that all remarkable non-linear distortions are solely generated in the non-linear DUT. Signal generator and receiver are synchronized using a 10 MHz reference signal, enabling coherent sampling for precise power measurements. The IIP3 of the total receiver RF front-end including RF, BB, and ADC non-linearity has been determined already in Section 2.7.2 and are listed in Table 2.5.

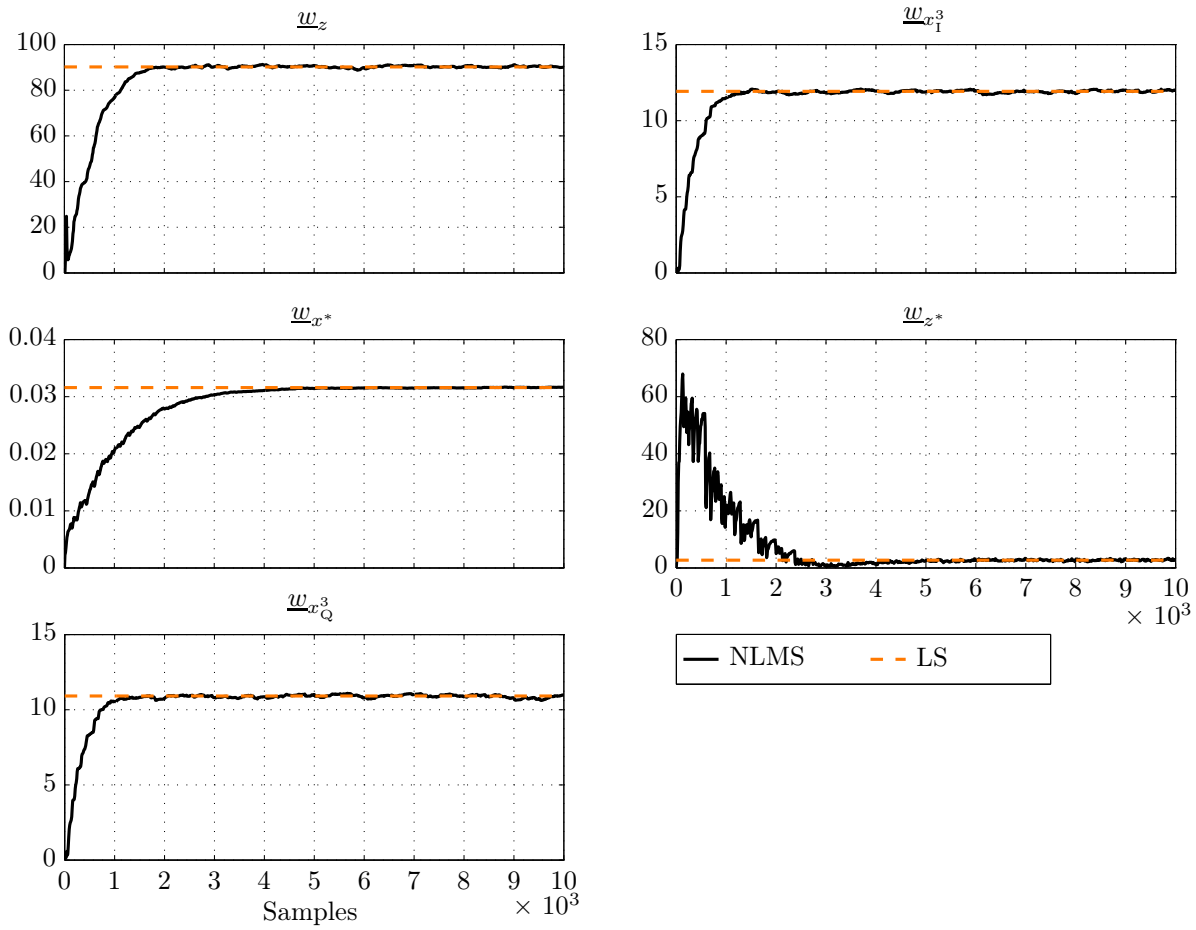


Figure 4.12: Adaptation of the AF coefficients with BPSK input (magnitude of complex coefficients) [GAM+14].

Figure 4.14(a) illustrates the mitigation performance achieved with a two-tone blocker input. The tone frequencies are the same as in simulations and have been chosen within the grid of the 1024-point FFT. The generator power was −39 dBm. The centre frequency of the receiver is 200 MHz, the BB bandwidth being 25 MHz. By using two taps for the AFs, the average suppression of non-linear distortion components is 25.4 dB in Figure 4.14(a). The additional taps have been introduced, as the real receiver suffers from memory effects that are excluded in the simulations. The memory depth of $M = 1$ found in Section 2.7.3, that corresponds to two AF coefficients, has thereby been validated through measured data that has been processed by the feedforward mitigation algorithm.

The obtained suppression at different input power levels is illustrated in Figure 4.15. Compared to simulations, RF distortion is milder in measurements which explains the relatively high suppression obtained by the BB-only model. Add to this, CD will occur long before the RF distortion products become remarkable. This is due to the gain settings of

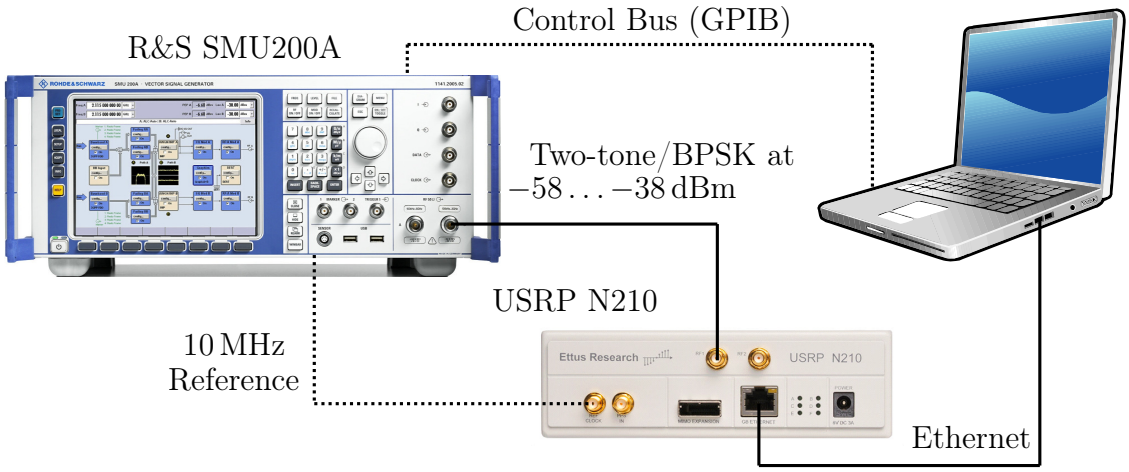
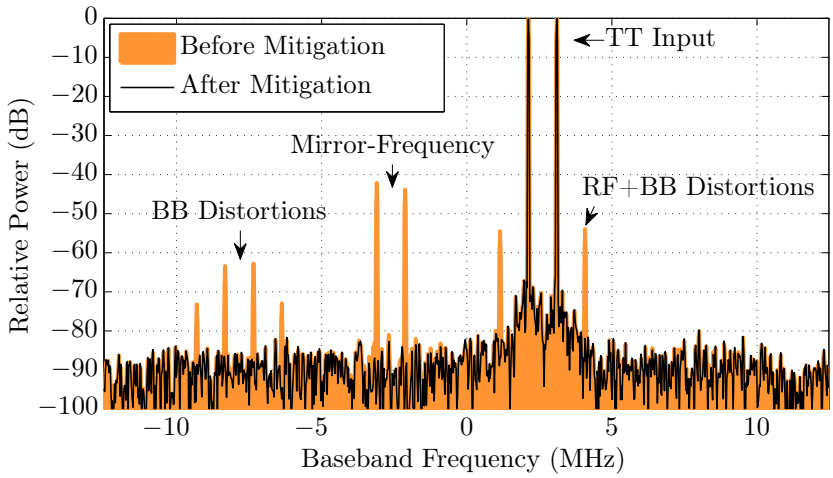


Figure 4.13: Sketch of the measurement setup [GAM+14].

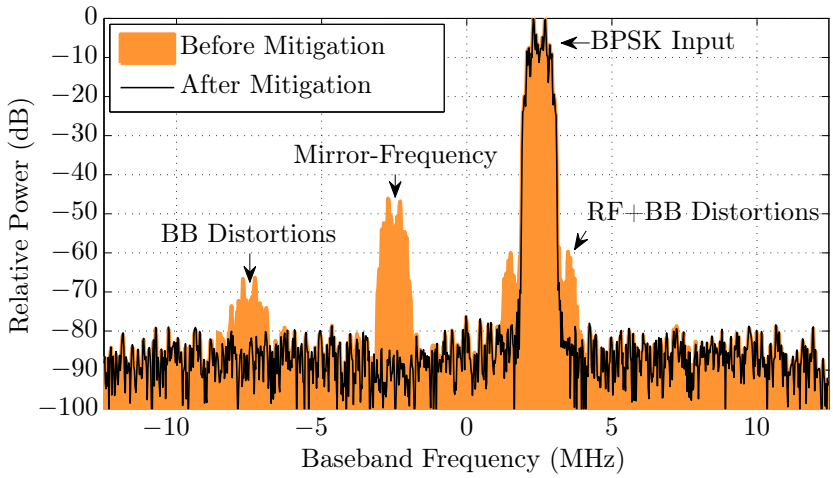
the WBX front-end that employs fixed gain at the RF and BB amplifier. Only the overall gain can be controlled through a variable attenuator at the RF input (see Appendix B). Unfortunately, the BB gain of 9 dB is much too high and cannot be easily changed to provoke more RF distortion. However, the cascaded model provides still the best mitigation performance.

Similar performance has been achieved with the BPSK-modulated signal, as shown in Figure 4.14(b). Here, the centre frequency and the BB bandwidth of the USRP were chosen to 570 MHz and 25 MHz, respectively. The BPSK signal has been generated with the same parameters as used in simulations, and the generator power and PAPR were −44 dBm and 3.35 dB, respectively. A mitigation gain of 7.1 dB is achieved using two taps for each AF.

To sum up, the co-existence of RF and BB distortions has been verified through real-world measurements, demonstrating the effectiveness of the total cascaded model. The suppression of the different distortion components is significantly better with the joint mitigation of the RF and BB stage non-linearities, compared to the previous solutions presented in the literature, which employ a model for only either of the stages. Moreover, joint mitigation of non-linear distortions and corresponding mirror components due to I/Q imbalance has been demonstrated by measurements. Thus, a complete model has been found capturing all essential distortion in the RF receiver front-end. Consequently, the digital feedforward mitigation method provides a high-performance linearisation solution for complete wideband DCR chains, enabling flexible sensing and processing of the RF spectrum in radio communication and radar devices.



(a) Two-tone scenario with −39 dBm input power.



(b) BPSK scenario with −44 dBm input power.

Figure 4.14: Mitigation results obtained with measured data [GAM+14].

4.5 Fundamental Limits

The digital feedforward mitigation principle is accompanied by some fundamental limits and challenges, which have to be considered when aiming for employing this algorithm into real devices. Basically, these limitations are imposed not only by the available technology, but also by the mitigation principle itself. These issues as well as some solution ideas are discussed in the following paragraphs. In addition, a summary like that presented in this section is not available in the state of the art, hence, it has been included in the publication [DGS+13].

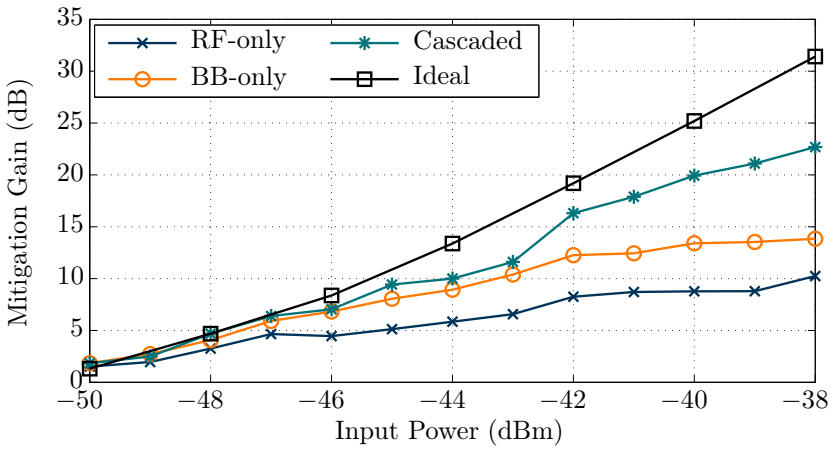


Figure 4.15: Average suppression of non-linear distortion components for different mitigation models obtained with measured data [GAM+14].

Model Mismatch significantly affects the achievable mitigation performance. First, only distortions created by the reference model can be mitigated by the AF. That is, the core distortion terms should match the physical structure of the RF front-end as close as possible to address all possible distortion. This has been shown by the joint modelling of RF and BB distortion generated in typical DCR front-ends. Otherwise, poor mitigation performance is achieved by the RF-only and BB-only models aiming for mitigating cascaded non-linearity in DCR front-ends. In such a case, adding more taps to the AF cannot compensate for severe model mismatch, since only the provided reference signal content is filtered.

Bandsplit Filters are adopted to derive the desired and reference signal for the AF stage and are therefore very important for the mitigation performance. Due to the high dynamic range conditions with weak signals and strong blockers, the filters should provide a high stopband attenuation. Thus, high filter orders and a significant amount of DSP resources may be required to implement them. However, achieving a good selectivity in the digital domain is still much easier compared to analogue filtering. For instance, their analogue counterparts suffer from component tolerances, design size and cost challenges, and thermal effects such as temperature drifts. Hence, the challenge of FIR filter is rather an implementation issue than a fundamental limit of the algorithm.

In-Band Distortion is a severe problem for the purely digital implementation of the feed-forward mitigation. It solely processes the blocker signals and subtracts them from the received desired signal. As the original transmitted blocker $x(t)$ is not known a priori at

the receiver, an estimate $\hat{x}(t)$ is filtered out of the received BB signal. However, the blocker itself does not only cause out-of-band distortion, but suffers also from in-band distortion, from the perspective of the blocker band. Thus, also the distortion estimates will differ from the original distortions contained in the desired signal branch.

The intuitive solution to that problem is to somehow obtain an ideal reference signal $\hat{x}(t) \approx x(t)$ and to generate the distortion estimates by using that signal. Finally, significant better mitigation performance and convergence characteristics can be expected. Several techniques are reasonable. For instance, the blocker waveform can be re-modulated after prior demodulation. For this, exact knowledge of the blocker waveform including its modulation, coding etc. or at least certain signal sequences need to be known. Likewise, a non-distorted version of the blocker can be gathered from a second linear RF path that employs higher attenuation to keep within the linear range. Also, digitisation requirements are relaxed as a reduced resolution auxiliary ADC may be sufficient to obtain a clean blocker signal. Similarly, an AF length can be increased to alleviate the in-band distortion problem. That is, the provided non-ideal distortion estimate is filtered until it matches the original distortion. High AF orders can shape the reference signal until it matches the desired signal. However, the other solution strategies are preferred compared to increasing the AF length, as it can result in high unnecessary filter lengths and unpredictable behaviour concerning the convergence characteristics and the step size setting.

A/D Interface is a significant bottleneck of the whole digital feedforward mitigation algorithm as it should provide a digitised version of the total BB signal including the strong distortion-producing blockers. If the distortion stems from a blocker that is outside the digitisation bandwidth, no reference signal is available for distortion regeneration. Add to this, there is the dynamic range problem when digitising a BB signal comprising several weak and strong carriers. In brief, there is a huge burden on the A/D interface of the receiver front-end concerning the ADC resolution and bandwidth. These requirements also depend on the selectivity employed at the RF input, e.g. by means of a preselection bandpass, and the actual target system specifications.

To overcome that problem, a large frequency bandwidth should by digitised as a whole by a single ADC or piecewise by parallel ADCs, each digitising a subset of the total bandwidth. However, the first approach is quite unrealistic with today's ADC technologies, especially due to the high power consumption [VSHGA+06]. However, a single wideband and high resolution ADC will likely become feasible in future, especially due to the fast growing digital circuit technology. In [KH08b; ZMS09], mixed-signal/hybrid techniques have been

proposed, that are discussed in Section 3.2. A clever solution is presented in [KH08b], where all possible distortion, that appears within the full receiver RF bandwidth, is included and only those distortion components falling into the desired bandwidth are further processed in the digital domain. However, beside the general benefits of the digital feedforward mitigation discussed in Section 3.2, no additional mismatch between the desired and reference signal is generated, compared to the mitigation architecture used in [KH08b]. Furthermore, a purely digital implementation also enables runtime reconfigurability of the mitigation algorithm, such as its FIR filter parameters or the AF step size.

Clipping Distortion (CD) cannot be mitigated by the presented algorithm due to the different nature of the distortion process. The feedforward algorithm at hand can only mitigate mild non-linear distortion where the distortion-producing blocker can be still recovered. Clipping causes severe distortion falling into the entire BB. Handling CD would require a non-clipped input signal, as discussed in Section 3.2. Such algorithms typically require hardware modifications of the RF front-end, e.g. a second RF path, and are out of scope in this thesis. Therefore, many practical RF measurements employ modulated signals with a reduced PAPR to strictly avoid CDs.

Multiple Blockers may be simultaneously present at the RF input, thus, parallelism of the digital feedforward architecture should be possible to exploit. That is, one blocker is processed by five parallel branches when employing the total cascaded model. Another blocker would require a further bunch of five branches that is arranged in parallel to the others. However, several blockers and their distortion products should not coincide to avoid the aforementioned time-domain correlation of the distortion branches. Otherwise, it results in severe convergence problems and difficult setting of the step sizes for the AFs. According to Figure 4.16, there should be a minimum frequency separation of strong blockers of at least the sum of the two blocker's bandwidth B_1+B_2 assuming dominant third-order non-linearity.

Cross-modulation effects are also neglected throughout the thesis and considered to be low, due to their dependence on the relative strength of the underlying signals [PC03; VSHGA+06]. The presented feedforward algorithm focuses on mitigation of the dominant IMD products.

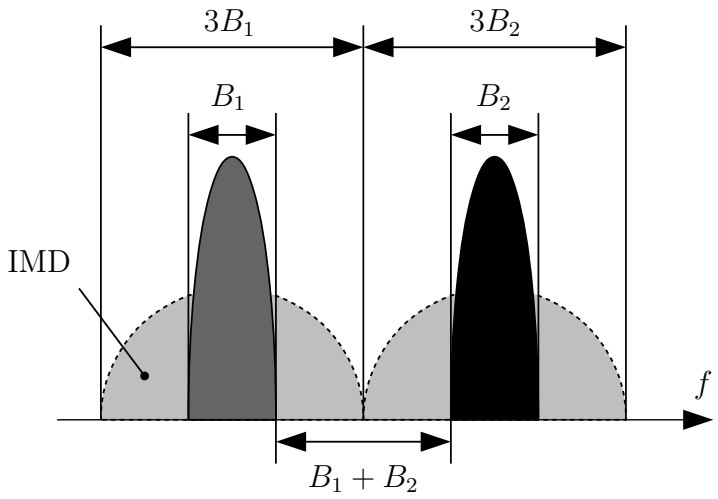


Figure 4.16: Minimum frequency separation for multiple blockers.

4.6 Inclusion of Memory Effects

As already mentioned in Section 2.5.7, the digital feedforward mitigation represents an adaptive Hammerstein model, comprising of a static reference non-linearity model followed by an adaptive FIR filter. In simulations, single-tap AFs are typically sufficient, due to the perfect match of the artificial model, mimicking the non-linear front-end, and the reference model, as well as due to the absence of memory. In measurements, at least two taps are necessary to obtain similar mitigation performance. In Section 2.7.3, it has been proven that the SDR under test, the USRP+WBX, manifests little memory. Hence, the AF stage can accomplish for linear memory effects by adding more taps. In that sense, the digital feedforward mitigation is able to cope with frequency-dependent characteristics in the RF front-end.

4.7 Summary of Digital Feedforward Mitigation

This chapter has presented the selected mitigation technique in detail. Starting from the basic concept of regenerating distortions and their adaptive subtraction in the digital domain, all necessary building blocks have been examined. In particular, a parallel mitigation architecture has been developed to mitigate cascaded non-linearities being present in DCRs with non-linear components at RF and BB. It has been shown that all generated distortion products can be suppressed significantly. The overall mitigation gain achieved with the proposed architecture outperforms state-of-the-art techniques. It is noteworthy that linear memory

effects can be considered due to the adaptive Hammerstein structure, composed of a static reference non-linearity model and the adaptive filtering stage. The effectiveness of this approach has been verified through extensive simulations and RF measurements with two-tone and modulated BPSK signals. Real-world measurements with the USRP confirm that RF and BB distortions co-exist. Finally, the proposed mitigation architecture enables a total clean-up of the BB signal. The chapter has been finished with an overview of fundamental (non-technological) limits including some solution ideas.

5 Mitigation under Fading Conditions

Up to now, the efficiency of the digital feedforward mitigation has been successfully verified in AWGN channels through simulations and RF laboratory measurements with cable connections. In principle, nothing changes in the mitigation architecture if the input signal level is varying due to a realistic radio channel manifesting all wave propagation phenomena, namely reflection, diffraction, scattering, and shadowing. In general, any fluctuation in the input signal level, also known as fading, is seen in the reference signal. However, the AF stage needs to adapt to these changes quickly. It is likely that there are bad channel conditions where the convergence characteristics and the mitigation performance of the algorithm are affected. Typically, an AGC is in charge of relaxing the fading effects and to provide a stable power level to the subsequent receiver stages. However, weak desired signals may get cut, while the AGC reduces the overall signal power depending on the blocker power level in order to prevent clipping. Employing digital feedforward mitigation has the potential to maintain and even increase the dynamic range for weak carriers while non-linear distortion effects are also alleviated. Finally, it is worth studying the impact of different fading phenomena on the level of the distortion products and the mitigation processing, in order to prove functionality of the algorithm in real-world scenarios, where an MS moves through a certain radio environment.

This problem is not considered in detail in the state of the art. Prior works in [VSHGA+06; AMV10b] just mention the role of the communication channel and suggest exploring dynamic channel effects in more detail. Time-varying waveforms occur due to channel conditions, MS movement, and user allocations. [KH08b] mentions the problem of the AF convergence time and concludes that the convergence should be faster than the minimum channel coherence time of a frame/block of the underlying standard of the desired signal.

Therefore, the influences of fading conditions and AGC on the distortion power levels and mitigation processing are investigated in more detail in the following sections. For this purpose, system-level simulations have been conducted including radio channel, receiver front-end simulation and BB processing. This work has been carried out within a master thesis [Dup13], where the interested reader is referred to for details. In addition, the results

have been published in [DGS+13]. This chapter will provide a brief overview about the methodology and the key findings.

5.1 Methodology

A typical scenario of interest is depicted in Figure 5.1, where an MS receives two adjacent FDMA channels belonging to different BTSs, a desired signal to be demodulated and a blocker signal. It is likely that there are conditions where the blocker is received with higher

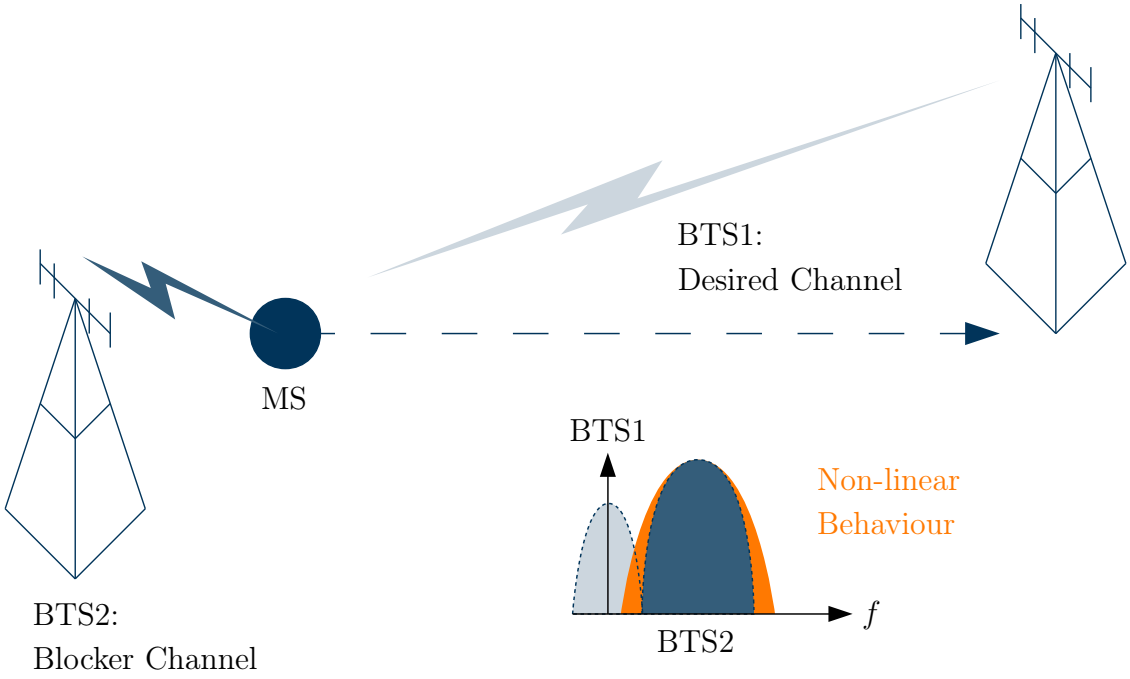


Figure 5.1: Typical fading scenario with two adjacent FDMA channels from different BTSs [Dup13].

power than the desired one, e.g. due to relative distances or unequal channel fading conditions caused by movement of the MS. Then, the blocker easily causes non-linear distortion that may fall on top of the adjacent weak desired signal, as sketched in Figure 5.1.

Such a scenario practically appears in GSM for railways (GSM-R) at the transition band between the public GSM extended band (GSM-E) and the GSM-R band at GSM 900 downlink frequencies. Details are given in Section 7.2.2. Following that practical use case, a signal configuration with a weak desired GSM carrier and a strong adjacent BPSK blocker has been assumed for the simulations with the parameters summarised in Table 5.1. The exact geometrical layout of the scenario under consideration is sketched in Figure 5.2. Assuming

Table 5.1: Detailed characteristics of the simulated signals [Dup13].

Parameter	**Desired**	**Blocker**
Modulation	GSM	BPSK
Symbol rate	256 kbps	2 Mbps
Bandwidth	200 kHz	2.7 MHz
Roll-off factor	-	0.35
Carrier frequency	0 Hz	1.55 MHz
PAPR	0 dB	5 dB
SNR	15 dB	67 dB

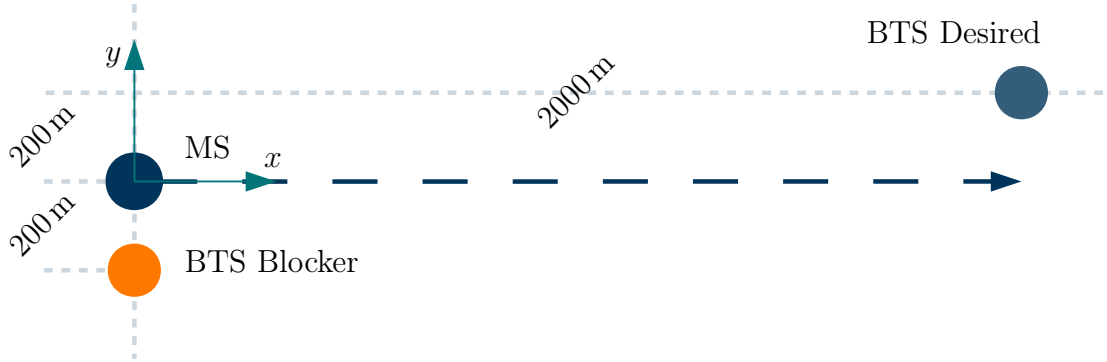


Figure 5.2: Geometrical layout of the fading scenario [Dup13].

Cartesian coordinates with the origin in the centre of the MS, the locations of the desired and blocker BTSs are $(2000, 200)$ m and $(0, -200)$ m, respectively. The MS is moving in direction x and thereby experiences different channel conditions and receives unbalanced power levels from both BTSs. Omnidirectional antennas have been adopted for all nodes as spatial effects are not of interest in the evaluation.

For the radio channel simulation, the wireless world initiative for new radio (WINNER) channel model [BESJ+05] is used. A tapped-delay line model mimics the channel in the simulator, based on coefficients extracted out of WINNER. There are different modelling techniques that can be classified into deterministic, stochastic, and geometry-based stochastic channel models (GSCMs). The deterministic model bases on Maxwell's equations or ray tracing models, which mimic reflection, scattering, and diffraction based on information of the shape and electromagnetic properties of surrounding objects. In stochastic models, the received power at different delays, Doppler shifts, and angles of arrival follow distributions derived from measurements. It provides a random number of multipath components with a random amplitude, delay, and phase. The GSCM technique

is adopted in WINNER, that randomly places scatterers and sums up all signal contributions of each scatterer. The GSCM technique is less expensive in terms of computation than conventional ray tracing and is widely accepted in industry. The model outputs the channel impulse response (CIR) at the desired sampling frequency.

WINNER offers a broad variety of propagation scenarios. The three scenarios summarised in Table 5.2 have been selected for studying the channel effects. These scenarios are mainly

Table 5.2: Selected WINNER propagation scenarios [Dup13].

Scenario	**LOS/NLOS**	**Speed** (km/h)	**Characteristic**
C3: Bad urban macro-cell	NLOS	20 60	Long delays, high frequency-selectivity
D1: Rural macro-cell	NLOS	20 60 120	Frequency-flat
D2a: Rural moving networks	LOS	20 60 120 300	Large Doppler variability, frequency-selective

distinguished by their line of sight (LOS) or non-line of sight (NLOS) characteristic, i.e. if there is dominant multipath propagation or not. Scenario C3 causes high time variation and is assumed to stress the algorithm due to rapid changing regrowth from snapshot to snapshot. Scenario D1 possess a low selectivity in frequency and low time variation and is likely to work without trouble. It constitutes a counterpart to the other scenarios. Large Doppler variability and high frequency-selectivity is caused in scenario D2a for representing moving networks, that should stress the convergence time of the AF stages significantly.

The DCR front-end is modelled by a static memoryless polynomial for modelling RF nonlinearity that is assumed to be dominating. Memory effects are not considered in this simulation, hence, single-tap AFs are sufficient. First, the impact of the channel on the spectral regrowth of the blocker is studied. Then, the convergence characteristics of the AF stage is investigated under varying blocker power conditions from snapshot to snapshot. Finally, the performance of the algorithm is verified and reasons for failures are given.

5.2 Impact of Fading and AGC

In general, the typical wave propagation phenomena lead to multipath (NLOS) propagation, where different delayed versions of the received signals superimpose. In addition, there is a frequency-dependent behaviour of the channel effects. Hence, multiple types of fading are distinguished, that are sketched in Figure 5.3. Large-scale fading denotes the path loss due

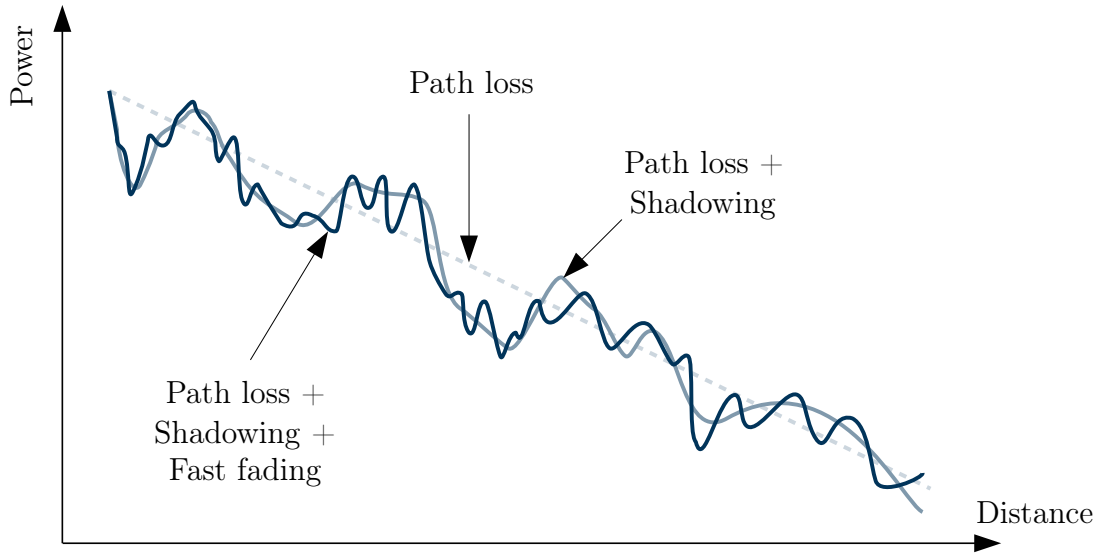


Figure 5.3: Qualitative behaviour of different fading phenomena [Dup13].

to the relative distance and shadowing of large objects like buildings. The MS moves through a *large* distance in relation to the BTS. Small-scale fading comprises fast variations on the amplitude of the received signal, mainly due to constructive and destructive superposition of multipath components. Here, the MS moves in *small* distances. Furthermore, frequency-flat and frequency-selective fading is distinguished, depending on the number of resolvable delays. In that context, the coherence bandwidth denotes the width of frequencies of the signal that are similarly affected by the channel. Finally, also Doppler spread occurs due to the movement of the MS. Depending on the speed, the Doppler spread causes slow or fast fading. In particular, in fast fading, the coherence time of the channel is smaller than the symbol period and causes a broadening of the original signal in time domain.

Due to these fading conditions, the non-linear distortion will vary from snapshot to snapshot in a non-linear manner. Also, signals from different BTSs experience different channel effects and sum up in the receiver. Thereby, highly unbalanced signal configurations with weak and strong signals may occur, especially in high frequency-selective channels, where signals can be even temporarily cancelled. Finally, the algorithm has to track the changes in these signals introduced by the channel characteristics, which requires a certain convergence time of the

AF stage. However, large scale fading is omitted from the analysis as it is obviously not a challenge for the algorithm to track these slow variations. Thus, the impact of small-scale fading is on focus in the simulations.

Reducing the channel effects and smoothing deep fades is typically the task of an AGC. It adapts the input power level to predefined levels to keep the components within their linear region, by comparing the average signal power to a reference power and adjusting a voltage-controlled amplifier at the AGC loop front. There are feedback and feedforward implementations of AGCs similar to the linearisation techniques discussed in Section 3.1.1. Basically, the feedback implementation is preferred due to its closed-loop regulation, although there is a longer delay introduced due to the decision. Receivers typically employ a stepped AGC at RF and a second AGC at IF or BB. A general drawback of the AGC is that the whole signal gets attenuated with regard to the high power signals. Hence, the overall dynamic range is reduced and weak signals may get cut, and thus covered by noise.

5.2.1 Simulation Setup

The simulation setup is depicted in Figure 5.4. The channel changes every snapshot having a

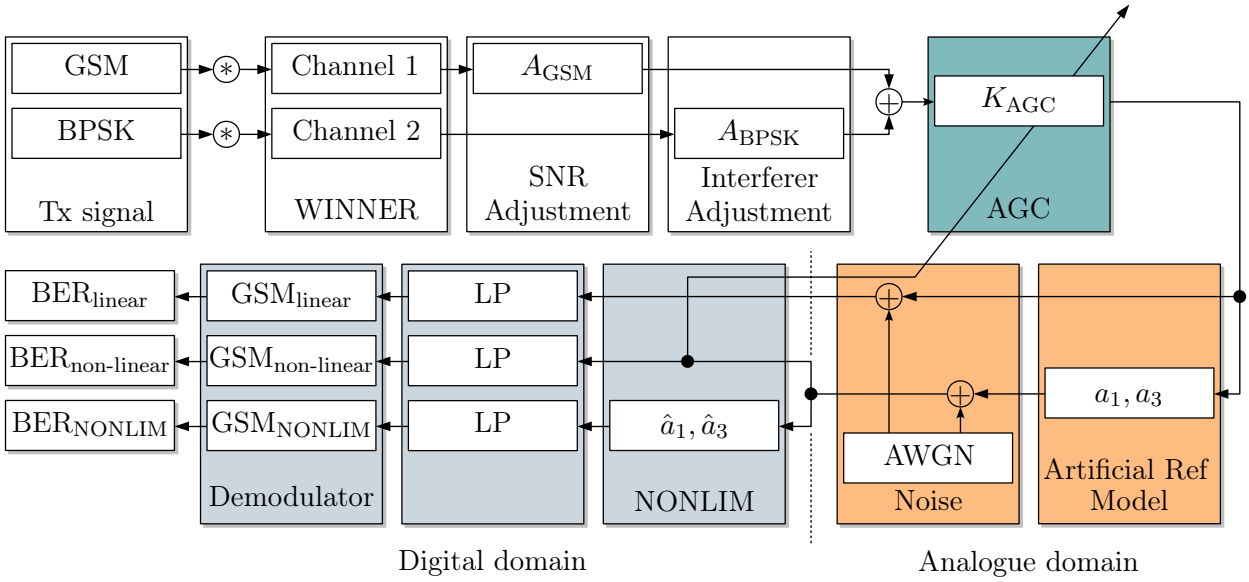


Figure 5.4: Block diagram of the fading simulation setup [Dup13].

length of a single GSM burst. Then, both blocker and desired GSM carrier are convolved with the individual CIRs ($*$). For this purpose, the channel coefficients are normalised to eliminate the influence of path loss and shadowing. Referring to Table 5.1, the SNR of the GSM is kept fixed at 15 dB to avoid bit errors due to channel effects and noise. Bit errors should only

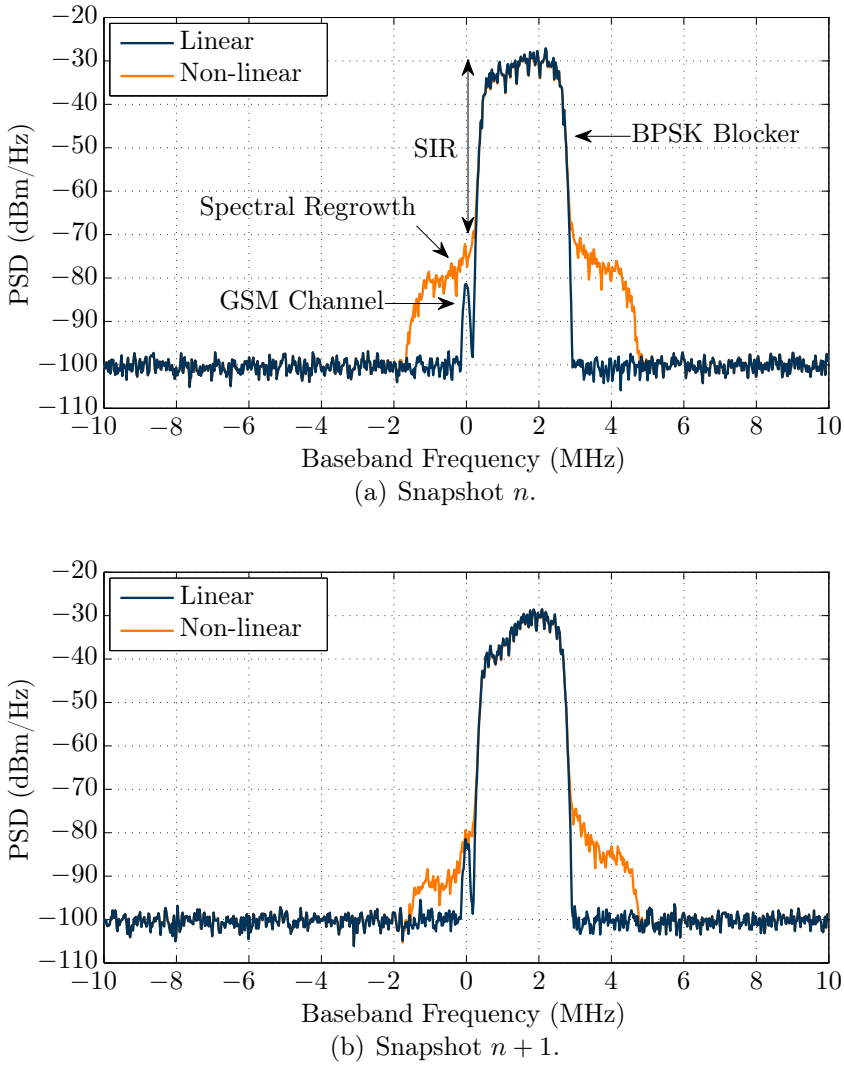
occur due to non-linearly induced interference. The power of the BPSK is set 52 dB higher than that of the GSM. The AGC averages the power over 75 samples which corresponds to 1 bit of a GSM burst (oversampling rate of 75). Then, the composite signal is fed through the artificial reference model with $a_1 = 1, a_3 = -200$ (equals to an IIP3 $= -7$ dBm) and an AWGN source that jointly mimic the non-linear DCR in the simulator. Before demodulation, the signal runs through a lowpass (LP) filter. The burst-wise BER analysis of the desired GSM carrier is conducted with the GSMSim simulation framework [EM97], that is also used and detailed in Section 7.2.2. In brief, it is an easy to use implementation of a transmitter and receiver front-end of a standard compatible GSM system. In total, 200 channel realisations are simulated, with 100 GSM frames per realisation.

5.2.2 Effects of Fading on the Non-linearities

Frequency-selective and fast fading conditions are most crucial operating conditions for the mitigation algorithm. In a frequency-selective channel, both blocker and desired signal are shaped from snapshot to snapshot in a different way, thereby changing the signal-to-interference ratio (SIR), here being defined as the ratio between the desired signal power level and the level of spectral regrowth of the blocker falling into the desired channel. The speed of the fast fading will be important for the convergence time of the algorithm as it determines the SIR variation rate. The algorithm should follow and mitigate these variations as fast as possible while keeping stability and providing sufficient mitigation gain.

Figure 5.5 illustrates an example of two consecutive snapshots that experience very different frequency responses due to a heavily frequency-selective channel. The wider the bandwidth of the blocker, the more harmful is its regrowth. From Figure 5.5(a) it is clearly visible that the weak GSM carrier is completely covered by the spectral regrowth induced by the BPSK blocker, resulting in a high BER. In addition, the channel shapes the blocker spectrum in a non-symmetric way, resulting in different levels for the upper and lower IMD.

Next, the statistics of the spectral regrowth are analysed as part of non-linear distortion is falling on top of the desired band and that is being processed by NONLIM. Especially the difference from snapshot to snapshot is expected to stress the convergence of the AF stages. In [Dup13], it has been shown that the spectral regrowth follows a generalized extreme value distribution, whereas the ΔSpectral_Regrowth is fitted best by a t location-scale distribution. By taking the desired signal power and the amount of spectral regrowth, the SIR and the corresponding ΔSIR from snapshot to snapshot are obtained. Figure 5.6



(a) Snapshot n.

(b) Snapshot $n+1$.

Figure 5.5: Impact of frequency-selective channel on SIR [Dup13].

illustrates the distribution of ΔSIR for the three different scenarios for a speed of 60 km/h. D1 shows only little variance of the SIR and looks almost impulse-like, mainly because of the low frequency-selectivity of the channel. On the contrary, D2a and C3 cause the high variance due to the frequency-selectivity. For D2a, this variance will increase with the speed of the MS.

5.2.3 Mitigation Performance under Fading Conditions

Now, the performance of the feedforward mitigation is evaluated by means of the SIR and BER improvement. Figure 5.7 shows the mitigation gain for the scenario C3, where each point represents one snapshot. In total, 20 000 snapshots have been simulated. Basically,

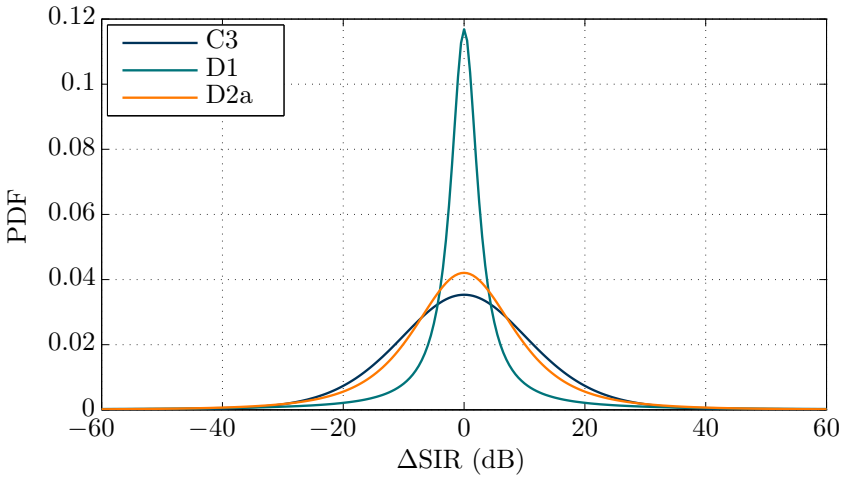


Figure 5.6: PDF of ΔSIR in different fading scenarios [Dup13].

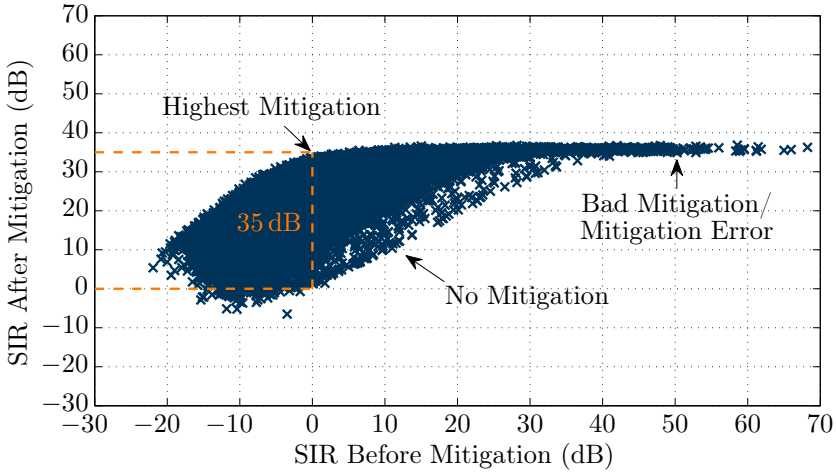


Figure 5.7: SIR enhancement analysis for the scenario C3 at 60 km/h [Dup13].

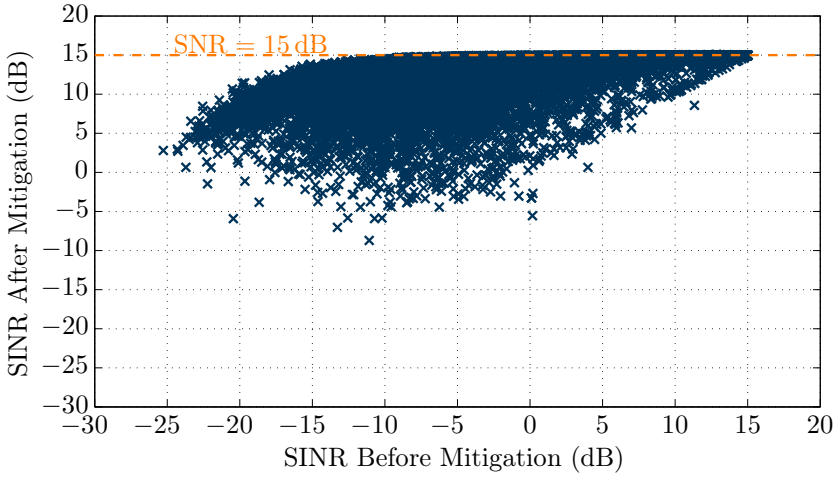


Figure 5.8: SINR enhancement analysis for the scenario C3 at 60 km/h [Dup13].

the mitigation gain is not constant and varies from snapshot to snapshot. The variation is mainly caused by the interference power, i.e. the amount of spectral regrowth in the desired band, as the desired signal power has been kept constant. The highest mitigation gain is approx. 35 dB with regard to the noise floor as the algorithm will not mitigate noise. There are cases without improvement of the SIR (no mitigation) due to convergence problems of the AF. Bad mitigation occurs for high SIR conditions, i.e. the power of the interference is even higher after applying the algorithm. Here, the algorithm introduces noise to the desired signal output, thereby increasing the level of interference. In such a case, the AF cannot find the correlating reference distortions in the desired signal branch, as the blocker has been received without distortions. Thus, the AF will not converge and adds mitigation noise to the output. To overcome this problem, the algorithm should be bypassed under high SIR conditions, when there is no need for mitigation in general.

The interference mitigation results also in an SINR improvement as illustrated in Figure 5.8. The upper limit is determined by the SNR = 15 dB. Hence, the ideal mitigation is denoted by a straight line at 15 dB, however, the algorithm does not always fully cancel the interference, which causes some variance.

Finally, the resulting BER after applying the mitigation algorithm is analysed. Three main performance cases can be distinguished

(i) Mitigation $\mathrm{BER}_{\mathrm{NONLIM}} < \mathrm{BER}_{\text{non-linear}}$,

(ii) No need for/ no mitigation $\mathrm{BER}_{\mathrm{NONLIM}} = \mathrm{BER}_{\text{non-linear}} = \mathrm{BER}_{\mathrm{linear}}$ or $\mathrm{BER}_{\mathrm{NONLIM}} = \mathrm{BER}_{\text{non-linear}} > \mathrm{BER}_{\mathrm{linear}}$,

(iii) Bad mitigation $\mathrm{BER}_{\mathrm{NONLIM}} > \mathrm{BER}_{\text{non-linear}}$.

Following Figure 5.6, scenario D1 is likely to provide the highest mitigation performance due to the low variation of the SIR. On the other hand, C3 and D2a often do not require mitigation due to the high SIR variation.

A figure of merit, the bit error mitigation performance (BEMP) is introduced with

$$\mathrm{BEMP} = 1 - \frac{E_{\mathrm{NONLIM}} - E_{\mathrm{linear}}}{E_{\text{non-linear}} - E_{\mathrm{linear}}}, \tag{5.1}$$

where E denotes the amount of errors per burst in the corresponding cases indicated by the subscripts. Note that there is a certain minimum error E_{linear} due to the noise and fading, the mitigation algorithm is not supposed to mitigate. A BEMP value near 1 indicates the best performance, whereas BEMP $\ll 1$ denotes a bad performance. From Figure 5.9 it can be concluded that in 90 % of all simulated snapshots, the BEMP is higher than 90 %,

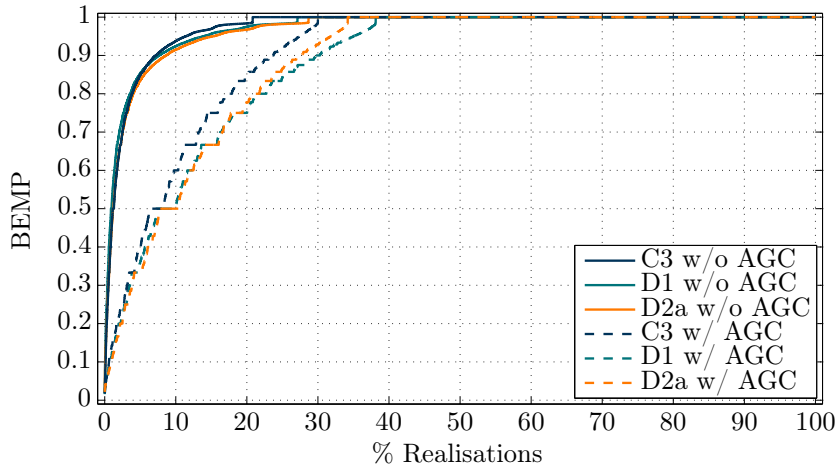


Figure 5.9: BEMP analysis for different scenarios at 60 km/h [Dup13].

independently of the scenario. When having the AGC enabled, the algorithm operates at high SIR regime and the AF coefficients converge more slowly. Finally, in only 75 % of the realisations the BEMP is higher than 90 %. In [Dup13] it is shown that the bad mitigation cases occur due to convergence problems caused by high ΔSpectral_Regrowth and where the algorithm is working in high SIR regime.

More interestingly is the actual BER enhancement evaluated on a burst-based analysis. For that purpose, the SINRs have been extracted from the simulations, re-ordered, and displayed with their corresponding BER. Finally, the mean BER per burst is computed, even though, there is not an equal number of realisations per SINR value available. Figure 5.10 shows a significant improvement of the BER performance independently from the scenario for SINR < 0. The lowest mitigation is achieved for C3 in the high SINR regime. Figure 5.11

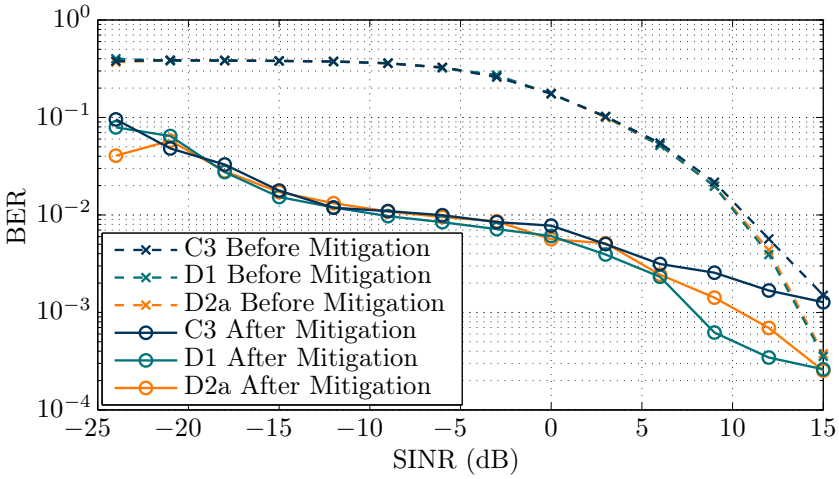


Figure 5.10: BER vs. SINR at different scenarios [Dup13].

illustrates the performance with AGC inclusion at the example of scenario C3. If the AGC

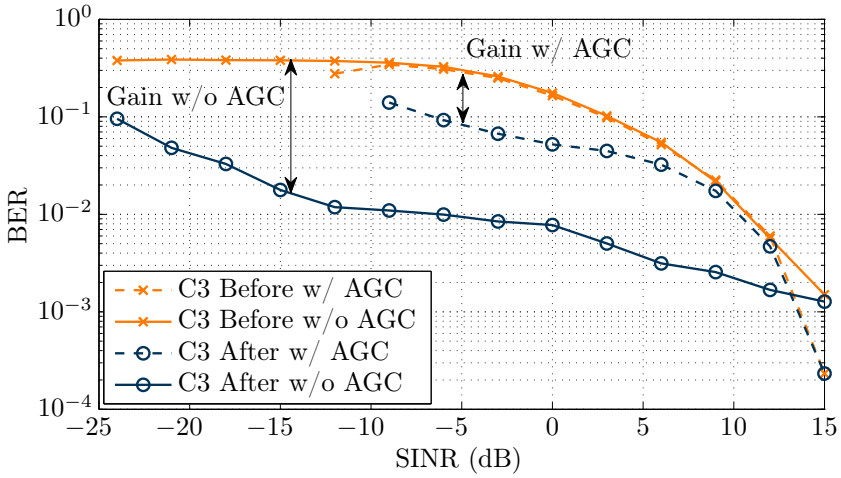


Figure 5.11: BER vs. SINR for scenario C3 with and without AGC [Dup13].

is enabled, the mitigation performance is reduced significantly. Although the AGC provides approx. 10 dB of SINR gain, the BER is higher compared to the case without AGC. This is due to the attenuation of the desired signal, which is affected significantly by noise.

To sum up, the radio channel causes a certain shaping of the blocker signal according to the characteristics of the scenario. In the statistical analysis of the behaviour of the NONLIM algorithm under random channels, it has been shown that there is not a significant performance degradation among the scenarios, although there is a dependence on the speed [Dup13]. In brief, the algorithm is applicable in receivers that operate in a real-world radio environment. However, the best performance is achieved in scenario D1 due to the less frequency-selective, low-variance interference. In addition, it has been found that the AGC reduces the overall system performance, although it prevents IMD generation. By disabling or relaxing the AGC under the presence of high-power signals/blockers, the dynamic range can be improved with the help of the feedforward algorithm. Furthermore, the algorithm should be bypassed under high SIR conditions.

The bypassing needs to be switched according to the interference power at hand. For this purpose, the input power can be gathered from the desired branch to estimate the significance of non-linear distortion. Another solution can be to perform a cross-correlation between the desired and reference branch to determine the amount of non-linear distortion present. Then, the algorithm can be enabled or disabled according to a threshold.

5.3 Summary about the Impact of the Radio Channel

This chapter has discussed an important practical issue when aiming for integration of the digital feedforward mitigation into real devices. It has been confirmed that the algorithm can be applied even under frequency-selective and fast fading conditions imposed by wave propagation in a real radio environment. In particular, it has been shown that the BER of distorted signals, as well as the energy of non-linearly induced interference decrease significantly after applying the algorithm. However, bypassing the algorithm under high SIR conditions seems to be inevitable to overcome convergence difficulties and the addition of noise. Eventually, the dynamic range and effective linearity of the RF front-end can be increased solely by the feedforward techniques with disabled or relaxed AGC.

6 Real-time Implementation

In the prior chapters, the feedforward mitigation capabilities have been demonstrated based upon offline post-processing of simulated and measured data in MATLAB. Practical aspects of embedding the DSP-based algorithm into the device's digital back-end are rarely discussed in the state of the art. Although the complexity of the algorithm has been analysed on a high level in Section 4.3, it is desirable to build an FPGA prototype of the purely digital feedforward mitigation to obtain further insights into practical challenges. An FPGA is well suited for this application due to its rapid prototyping.

Embedding the algorithm into the device offers tremendous advantages. The receiver self-regulates the induced non-linear distortion and thereby improves its effective linearity. In addition, the host PC is offloaded and can solely perform the specific BB processing since the digital back-end transfers the corrected I/Q data directly to the host.

There is no implementation of the purely digital feedforward mitigation algorithm available in the state of the art. The only prior implementation of the feedforward mitigation algorithm is reported in [KH08b], where the distortion regeneration takes place at RF level. This kind of mixed-signal implementation, that has been discussed in Section 3.2, requires a custom hardware design with considerable effort. General implementation of an LMS AF on FPGAs can be found in [MB04]. An optimisation of the LMS tap-weight update by simple bit-shift operations is proposed in [OKA06].

In this chapter, a brief overview of a purely digital implementation on the FPGA of the USRP N210 [Ett] is given. This work has been carried out within a master thesis [Sch13], to which the interested reader is referred for more details. In addition, the results have been published in [SGT13].

6.1 Target Architecture and Basic Scenario

The design is implemented on the FPGA Xilinx Spartan-3A DSP (XC3SD3400A) of the USRP N210 motherboard [Ett; Xilb]. An FPGA is an integrated circuit that

holds programmable logic. For implementing digital circuits on an FPGA, a high-level description of the circuit is mapped to the FPGA architecture with the help of hardware description languages (HDLs) (very high speed integrated circuit HDL and Verilog) and software tools for electronic design automation (EDA) [Sch13]. FPGAs are very popular due to their flexibility and very high speed supporting parallelism of algorithms. The digital circuit is implemented based on configurable logic blocks (CLBs), which comprise look-up tables (LUTs), flip-flops (FFs), multiplexers (MUXs), and carry/arithmetic logic. Beside this general FPGA fabric, the target architecture offers dedicated memory (Block random-access memory (RAM)) and DSP48A slices that are well suited for DSP applications [Sch13]. A DSP48A slice is composed of a pre-adder, a signed 18×18 bit multiplier, and a post-adder. These slices can implement multiply-accumulate (MAC) architectures such as FIR filters, where a multiplication is followed by an addition.

The number representation is a challenge for implementing DSP-based algorithms on an FPGA. A fixed-point or floating-point arithmetic with a certain bit width can be implemented. However, a fixed-point implementation is more common due to its increased efficiency. Nevertheless, the designer has to deal with bit growth, overflow, and scaling/ rounding methods. For example, if two numbers are added, the word length needs to be increased by 1 bit in order to avoid potential overflow. Likewise, multiplying N and P wide numbers result in a bit width of $N + P$ for maintaining full precision. Overflows need to be strictly avoided, as the number can wrap around from positive to negative or vice versa and may cause abrupt jumps in the waveform. Consequently, the result of any arithmetic operation needs to be rounded and truncated. Beside other techniques, non-symmetric rounding to positive is applied for this design, that adds a binary 1 to the rightmost bit position after truncation and right-shifting by k bit positions (scaling by 2^{-k}) [Sch13].

Basic Scenario and Simplifications

The basic scenario and the block diagram for the real-time implementation is illustrated in Figure 6.1, similar to Figure 4.1. In the sketched spectrum ①, a simple two-tone blocker is assumed, whose IMD products may hit an adjacent target frequency band. Hence, a bandstop and a bandpass filter form a bandsplit filtering stage assuming fixed filter characteristics. Thereby, the runtime reconfigurability of the filters on the FPGA is omitted for the sake of simplicity. As a further simplification, only a third-order RF non-linearity model is implemented as the RF non-linearity, or the distortion in the fundamental zone, is assumed to be dominating. Further distortion at BB is not taken into account to keep the model as simple as possible. Moreover, a separated I/Q processing of the distortion

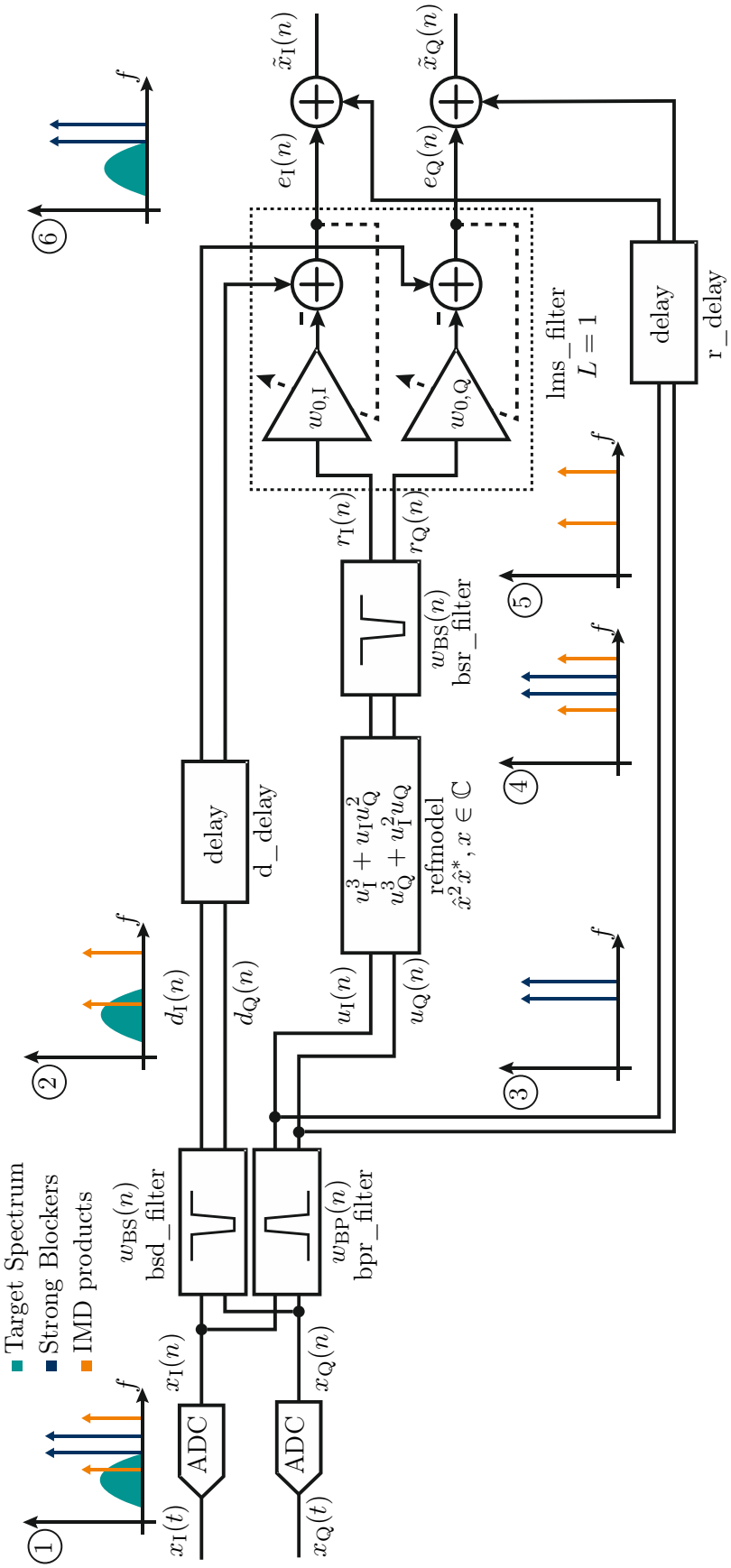


Figure 6.1: Block diagram of the NONLIM FPGA implementation by means of a two-tone example [Sch13].

estimates is performed in the reference branch, that shows already sufficient mitigation in practice [GSH+12a].

The architecture is composed of the FPGA module instances bandstop-desired filter (*bsd_filter*), bandpass-reference filter (*bpr_filter*), bandstop-reference filter (*bsr_filter*), reference model (refmodel), and the LMS AF (lms_filter). The band-split filter separates the blocker and the useful signal to ② and ③, whereas a complex implementation for the *bpr_filter* is chosen due to the mirror-frequency interference being present in real-world measured signals (cp. Section 4.4). The reference model computes the RF non-linearity term x^2x^* according to (2.22) and outputs a signal such as ④. For separated I/Q processing, this term is decomposed into (2.23) and (2.24). The exact coefficient $3a_2$ in (2.22) is omitted without affecting the performance of the algorithm, as the AF stage will find the correct value any way. Next, the distortion estimates are fed through a real-valued bandstop filter (*bsr_filter*) to suppress the blocker and to obtain only the distortion. An equivalent filter is employed in the desired branch, which has an inverse characteristic to the *bpr_filter*. Then, the output signal ⑤ enters the AF stage that is depicted with a filter order $M = 0$ ($L = M + 1 = 1$). That is, only scaling of the estimates is conducted in the ideal case having a perfect match of the reference non-linearity and the real-world non-linear behaviour of the front-end. Here, a real-valued LMS implementation is chosen for reasons of simplification. A complex-valued LMS implementation might outperform the real-valued one due to the capability to mitigate AM/PM distortion and I/Q imbalance effects that are present in real-world signals. Alternatively, the AF length could be increased to compensate for the missing cross-products between I and Q. Finally, as depicted in ⑥, the blocker may be added back in order to satisfy a spectral sensing use case (Section 4.1).

There are certain processing delays introduced by NONLIM that need to be compensated by delay blocks (*d_ delay* and *r_ delay*). There is a pure processing delay of the processing chain as well as the settling time of the FIR filters (length of $M/2$). The signals of the desired and reference branch need to be aligned to provide the AF with corresponding samples. The delay block *d_ delay* accounts for the relative delay between the desired and reference branch. Another delay block *r_ delay* compensates for the absolute delay of the reference branch to synchronously add the blocker to the AF error signal.

6.2 Modules Implementation

Figure 6.2 illustrates the existing direct-down conversion (DDC) chain of the USRP with the integrated NONLIM FPGA module. Basically, even in a modern SDR, all the high-speed

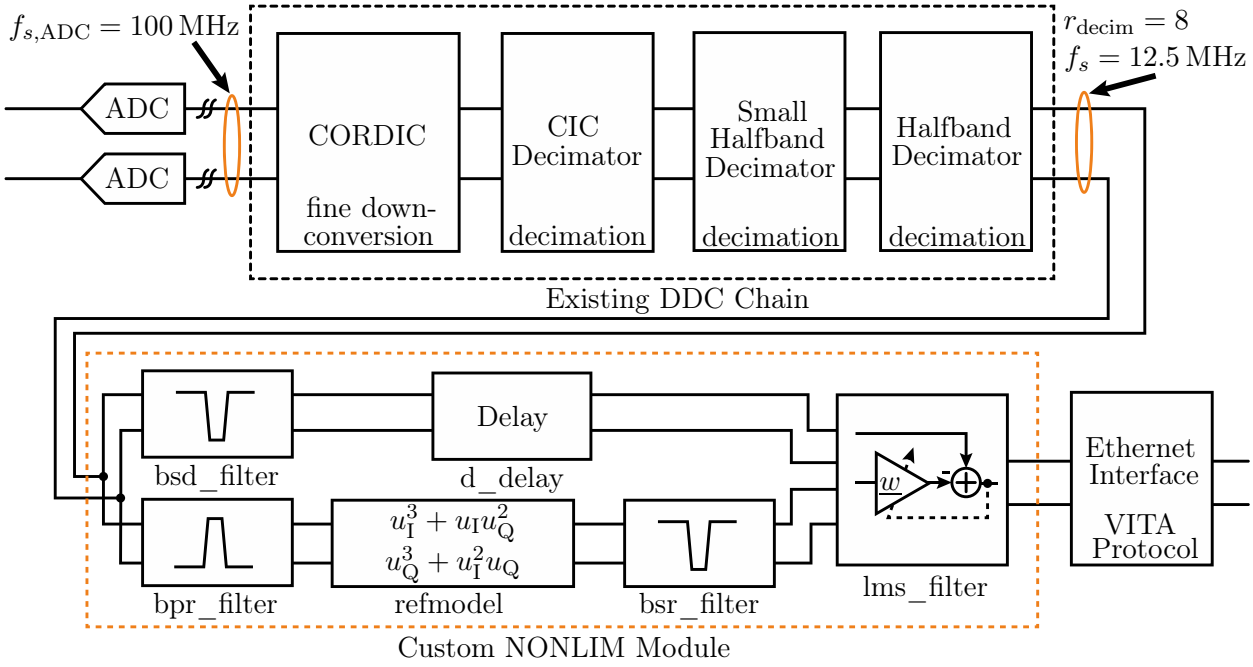


Figure 6.2: Existing DDC chain with the integrated NONLIM module [Sch13].

processing tasks such as decimation, filtering, and interpolation still remain on the FPGA. The main task of the existing DDC chain is to perform frequency conversion to zero-IF and decimation, such that the samples can get transferred to the PC. The ADC clock with 100 MHz is much too high to directly stream the data to a PC in real-time, i.e. a decimation by at least 4 needs to be performed in order to transfer them via gigabit ethernet (GigE) [Ett; Sch13].

The coordinate rotation digital computer (CORDIC) denotes an iterative algorithm performing approximated computations of trigonometric functions to form a NCO that conducts the final down-conversion. The cascaded integrator comb (CIC) is a hardware-efficient implementation for decimation and interpolation filters by using only additions and subtractions. However, the CIC causes aliasing components and a roll-off in the passband that makes subsequent filtering essential. In order to compensate for the CIC roll-off, a two-stage halfband filter is employed with symmetric FIRs with every second coefficient being zero. This allows for using the hardware multipliers in a time-multiplexed manner. In order to efficiently compensate for the CIC roll-off, a decimation rate as a multiple of 4 should be chosen [Sch13]. Only in the case this constraint is met, both halfband filters are activated. To relax the requirements of the real-time processing of the samples in the custom NONLIM module, a decimation rate of $r_{\mathrm{decim}} = 8$ is used. Hence, the custom module is placed after the decimation as indicated in Figure 6.2. This allows for a processing time of 8 clock cycles until the next sample appears at the input. Also, BB distortion is likely to be filtered out if

the blocker IF frequency is chosen to be slightly greater than $f_s/6$ (cp. Section 4.2). Thus, distortion in the fundamental zone are on focus in this implementation.

The HDL Verilog has been used for the implementation, as the existing USRP FPGA firmware is already written in Verilog [Ett]. To verify the correct functionality of each sub-module, a top-down design flow is adopted [Sch13]. Starting from a fixed-point high-level simulation in MATLAB, an HDL-MATLAB co-simulation is performed by using a custom file interface and the ISim behavioural simulator that is included in the Xilinx EDA tool ISE. Finally, the modules are validated running on real hardware. The following subsections provide further details of the implementation of the main sub-modules.

FIR Filter

There are three instances of FIR filters, namely *bsd_filter*, *bpr_filter* and *bsr_filter*, as summarised in Table 6.1. These are designed with the given characteristics using the HDL-

Table 6.1: FIR filter implementation for a two-tone blocker centred at $f_{\mathrm{IF}} = 3\,\mathrm{MHz}$ [Sch13].

Filter instance	**BW**	$\mathbf{f}_{\mathrm{cutoff1}}$	$\mathbf{f}_{\mathrm{cutoff2}}$	$\mathbf{A}_{\mathrm{stop}}/\mathbf{A}_{\mathrm{pass1/2}}$	**Filter order** M
bsd_filter	640 kHz	320 kHz	600 kHz	85/0.2 dB	161 (M_{bsd})
bpr_filter				40/0.2 dB	95 (M_{bpr})
bsr_filter					95 (M_{bsr})

MATLAB interface and the Parks-McClellan algorithm (Section 4.2). The implementation is established using the intellectual property core generator of the Xilinx EDA tool, that efficiently map a pre-defined module with a specific function to the target device.

Basically, there are two main possible implementation structures: MAC (symmetric systolic FIR) and distributed arithmetic (DA). The MAC structure uses multiplications and additions, as illustrated in Figure 6.3, that are mainly implemented by concatenated DSP48A slices. This type of implementation is used for the *bsd_filter* and *bsr_filter*. On the contrary, the DA approach substitutes the multiplications by bit shifts, additions, subtractions, and LUTs. It is used for the *bpr_filter* to save DSP48A slices at the expense of general FPGA fabric (CLBs). A complex-valued implementation has been chosen for the *bpr_filter*. Although resulting in twice as much resources, a non-symmetric filtering characteristic at both positive and negative frequencies cannot be tolerated with focus to the mitigation performance. Basically, a very pure reference signal is desirable for the reference branch to

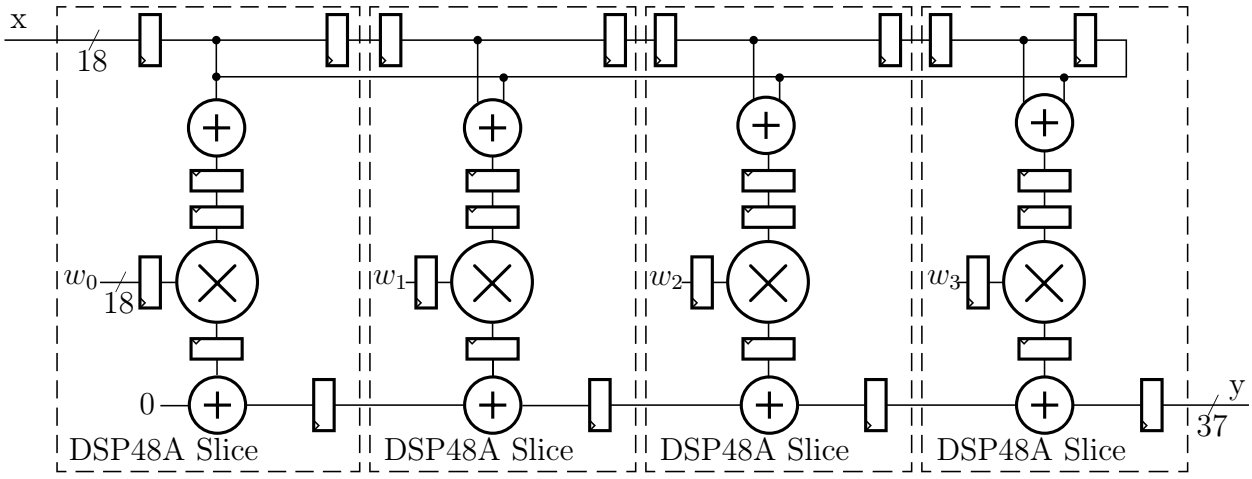


Figure 6.3: Block diagram of a symmetric systolic FIR implementation [Sch13].

make the AF stage work efficiently. Symmetric (real-valued) filtering is, on the other hand, acceptable for other bandstop filters with regard to the more efficient resource utilisation.

A general challenge of FIR filter implementation is the quantisation of the filter coefficients using a fixed-point number format. It results in small deviations in the frequency response characteristics compared to their floating-point counterparts. In [Sch13], these deviations have been analysed and reported to be negligible for the chosen number format and filter characteristics.

Reference Model

The block diagram for implementing the reference non-linearity, illustrated in Figure 6.4, exactly follows (2.23) and (2.24), where the initial value $3a_2$ is omitted to save another multiplier. Basically, the implementation performs the multiplications and additions sequentially. In order to keep the results within the given bit width, a non-symmetric rounding to positive and a subsequent bit truncation of the least significant bits (LSBs) is established. In total, the implementation of the reference model requires 6 DSP48A slices.

LMS Filter

Regarding the LMS filter, a real-valued implementation is chosen that is decomposed into an FIR and coefficient update part, as depicted in Figure 6.5. The FIR is implemented by L parallel multipliers and is followed by a hard truncation of SF bits. Here, no rounding is performed as it is expected that the AF stage accounts for the imposed bias. A similar

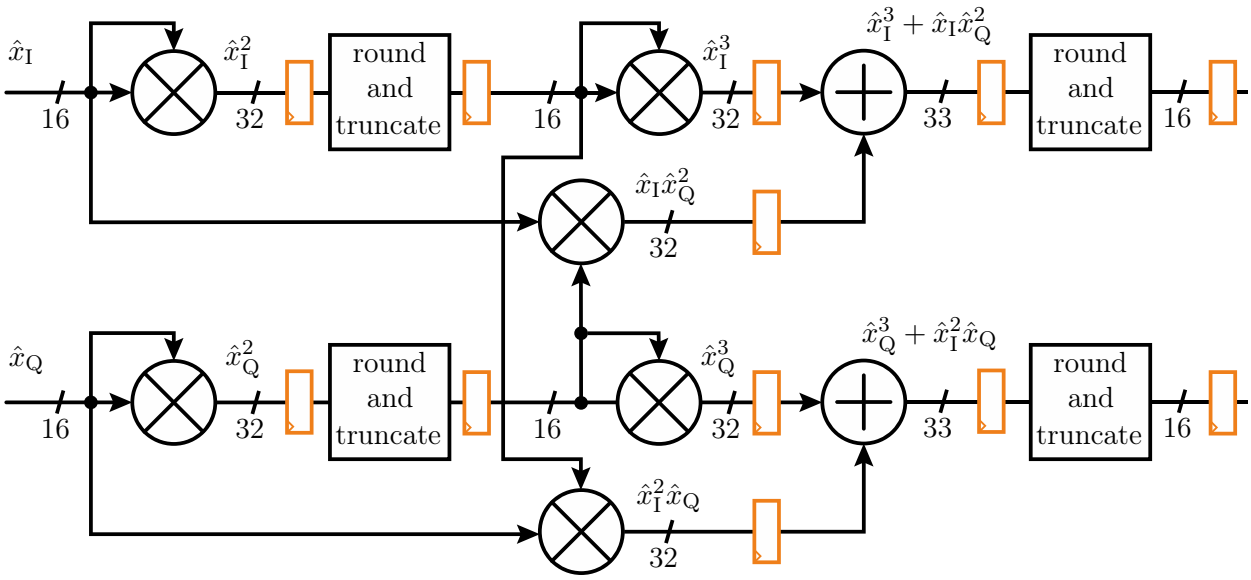


Figure 6.4: Block diagram of the pipelined implementation of refmodel [Sch13].

structure is adopted for the coefficient update part with a bit truncation factor MU. For that purpose, the approach of [OKA06] is followed that approximates the multiplication with the step size by a simple bit shift operation. In [Sch13], rules for choosing the truncation factors SF and MU are given. They need to be adapted during runtime and may cause overflow conditions in case they have been chosen to be too low. For example, both values SF and MU need to be as small as possible and have to obey the rule $SF + MU \geq 26$ [OKA06], in order to achieve a low residual error. For the scenario of interest, a setup with $L = 4$ provided the best mitigation performance. That is, the DSP48A slices are used in a parallel manner to perform the entire algorithm within 8 clock cycles.

On the other hand, a complex-valued LMS architecture can be used to account for I/Q imbalance and AM/PM distortion. That is, lower AF order and therefore less resource utilisation can be achieved. Table 6.2 compares the real- and complex-valued structure in terms of their resource utilisation. Indeed, when setting the AF order $L = 2$, 4 DSP slices

Table 6.2: DSP48A Utilisation of the real- and complex-valued LMS implementation [Sch13].

L	**Real-valued**	**Complex-valued**
In general	$2L + 2L = 4L$	$4L + 2L = 6L$
1	4	6
2	8	**12**
3	12	18
4	**16**	24

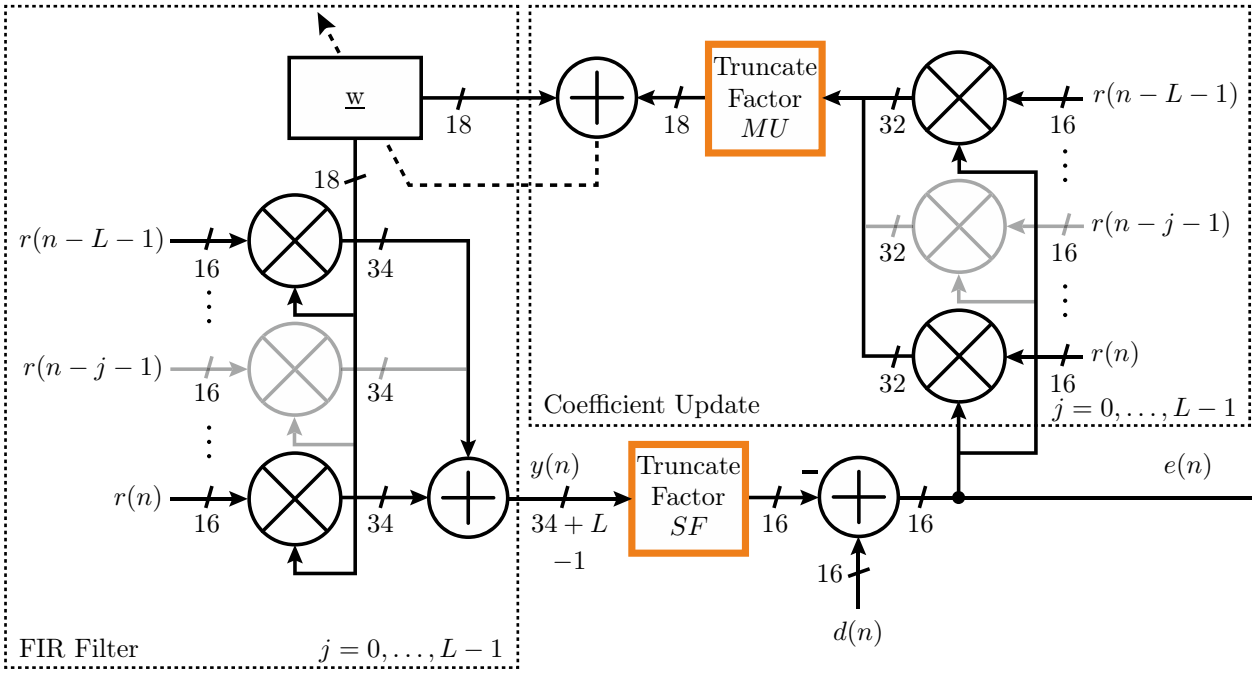


Figure 6.5: Implementation structure of the LMS filter [Sch13].

could be saved with a complex-valued implementation.

Delay Blocks

There are two instances of delay blocks, *d_delay* in the desired branch and *r_delay* for adding back the blocker to the final AF error signal. Both are implemented by block RAM following the first in – first out principle.

Design Report

The resource utilisation of each FPGA module is detailed in Table 6.3, though fabric for interconnection of the modules is not considered here. Table 6.4 summarises the amount of used resources with and without the custom NONLIM module. Although there is a large amount of resources available, timing violations turn out to be the challenging problem in practical FPGA implementations when approaching a high resource utilisation of the total available resources. Fortunately, all timing constraints have been met with the aforementioned design.

Another figure of merit is the introduced processing delay of the NONLIM module, which is of high interest in time division multiple access (TDMA)-based communication systems like

Table 6.3: Resource utilisation of the individual sub-modules [Sch13].

Type\Module	***bsd_filter***	***bpr_filter***	***refmodel***	***bsr_filter***	***r/d_delay***	***lms_filter***
LUTs	1158	6796	160	697	0	562
FFs	1478	7294	352	883	0	878
DSP48A	22	0	6	13	0	16
Block RAM	0	0	0	0	1 (each)	0
Filter order	161	95	-	95	0	4
Architecture	MAC	DA	MAC	MAC	-	MAC

Table 6.4: Relative resource utilisation of stock and custom FPGA firmware [Sch13].

Type\Design	**USRP N210 (stock)**	**NONLIM**	**USRP N210 + NONLIM**
LUTs	31649/47744 (66 %)	9783 (20 %)	40578/47744 (85 %)
FFs	20007/47744 (41 %)	11446 (24 %)	31179/47744 (65 %)
DSP48A slices	31/126 (24 %)	57 (45 %)	88/126 (70 %)
Block RAM	41/126 (32 %)	2 (2 %)	43/126 (34 %)

GSM. The delay of the individual modules in FPGA clock cycles is detailed in Table 6.5. The total delay accounts for the sum of all sub-modules in the reference branch plus the

Table 6.5: Delay of the invidual FPGA sub-modules [Sch13].

Module	**Delay (cycles)**
bsd_filter	33
bpr_filter	16
bsr_filter	24
lms_filter	6
refmodel	10
r_delay	$49 \cdot 8$
d_delay	$16 \cdot 8$

settling time of the FIR filters, delays by pipeline stages, and rounding operations. The total delay amounts for 201 clock cycles, which corresponds to 2.01 µs with a master clock being $f_{\mathrm{CLK}} = 100\,\mathrm{MHz}$. Regarding this consideration, the convergence time of the AF is not included, as it depends on the signal configuration and chosen step size at hand. Basically, the processing delay of the FPGA prototype is likely to be negligible in modern communication systems.

Another challenge is to enable runtime reconfigurability of the algorithm for changing the algorithm's parameters, which depend on the respective signal scenario. The presented FPGA prototype employs a rather static design. In practice, all filters need to be adjusted to the centre frequency and bandwidth of the (modulated) blocker. Furthermore, the LMS filter parameters L, SF, and MU are to be properly chosen to obtain convergence of the coefficients and finally sufficient mitigation performance. The truncation factors SF and MU can be easily adjusted by a control interface in the USRP architecture (settings bus interface). Besides, flags for adding back the interferer and bypassing the algorithm can be defined. For that purpose, a user interface has been designed in GNURadio [BCB+]. On the other hand, the AF length and the FIR characteristics cannot be easily adjusted. While the FIR coefficients may be changed using a block RAM, the maximum length is fixed in the synthesized design. Also, the filter design should be established on a PC, as it would not be efficient to run it directly on the FPGA. Then, the new set of filter coefficients can be transferred via the Ethernet interface to update those in the FPGA design. Eventually, the general filter structure may be changed during runtime by partial reconfiguration of the FPGA [Xila].

6.3 Experimental Results

This section analyses the real-time performance of the algorithm with real communication signals that are fed to the device and processed with the custom FPGA firmware. For that purpose, a two-tone and a BPSK-modulated blocker have been generated, similar to earlier post-processing-based experiments. In both cases, the AF stage comprises $L = 4$ taps that are fixed in the FPGA design.

The setup for the two-tone scenario is depicted in Figure 6.6. Two sinusoidal signals with a spacing of 600 kHz are fed to the USRP+WBX device using a power combiner [Ett]. A 10 MHz reference clock is adopted for coherent interaction between the devices. The power spectral density (PSD) with the stock and custom firmware is illustrated in Figure 6.7. Each curve depicts the average of 15 FFT blocks with Hann windowing and length 1024. The corresponding truncation factors are $SF = 16$ and $MU = 11$. It is apparent that the upper and lower third-order IMD products (IMD3) have been mitigated. However, the algorithm is also affecting the negative side of the BB due to the real-valued implementation of the *bsd_filter*. A complex implementation would not affect the mirror frequency range, but results in a larger amount of required logic resources. Unfortunately, there are notches appearing at the locations of the IMD products after mitigation. A detailed analysis revealed

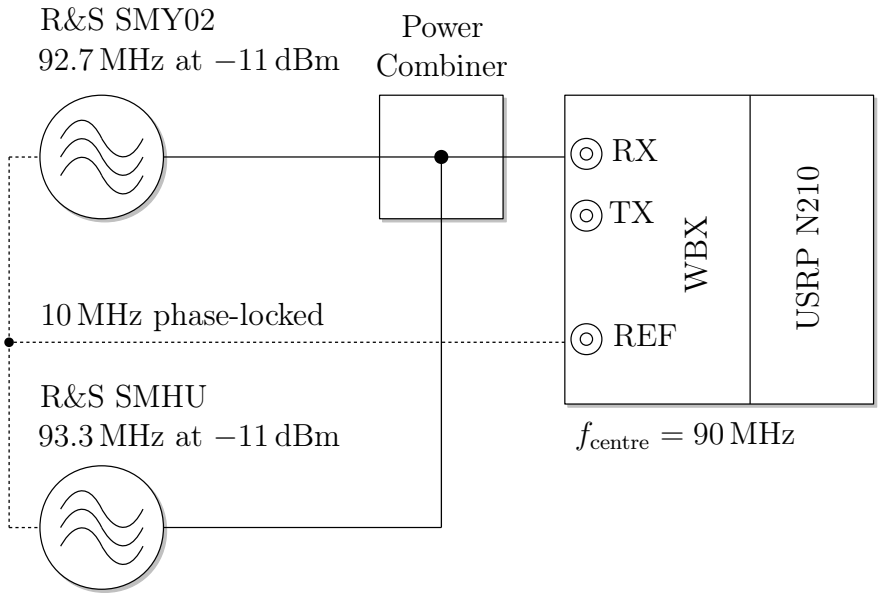


Figure 6.6: Experimental setup of the two-tone scenario.

that these are only visible with measured two-tone signals and if averaging is applied over several FFT blocks. It is likely to be caused by the phase noise of the signal generators and the finite stopband attenuation and finite slope steepness of the FIR filters. Hence, good filter characteristics are essential for providing a clean reference signal to the AF stage. In practice, the reference distortions contain also other components than the IMD products. An increased filter order is desirable to obtain a higher stopband attenuation and a lower transition bandwidth. However, the effect is accepted in this case, due to the large amount of required resources. In the subsequent paragraph, it will be seen that such effects do not appear with modulated blockers.

Next, a BPSK-modulated blocker is combined with a weak desired GSM channel and fed to the USRP+WBX configured with the custom firmware. The measurement setup, depicted in Figure 6.8, consists of two vector signal generators. The BPSK signal has a symbol rate of 450 ksym/s and is filtered with a raised-cosine, which has a roll-off factor of 0.5. In addition, a GSM channel is generated by a second vector signal generator with all 8 time slots being active. As seen in Figure 6.9, the weak GSM channel is completely masked by spectral regrowth of the BPSK signal. After mitigation, the correlating non-linear distortion on top of the weak GSM channel is cancelled. It is worth noticing that the cleansed spectrum perfectly encloses the GSM channel and no notch-like effect is occurring. The corresponding truncation factors are $SF = 18$ and $MU = 9$. In Chapter 7, it will be demonstrated that the BER of a weak GSM channel, which is affected by non-linear distortion induced by nearby stronger channels, can be improved significantly.

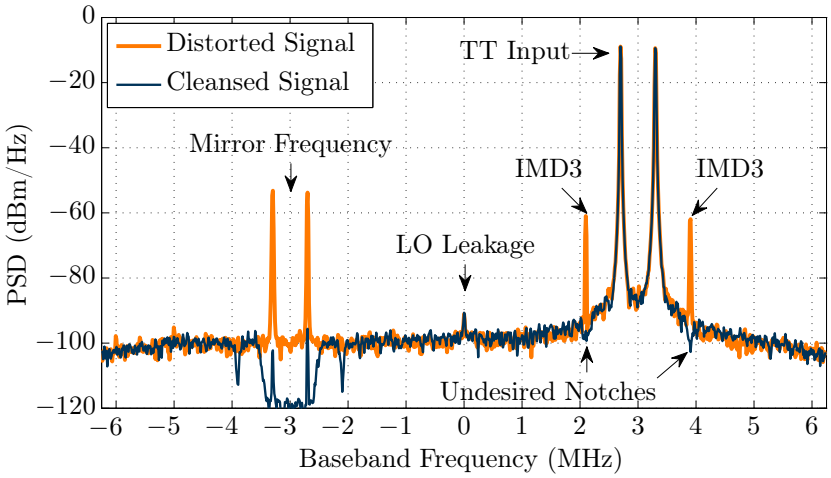


Figure 6.7: Two-tone mitigation performance with the custom FPGA firmware [Sch13].

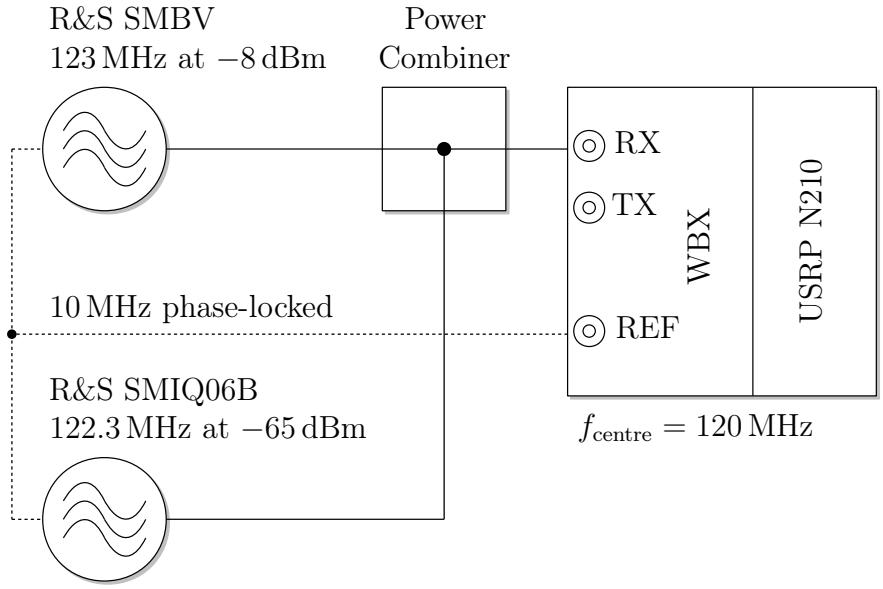


Figure 6.8: Experimental setup of the BPSK scenario.

6.4 Summary of Real-time Implementation Issues

This chapter has presented a resource-efficient implementation of the digital feedforward mitigation on the FPGA of the USRP N210 [Ett]. By embedding the mitigation algorithm into the radio, the cleaned signal is directly forwarded to the host PC for further BB processing. Thus, the receiver self-regulates its non-linearity, whereas the host PC is offloaded. The proposed custom FPGA module adds negligibly small processing delays, hence, TDMA-like communication systems are not disturbed by the mitigation processing. Even the simplified FPGA prototype shows a promising mitigation performance through RF

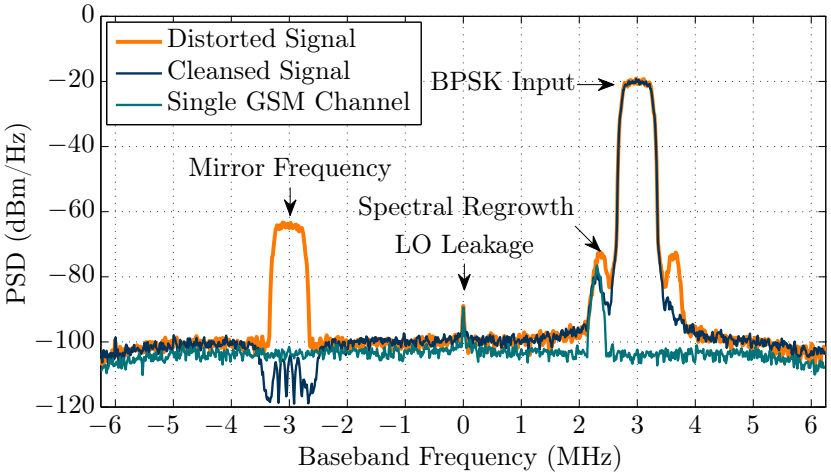


Figure 6.9: BPSK mitigation performance with the custom FPGA firmware [Sch13].

device measurements with two-tone and BPSK blocker signals. The mitigation performance may be improved further by employing the extended cascaded model with a complex-valued LMS implementation, as it has been proposed in Chapter 4.

7 Application Scenarios

The main scope of this chapter is to present some use cases of the digital feedforward mitigation algorithm. Spectrum sensing in CR, GSM downlink reception, and GSM-based passive radar are discussed as examples. The results have been published in [GSH+12c; GSH+12b; GSH+12a; GSH+13].

7.1 Spectrum Sensing in Cognitive Radio

7.1.1 Cognitive Radio Principle

The CR was first coined by Joseph Mitola in his famous article [MM99] following his previous article on SDR [Mit95]. It is a promising concept to improve the spectrum utilisation in communication systems that have a fixed frequency allocation. Typically, many frequency bands in the spectrum are largely unoccupied most of the time, whereas a few are heavily used [Hay05]. A CR is aware of its radio environment and is able to adapt its internal parameters to use available radio resources. Thereby, a secondary use of licensed bands is done on a basis of non-interference with the primary users (PUs) that own the the licensed bands [Hay05; YMC+08].

Whereas a SDR supports multiple standards, a CR adds cognition to it and arranges the adjustments of its parameters [Law10]. Thus, a CR can be considered as an intelligent extension of a typical SDR. To make this concept work, the radio needs to fulfil several capabilities that can be depicted in the cognition cycle [Hay05] illustrated in Figure 7.1. First, the radio needs to perform spectrum sensing to be aware of the radio environment, reflecting the capability *observe*. Then, an intelligent processing is performed on the observed data, reflecting the capabilities *orient* and *plan*. Based on this processing, the CR *decides* on the spectrum opportunities and adjusts its physical transmission parameters, such as the transmission power, frequency, and modulation. Finally, the CR nodes exploit these opportunities under the constraint to avoid interference to the PUs, that corresponds to the capability *act*. The reasoning and learning is a central part of a CR node and is therefore in

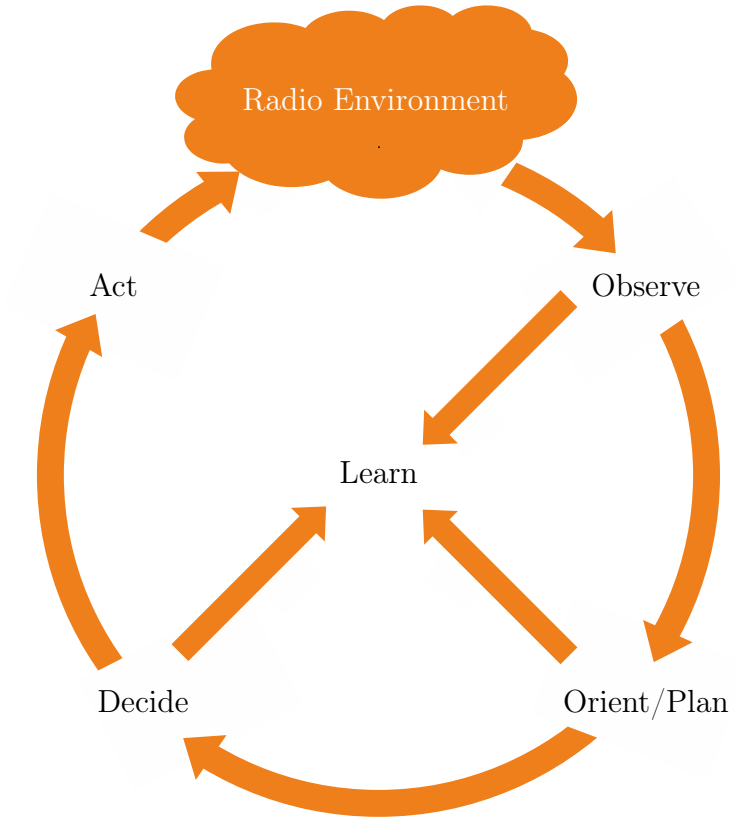


Figure 7.1: CR cognition cycle [Hay05; PKM+11].

the centre of the circle. It is often referred to as cognitive engine that takes decisions based on a knowledge base, a radio environment map, and a policy engine.

There are many challenges of the CR concept, but those related to the RF front-end are on focus here. Basically, CRs impose very tough requirements on the transceiver's electronics [CB05; Raz09; MKH+10]. They need to work under huge dynamic range conditions of up to 100 dB and require a high resolution A/D interface as well as highly linear front-end components. Even more, a CR is supposed to operate at any frequency under a wide range from tenths of MHz to about 10 GHz. Due to the wide bandwidth, also down-conversion of interferers at LO harmonics may be crucial for CRs. Moreover, a CR requires fast digital radio processors that enable fast sensing and adaptation without causing harmful interference to other users. Eventually, a CR need to tolerate interferers at any frequency band, causing high requirements to the mixing purity.

There are certain applications of CRs from TV white spaces, public safety networks, and cellular networks [WGC11]. TV white spaces denote the frequency bands allocated to broadcasting services, but which are not used most of the time at a certain location. They offer a good playground for researchers to verify the CR principle in real-world scenarios until they

move to heavily used licensed bands. In fact, the CR concept is far from being mature and requires additional experimental-driven research to convince regulators to allow CR in future. However, there is also a first CR standard, the IEEE 802.22 wireless regional area network, a broadband internet service designed for use in rural areas. In public safety networks, CR can assist emergency responders of a disaster scenario to overcome the incompatibility of their radios, by serving a facilitator between different devices. In general, CR is applicable in cellular networks to cope with the growing number of services and demands while remaining within the given spectrum mask.

For details on the CR principle, the interested reader is referred to [Fet09; Ars07].

7.1.2 Sensing Algorithms

It is obvious that a CR requires reliable spectrum sensing algorithms to detect the PU activity and to minimize the interference to existing communication systems [CGG+10]. There are three main sensing algorithms known from literature: energy detection, matched filter detection, and cyclostationary feature detection [Ars07]. Energy detection is a simple common technique that does not require a-priori knowledge about the signals to be detected. Instead, matched filter or autocorrelation-based detection is more complex and requires the signal characteristics, i.e. for every PU system. The feature detection seeks "signatures" produced by the modulation schemes, e.g. periodic signal features that are visible only in higher order statistics. Basically, the more knowledge of the signals to be sensed is available, the weaker signals can be detected [WM10].

However, there are several challenges of sensing. First, the effects of the radio channel need to be considered, especially fading conditions and the so called hidden terminal problem due to shadowing effects. Second, the time to arrive at a reliable decision without causing interference should be significantly small. In practice, interference is almost unavoidable, as the PU can be only detected after it already transmits. Eventually, future available radio resources in time, frequency, and space need to be predicted properly. There are several techniques to improve the sensing results, e.g. by multi-antenna-based sensing techniques [CB05] or by collaborative spectrum sensing among multiple nodes [CB05; QCP+08]. In the latter approach, the spatial diversity of the nodes is used and data fusion is subsequently performed to overcome the hidden terminal problem.

Energy-based Detection

Here, the main focus is on energy-based detection because it is the most common approach for detecting unknown signals in CR applications especially due to its ease of implementation [Urk67; DAS03; MGC10]. A major drawback of an energy detector (ED) is its large sensing time when operating in low SNR conditions. Also, an ED is not able to sense weak signals that are covered by noise, however, they have to detect SNRs as low as -20 to $-30\,\mathrm{dB}$ [Raz09].

The ED, illustrated in Figure 7.2, measures the energy of the input signal over a specific time interval T and compares it with a threshold ξ [MGC10; GKM+11a; GKM+11b; Urk67]. The

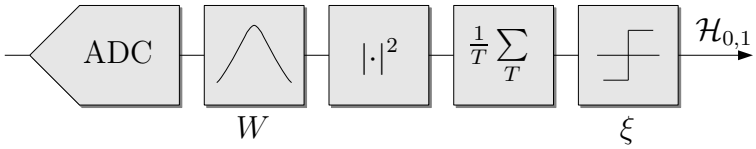


Figure 7.2: Block diagram of a discrete energy detector.

detector usually compares between two hypotheses "signal present" ($\mathcal{H}_1$) and "signal absent" ($\mathcal{H}_0$):

$$\begin{aligned} \mathcal{H}_0 : x(t) &= n(t) \\ \mathcal{H}_1 : x(t) &= n(t) + s(t), \end{aligned}$$

where $x(t)$ denotes the received signal, $s(t)$ the band-limited (W) signal of the PU, and $n(t)$ AWGN. According to Figure 7.2, a discrete-time test statistic [MGC10]

$$V \mathrel{\widehat{=}} \frac{1}{\sigma_0^2} \sum_{i=1}^{N} |x_i|^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \xi \tag{7.1}$$

can be deduced, where $N = TW$ is the total number of samples acquired in the sensing time T and σ_0^2 is the noise power of the detector. There are closed-form expressions for the distribution of both hypotheses depending on the type of modelling for the input signal. The probability of detection P_D and the probability of false alarm P_{FA} are the most common performance figures for an ED and can be expressed by

$$\begin{aligned} P_{FA} &= P(V > \xi \mid \mathcal{H}_0) \quad \text{and} \\ P_D &= P(V > \xi \mid \mathcal{H}_1). \end{aligned}$$

Setting the threshold ξ for a desired performance (P_{FA}, P_D) requires knowledge about the noise power σ_0^2 of the ED [YMC+08]. In theory, the noise power is assumed to be perfectly known but it should be estimated in real systems based on recorded samples from a noise-only observation $\mathcal{H}_0$. In general, low SNR signals can only be detected by sensing a large T like in a spectrum analyser [Ars07]. For long sensing duration T, the aforementioned Gaussian distribution of the noise is valid, and hence the (7.1) in both hypotheses can be approximated as Gaussian [Sha10]:

$$V \sim \begin{cases} \mathcal{N}(\mu_0, \sigma_0^2) & \mathcal{H}_0 \\ \mathcal{N}(\mu_1, \sigma_1^2) & \mathcal{H}_1 \end{cases}$$

with $\mu_0 = 2N$, $\sigma_0^2 = 4N$, $\mu_1 = 2N(1 + SNR)$ and $\sigma_1^2 = 4N(1 + 2SNR)$. In this case, P_D and P_{FA} can be calculated with

$$P_{FA} = \frac{1}{2}\text{erfc}\left[\frac{(\xi - \mu_0)}{\sigma_0\sqrt{2}}\right] \text{ and } P_D = \frac{1}{2}\text{erfc}\left[\frac{(\xi - \mu_1)}{\sigma_1\sqrt{2}}\right], \tag{7.2}$$

where $\text{erfc}(\cdot)$ denotes the complementary error function. In fact, assuming Gaussian random processes provides closed-form expressions for computing the threshold for a given performance (P_{FA}, P_D).

7.1.3 Problem Definition

Compared to the use cases considered in the state of the art, the focus of this work is on analysis and mitigation of receiver non-linearity for reliable detection of the radio spectrum in CR, such that weak signals are properly detected as well as transmit opportunities are not missed. Consequently, distortions are mitigated in the whole BB spectrum to make it suitable for spectrum sensing, instead of eliminating distortions only created on a single information-bearing signal of interest for correct demodulation. In fact, contrary to the signals that have been assumed in the previously mentioned references, this work aims at mitigation of interference caused by realistic modulated signals. In keeping with this objective, realistic constellations with wideband CW user and blocker signals are considered for the interference analysis and mitigation. A feedforward approach with time- and frequency domain implementation of an AF stage is investigated in order to cleanse the BB of distortions. For this purpose, the spectral sensing information about level and spectral location of strong interferers is used as an input for the mitigation algorithm to handle the most critical distortions.

The main objectives for considering non-linear distortion in the operation of a CR are

- Reliable spectrum sensing under huge dynamic range conditions and
- Proper demodulation of weak desired signals for communication.

Signal constellations with weak user and strong blocker signals are on focus in the following discussion. However, from the CR perspective, strong interferers can be PU signals and the weak user signals might be associated with secondary users (SUs).

First, IMD products can show up as unwanted signals in free frequency bands, where the CR may operate as SU. As a consequence, simple energy-based spectrum sensing would detect this band as occupied and the radio would miss its transmit opportunity [GKM+11a]. Figure 7.3 depicts the distorted BB spectrum $\widetilde{y}(t)$ of Figure 2.33 with 1000 averages and the thresholds in 1024 sub-bands for energy detection. For threshold calculation, a noise-only measurement and a false alarm probability $P_{FA} = 0.01$ have been used. The sub-bands are filtered from the input signal by a 1024-point FFT. With respect to the lower IMD region in Figure 7.3, "signal present" was detected in all realisations ($P_D = 1$). These are actually the false alarms due to interference induced by the non-linear wideband receiver, decreasing the spectrum sensing reliability and causing the loss of opportunity for secondary communication.

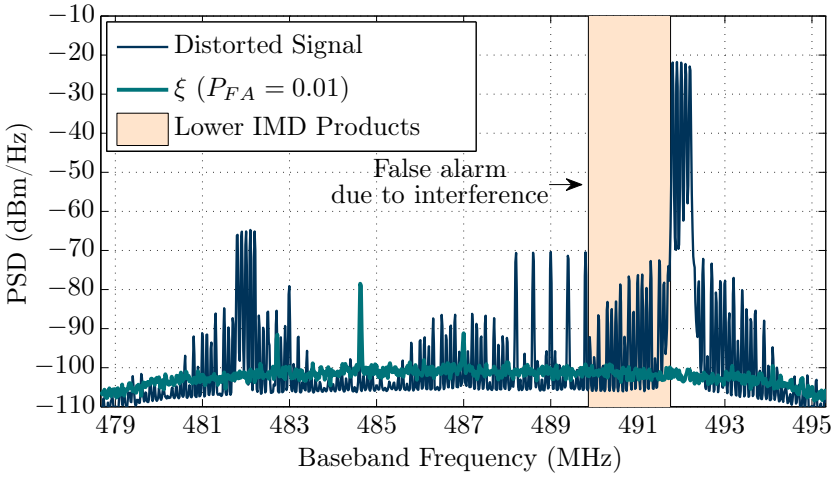


Figure 7.3: Averaged BB spectrum and energy detection thresholds [GSH+12a].

Second, IMD- and XMD-products may coincide with a desired SU signal, causing an increase of the BER and hence difficulties during demodulation.

To alleviate these effects, either interferers themselves have to be cancelled or their distortion products need to be reduced. The interferer itself can only be fully compensated if it is sub-

tracted in the analogue domain at the front-end input, i.e. before the signal passes through any non-linear component. In theory, the analogue cancellation with digital generation of an interferer replica is possible, but not feasible as the interferer must be known in advance (Section 3.4). This approach might work for deterministic or slow-changing (modulated) signals, which can be predicted with high accuracy. Mitigating the distortions is desirable and takes also the interference to other weak users into account. Nevertheless, information about the spectral location of strong interferers could be considered in the sensing algorithm or exploited for secondary transmissions where adjacent frequency ranges being affected by the distortion products are not used [MKH+10]. Secondary transmissions very close to a strong incoming signal can cause harmful intermodulation in the receiver of the mutual secondary communication partner. Instead, frequency ranges that are further away can be used for transmissions.

Adopting digital feedforward mitigation especially for cyclostationary feature detectors is reported in [RGV+13; AMV+13]. Alternative ideas are presented in [MKH+10; Mar09]. The idea in [MKH+10] is to assist the sensing by determining the spectral location of distortion products, so the CR can choose an IMD-free spot for its operation. A similar approach is presented in [Mar09], that proposes pre-selection filters and linear circuits to prevent non-linear distortion and to avoid affected frequency ranges like being occupied by other users.

7.1.4 Detection of Mitigated Multi-carrier Signals

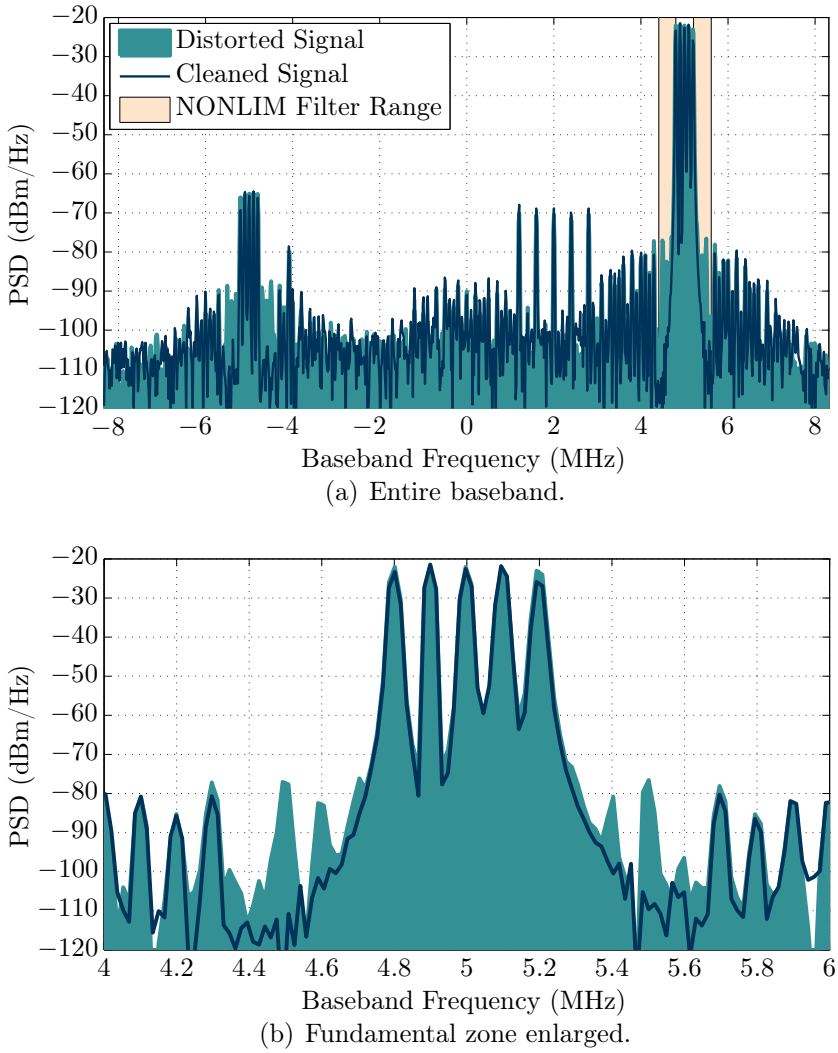
In order to verify the sensing algorithm under conditions of receiver non-linearity, measurements have been conducted with the USRP N210+WBX [Ett]. A multi-carrier CW signal configuration as sketched in Figure 2.33 is assumed with a

- Weak signal: 5-tone sine at 2 MHz with −68 dBm (400 kHz tone spacing), and a
- Blocker: 5-tone sine at 5 MHz with −20 dBm (100 kHz tone spacing).

The signal and carrier frequencies were chosen arbitrarily. The large difference in power of 48 dB was used to provoke non-linearly induced interference at the receiver. The centre frequency and the sample rate were 487 MHz and 16.67 MHz, respectively

The data acquisition and the processing were implemented in MATLAB using NONLIM. Finally, the suppression of the IMD products are examined on the basis of different filter orders and step sizes of the AF stage.

Figure 7.4 illustrates the achieved mitigation performance with conventional LMS filtering after 10 240 iterations. Because odd-order IMDs around the blocker are superimposed on



(a) Entire baseband.

(b) Fundamental zone enlarged.

Figure 7.4: Mitigation performance with measured multi-carrier signals.

each other, they are jointly denoted by spectral regrowth in the remaining discussion. The suppression of the spectral regrowth in the fundamental zone is on average 23 dB, as seen in Figure 7.4(b), indicating the possibility to cleanse the BB of regrowth for reliable spectrum sensing. The mitigation has been performed with a simple BB non-linearity model manifesting 2^{nd}, 3^{rd}, and 5^{th}-order distortion (Section 2.5). The operating range of the mitigation algorithm is depicted by the orange-coloured surface in Figure 7.4(a). For the AF stages, different step sizes have been used: $\mu_2 = 20$, $\mu_3 = 30$, and $\mu_5 = 10$ (subscript n in μ_n denoting the order of non-linearity). The chosen step sizes and the filter order are relatively large because the distortion products had very low amplitudes in general. A substantially long AF length of 128 was necessary to mitigate the distortions, attributed to the simplified I/Q-based processing and the model mismatch.

Now, the ED performance is to be compared before and after application of the implemented mitigation algorithm on a specific bandwidth of interest (260 kHz) in the fundamental zone. Figure 7.5 depicts the original and cleaned baseband spectrum, the band

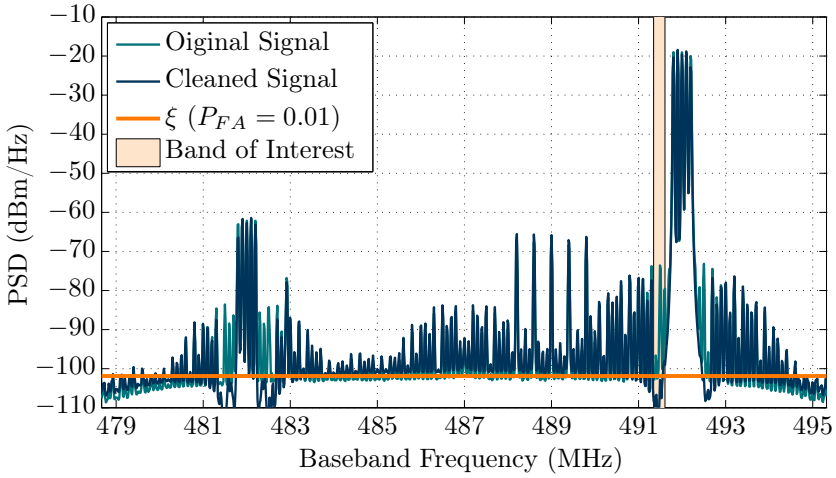


Figure 7.5: Original and cleaned baseband spectrum with energy detection thresholds.

of interest, as well as its ED threshold for $P_{FA} = 0.01$. Figure 7.6 illustrates the receiver operating characteristic (ROC) for both cases, showing the detection of interference [GKM+11a]. Note that the ROC is typically presented for the desired detections. In

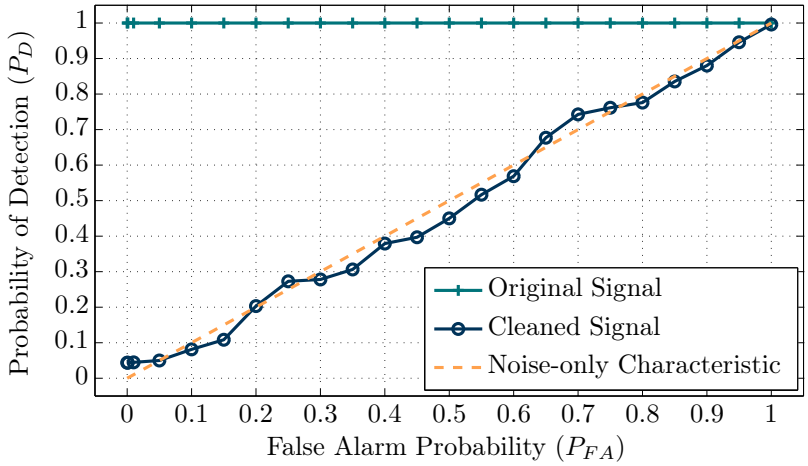


Figure 7.6: ROC curves before and after mitigation.

this case, it shows the detection of non-linearly induced interference with the aim to minimise these detections. While the interference detection probability P_D, or equivalently the false alarm probability due to interference, is always one for the original signal, the ROC curve of the cleaned signal is approaching the ideal noise-only characteristic ($P_D \approx P_{FA}$), indicating that there is no more interference signal present in that band. In fact, the mitigation

aims for avoiding any detection in the band of interest, as only non-linear distortions appear there. After mitigation, the effect of the non-linearly induced interference has been fully compensated, thus improving the reliability of spectrum sensing. In some cases P_D is even below P_{FA}, attributed to the long AFs that suppressed all reference signal content being present in the desired signal. Figure 7.4(a) clearly shows that the algorithm in the used implementation only works locally compared to the wideband impact of the distortions.

7.2 GSM Cellular Radio

7.2.1 GSM Physical and Link Layer Overview

The GSM is the most common mobile communication system worldwide and denotes the second generation (2G) of the digital cellular networks [EVB+09]. GSM implements a Gaussian minimum-shift keying (GMSK) with a modulation index of $h = 0.5$ [Etsd]. It is a variant of the minimum-shift keying that encodes each symbol as a half sinusoid resulting in a constant envelope modulation. In the GMSK case, the data stream is additionally filtered with a Gaussian pulse-shape filter, having a filter length bit period product $BT = 0.3$, to reduce the side-lobe power. This modulation has been chosen to be resistant to any non-linear distortion in the transmitter or receiver, enabling the deployment of low-cost amplifiers in the MS RF front-end.

The symbol rate accounts for $r_b = 1625/6 \cdot 10^3\,\text{bps} = 270.833\,\text{kbps}$. There are four different GSM bursts, namely normal burst (NB), frequency correction burst, synchronisation burst, and dummy burst (DB) [Etsb]. In total, 148 bits are grouped together in one burst, and a guard period of 8.25 bits is left in between the bursts to avoid collisions. The burst duration is approx. 577 µs. Figure 7.7 illustrates the structure of a NB, where $2 \cdot 57 = 114$ are pure data bits. Regarding the access to the network, the GSM implements a TDMA scheme,

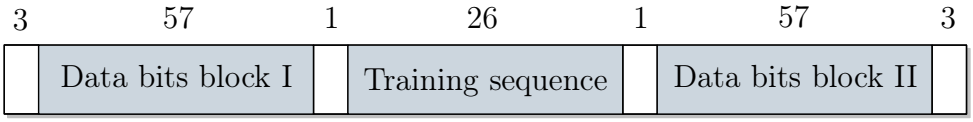


Figure 7.7: GSM NB structure.

as illustrated in Figure 7.8. Eight time slots (TSs) are grouped in frames that are again integrated into higher hierarchy frames. A single user can assign one or more TSs.

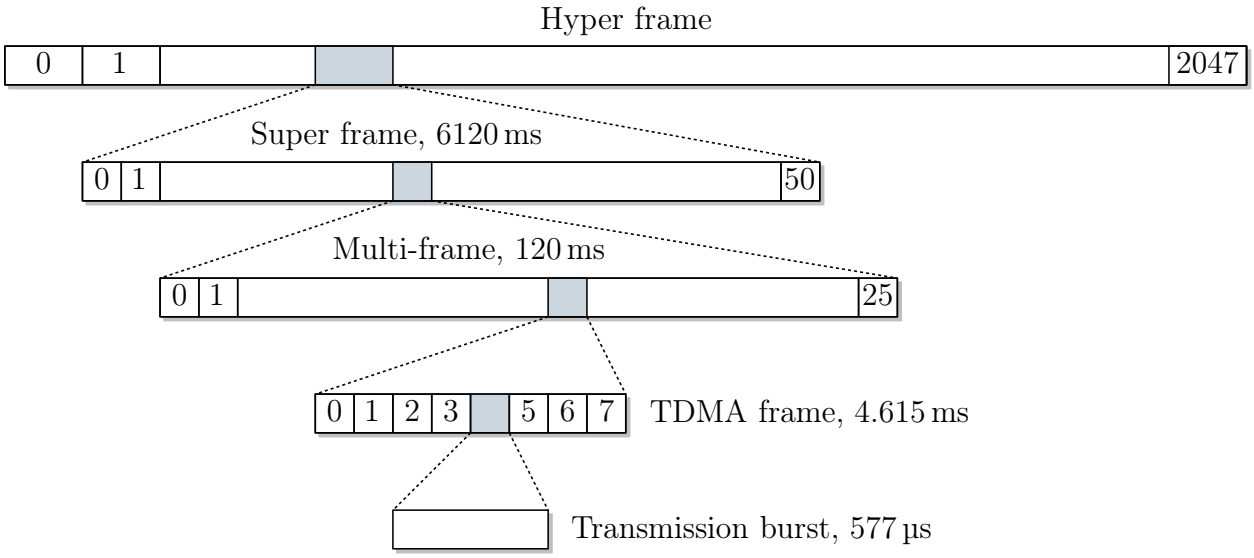


Figure 7.8: GSM frame structure.

In terms of duplexing, FDD is widely used, i.e. there are separated frequency bands for up- and downlink. These bands are divided into 200 kHz wide channels where the GSM carriers are placed.

There are two different types of RF channels that are distinguished in GSM: traffic channels (TCHs) and control channels [Etsb]. The complete control channel, containing the broadcast control channel (BCCH) and other channels, is here referred to as "BCCH". These are beacon channels which transmit continuously at full power. For example, they are used by the MS for synchronisation, network registration, and paging. Every cell in the GSM network has one BCCH and several TCHs. Table 7.1 shows its TDMA structure, which is clearly defined in the standard [Etsb]. In the Table only the first 10 TDMA frames,

Table 7.1: Excerpt of TDMA frame mapping of the BCCH, cp. [Etsb]

	TS0	**TS1**	**TS2**	**TS3**	**TS4**	**TS5**	**TS6**	**TS7**
0	FCCH	DB	DB	DB	DB	DB	DB	DB
1	SCH	DB	DB	DB	DB	DB	DB	DB
2	BCCH	DB	DB	DB	DB	DB	DB	DB
3	BCCH	DB	DB	DB	DB	DB	DB	DB
4	BCCH	DB	DB	DB	DB	DB	DB	DB
5	BCCH	DB	DB	DB	DB	DB	DB	DB
6	CCCH	DB	DB	DB	DB	DB	DB	DB
7	CCCH	DB	DB	DB	DB	DB	DB	DB
8	CCCH	DB	DB	DB	DB	DB	DB	DB
9	CCCH	DB	DB	DB	DB	DB	DB	DB

each consisting of 8 TSs or "bursts", are illustrated. TS 0 is always used for control information, e.g. the frequency correction channel (FCCH), or the synchronization channel (SCH). The remaining seven TSs are reserved for TCHs, but are usually not used. In the idle state, DBs are transmitted, which have a predefined bit sequence.

For more details on the GSM system, the interested reader is referred to [EVB+09] and references therein.

7.2.2 Robust Train Radio Reception (GSM-R)

The work reported in this section provides insight into the degradation of performance of the GSM system by RF impairments in the receiver under realistic interference conditions. The feedforward mitigation algorithm has been applied to real measured GSM signals, distorted due to receiver non-linearity. A significant reduction in the BER for the desired channel has been achieved by reducing non-linear distortions in the receiver caused by two strong neighbouring interferer signals. Thereby, the principle of DSP-based mitigation of receiver non-linearity was successfully demonstrated for real communication signals. The result of these studies have been published in [GSH+13].

Although the GSM system is very mature, interference problems in GSM can still occur. Typically, co-channel interference due to bad network planning, coexisting cellular systems, and radio channel effects are of major concern [Sch94]. The focus here is on downlink interference in GSM, caused by adjacent channels, channels located further away, or even out-of-band signals. All these types of interference signals, herinafter referred to as neighbour channels, can create harmful interference on top of weak desired channels. The interference is induced by the spectral leackage of the neighbour channel or by non-linear distortion in the receiver front-end of the MS. Although the GSM signal itself is not sensitive to non-linearities due to its constant-envelope modulation, IMD may occur in unfavourable signal configurations with at least two strong GSM neighbour channels. Receiver non-linearities seem not to be critical if two adjacent channels belong to the same BTS, since power control mechanisms avoid co-channel interference. However, a crucial coexistence scenario practically appears in the adjacent downlink bands of GSM-R, containing a weak desired channel, and public GSM (GSM-E) that may include various strong channels. Thus, distortions are easily generated in wideband direct-conversion receivers without filtering as used in SDR, but also in narrow-band superheterodyne receivers with limited selectivity and linearity. Distortions may fall top of the desired channel and thus increase its BER, causing a poor speech quality, low data rate, or even dropped calls and a loss of connectivity to the serving BTS. The

signal quality becomes unacceptable if the BER exceeds a certain threshold, or equivalently, a certain receiver quality (RXQUAL) level [Etsc]. Because distortions on top of the desired channel cannot simply be filtered out, clever signal processing is required. The aim is on design and analysis of DSP algorithms to mitigate non-linear distortions in GSM in order to improve the BER performance for maintaining reliable GSM communication.

The main principle of the DSP-based mitigation algorithm has been already discussed in Chapter 4. Up to now, DSP-based mitigation algorithms alleviating receiver distortions have been applied to CW signals (sinusoid, multi-carrier) and simple amplitude modulated signals. There is no work in the state-of-the-art literature applying this kind of DSP-based mitigation to real communication signals. It will be discussed, how this algorithm can be used for GSM signals, aiming for a future employment into real devices.

The GSM communication scenario is sketched in Figure 7.9. The MS is communicating with

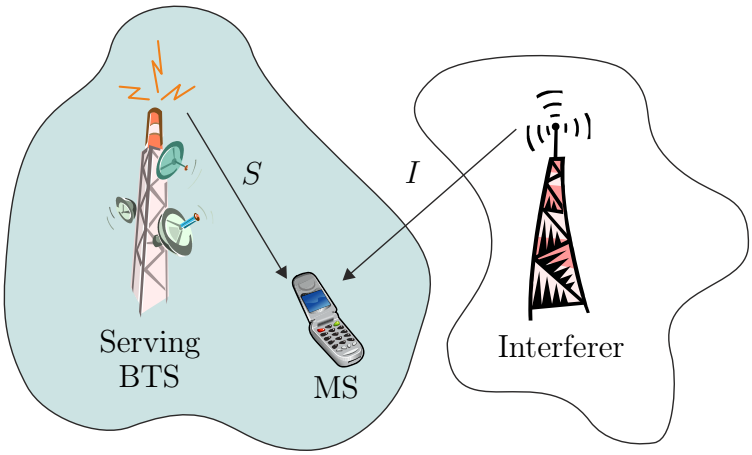


Figure 7.9: Considered interference scenario.

its serving BTS, while it is interfered by another BTS operating in a neighbour channel. The amount of non-linearly induced interference depends on the signal level S of the serving BTS, and the signal level I of the interferer. The SIR is a widely used parameter to describe the amount of interference. These signals vary over time, if the MS is moving (fading). Different types of interfering signals may be present: other GSM signals, UMTS, or even Long Term Evolution (LTE) signals. Figure 7.10 shows a simulated baseband spectrum with a weak desired GSM channel and two strong GSM neighbouring channels. Considering GSM interferers, at least two strong signals need to be present to cause harmful IMD on top of the desired channel. Conversely, single neighbouring UMTS or LTE signals with high PAPR would cause severe regrowth that may mask the narrow GSM signal completely. The spectrum has been computed by a 8192-point FFT with Hann window. In Figure 7.10, the

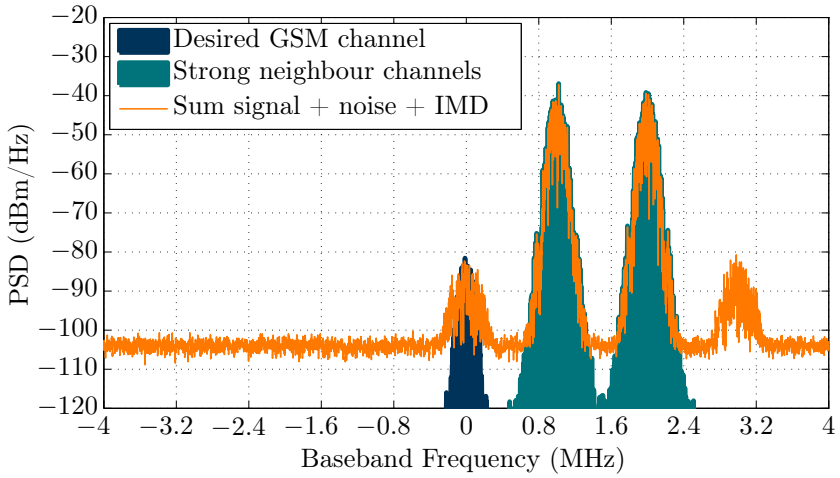


Figure 7.10: Complex baseband spectrum with a weak channel and two strong neighbouring GSM channels, creating IMD on top of the weak channel.

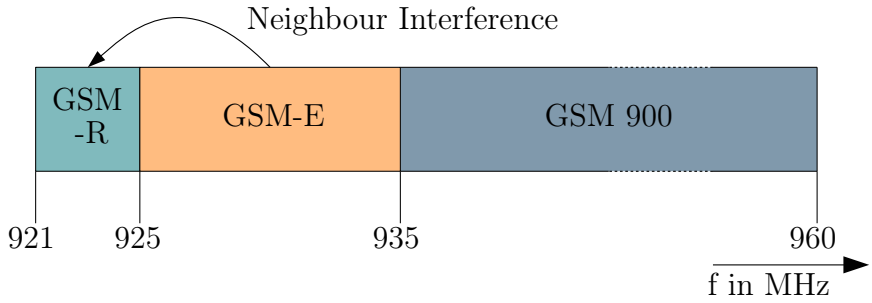


Figure 7.11: Radio spectrum frequency allocation for GSM 900 downlink in Europe.

SIR was set 0 dB, resulting in a BER for the desired channel of 0.38 (103 600 total bits) that makes the signal already unusable.

The scenario illustrated in Figure 7.10, for example, practically appears in the GSM-R [KBN+06; BFM+10]. GSM-R is based on the common GSM standard and uses reserved frequencies for railway applications, e.g. 876-880 MHz in uplink, and 921-925 MHz in downlink in Europe [Etsd]. The frequency allocation for GSM 900 downlink band in Europe is illustrated in Figure 7.11, where interference in the downlink from public GSM-E band to the GSM-R band is of major concern. The problem became more serious since the refarming of the GSM 900 band, since operators are allowed to employ UMTS or LTE systems in GSM-E band. The consequences of this close frequency allocation for the MS in downlink are less considered by regulators or network planners. Although operators keep themselves within their assigned frequency ranges and power levels, it is likely that non-linear distortions in the MS can easily cause harmful interference. A detailed description of GSM-R sensitivity to external wireless interference is given in [BFM+10].

Application of Feedforward Mitigation

The main idea in order to decrease the BER of GSM signals in aforementioned interference scenarios is to mitigate the effects of receiver non-linearity. This is done by DSP at I/Q signal level, which is inserted in between the ADC of the receiver and the GSM specific demodulation. The basic system architecture is shown in Figure 7.12. The pure RF antenna

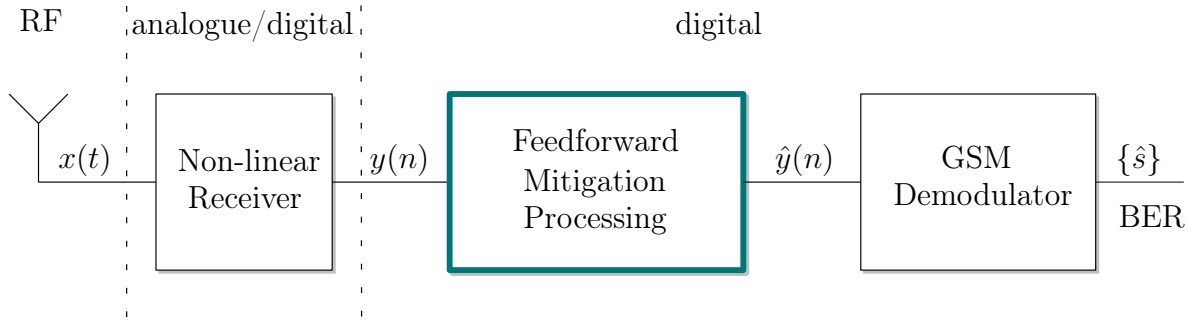


Figure 7.12: Basic system architecture, treating with non-linear distortions in a GSM receiver.

signal $x(t)$ is amplified, down-converted, and digitised by a non-linear receiver to $y(n)$. It does not only contain an amplified baseband version of $x(t)$, but also distortion products added by the receiver. In the very simple case, the non-linear behaviour of the receiver can be modelled by a memoryless polynomial, as discussed in Section 2.5. IMD products generated by mixing of the desired and the interference signal are particularly annoying [Ken00]. The mitigation DSP block tries to eliminate these distortions, outputting the signal $\hat{y}(n)$. The signal is then further processed by the GSM demodulator including synchronisation and channel equalisation. Finally, the BER can be computed with the received bits $\{\hat{s}\}$, if the transmitted bits are known.

Referring to Chapter 4, a feedforward mitigation algorithm has been applied to GSM signal in order to handle third-order distortions of the LNA at RF, falling on top of the desired signal. Figure 7.13 illustrates the main architecture, where the basic idea is to isolate the interferers and to regenerate the distortions they cause. A lowpass/bandpass (h_{LP}^{R}/h_{BP}^{C}) pair splits the input signal into a reference signal y_{ref}, containing the interferer only, and the desired signal y_{des}, including the desired signal and the distortions caused by the interferer. The reference signal is then fed into a reference non-linear model, in this case a simple memoryless polynomial. The term x^2x^* denotes the baseband representation of the third-order distortions induced by the real-valued RF passband signal at the RF amplifier (Section 2.5). The distortion products are then further processed by an AF to adjust their magnitudes and phases to exact values [Hay02]. The common error signal $e(n)$ is used for adapting the

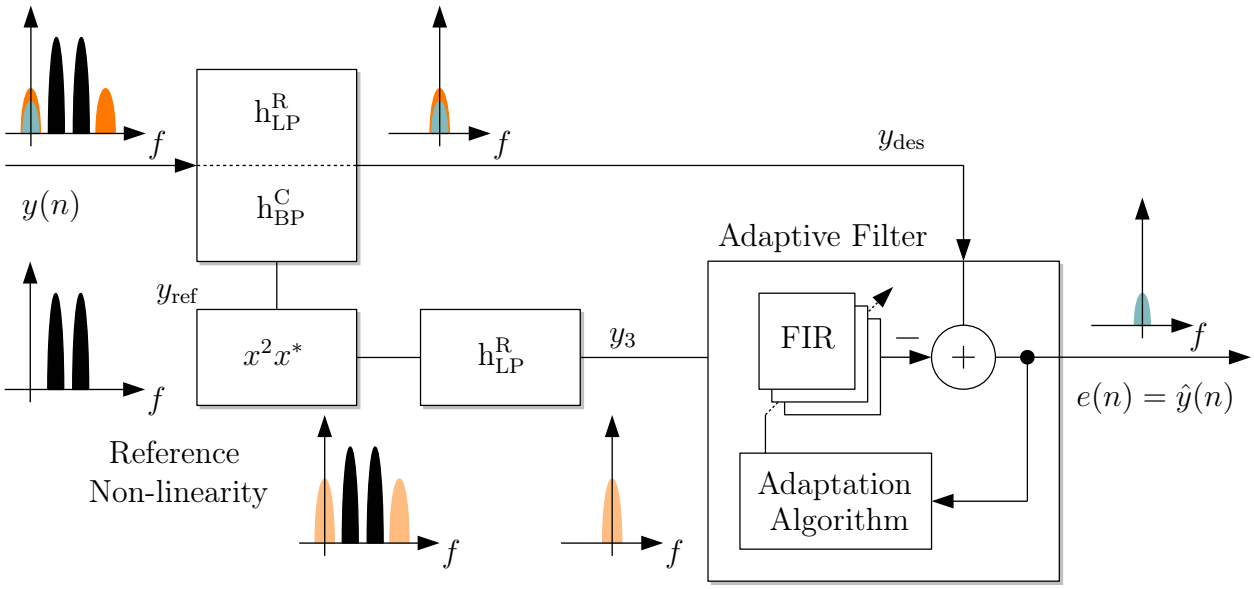


Figure 7.13: Block scheme of the implemented mitigation algorithm for GSM.

FIR filter taps and represents the main output of the mitigation processing. Finally, the reference distortions are subtracted from the desired signal, hence reducing the non-linear distortions falling onto the desired signal. Both I and Q components of the complex BB signal are processed separately for reasons of simplification.

Due to the sample-by-sample operation of the adaptive filtering and the constant-envelope modulation, the TDMA structure does not affect the convergence behaviour of the AF. Instead, the characteristics of the lowpass/bandpass pair as well as the quality of the reference signal y_{ref} dictate the mitigation performance. Table 7.2 summarizes the used filter parameters for a best case estimation. The filter order depends on the dynamic range between

Table 7.2: Filter parameters for the considered GSM interference scenario.

Parameter	Value
Passband h^R_{LP} (real)	$-0.25\,\text{MHz} \ldots +0.25\,\text{MHz}$
Passband h^C_{BP} (complex)	$-2.6\,\text{MHz} \ldots -0.8\,\text{MHz}$
Transition bandwidth Δf	50 kHz
Stopband attenuation	80 dB
FIR order	1200

interfering and desired channel at hand. In this example, it is chosen sufficiently high to obtain high stopband attenuation, and thus pure desired and reference signals. In general, sufficient knowledge about the interferer signal is required, such as its modulation and fre-

quency location. The reference distortions are imperfect as the interferer itself suffers from non-linear distortion (in-band distortion). For exact distortion regeneration, an undistorted version of the interferer, like at the antenna input, is desirable. Suitable approaches for obtaining a better reference signal would be using a second receive chain for the interferer with additional attenuation for linear reception [ZMS09], or re-modulating the interferer after demodulating its bits.

Demodulation and BER Detection

In this work, DBs of the GSM BCCH are used as a desired signal. That is, the transmitted bit sequence is known and used to compute the BER at the receiver. The mitigation processing as well as the GSM demodulation have been implemented in MATLAB using the existing *GSMsim* toolbox that is fully compliant with the GSM standard [EM97]. Figure 7.14 shows a block diagram of the *GSMsim* framework that enables the implementation of a complete GSM communication system. Random bits are generated and passed through the channel

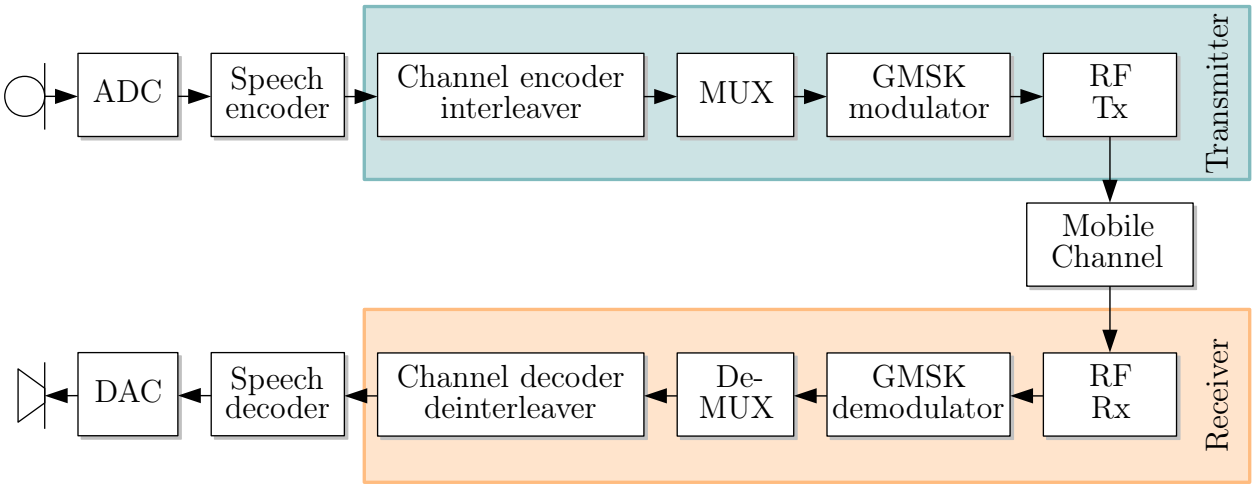


Figure 7.14: Block diagram of the *GSMSim* framework [EM97].

encoder/ interleaver and are eventually compared to the received bits delivered by the channel decoder/ de-interleaver. The voice interfaces are not included in the toolbox and are just depicted to complete the communication system.

The transmitter adds channel encoding to assist the receiver for detecting transmission errors. The task of the interleaver is to shuffle the bits and to distribute them over a number of bursts, so the errors appearing in a received data block are uncorrelated. Also, a training sequence is added by the MUX to enable channel estimation at the receiver. Finally, the

bit stream is modulated with GMSK, performing differential encoding of the incoming burst with a non-return to zero sequence.

The detailed block diagram of the GSM receiver is illustrated in Figure 7.15. It represents

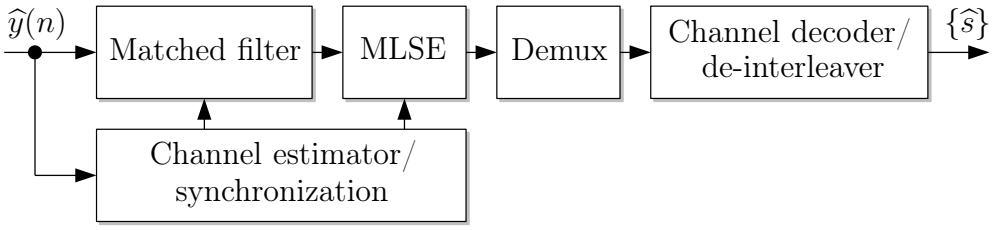


Figure 7.15: Block diagram of the receiver implementation used in *GSMsim* [EM97].

not a complete front-end as, e.g., filtering of the desired channel has to be applied manually. The demodulator estimates the most probable received symbol sequence with the help of the training sequence. First, channel estimation, synchronisation, and matched filtering of the received GSM bursts are performed. The delay and the channel coefficients are determined by cross-correlating the known training sequence with the actual received one. The sequences provided by the matched filter, containing one sample for each transmitted symbol, are then further processed by the minimum least square error (MLSE) detector. The MLSE detector is implemented as a Viterbi equalizer which provides an estimate of the most probable sequence of transmitted binary symbols. Finally, the estimated bits are extracted by de-multiplexing, de-interleaving, and channel decoding.

Simulation Setup

In order to evaluate the performance of the mitigation algorithm, a complete GSM link including receiver non-linearity has been simulated. The non-linear receiver is represented by a third-order polynomial and an AWGN source, whereas the radio channel is not considered herein. The signal configuration as sketched in Figure 7.10 has been simulated with different SIRs for the desired channel, taking into account fixed SNR of 15 dB (constant signal and noise power for the desired channel). In total, 700 DBs (103 600 bits) have been analysed. The adaptive filtering was performed using conventional time domain LMS filtering with a single tap, that has been found to provide sufficient convergence speed and low residual error.

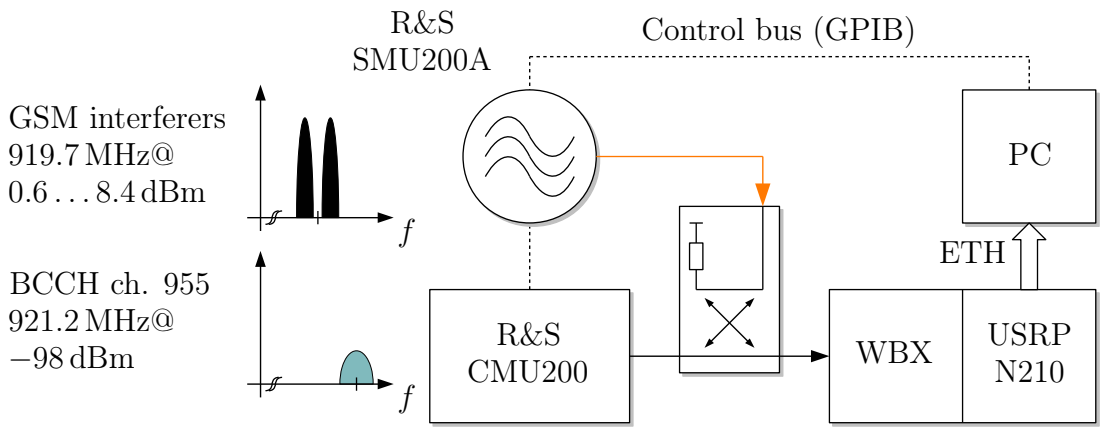


Figure 7.16: Laboratory setup reproducing the considered interference scenario.

Measurement Setup

An experiment with same parameters as in the simulations has been conducted by using an automated setup, as illustrated in Figure 7.16. Fully standard compliant BCCH bursts were generated with a BTS emulator. The interferer signal with two strong neighbouring GSM signals (1.5 MHz offset) was coupled into the RF path using a directional coupler. Both signals were received by the SDR USRP N210, equipped with the WBX wideband front-end [Ett]. The linearity of WBX with a third-order input-referred intercept point IIP3 = +13.8 dBm is considerably good [GSH+12a] and required high power levels for the interferer. The centre frequency and sample rate of the device were 925 MHz (LO offset of 3.8 MHz) and 25 MHz, respectively. The I/Q data has been transferred to a PC via GigE and processed offline as described in Section 7.2.2.

Results

The BER performance before and after mitigation (BM/AM), as well as the ideal BER level without interference (linear receiver), are shown in Figure 7.17 for simulated and measured data. Basically, the BER is saturating at around 0.007 (RXQUAL = 2), which is the lower bound dictated by the fixed SNR of the desired channel. The simulated BER performance after mitigation is approaching to the ideal BER level, indicating that almost all IMD on top of the desired signal has been mitigated. The BER has been reduced significantly to less than 0.03 (RXQUAL = 4) for all SIR conditions. That is, the GSM receiver provides a better detection of the transmitted symbols after applying the mitigation processing. However, the mitigation gain for the measured data is a bit less in low SIR conditions, mainly due to the inaccurate distortion regeneration. In contrast to simulations, an AF of order 6 has been

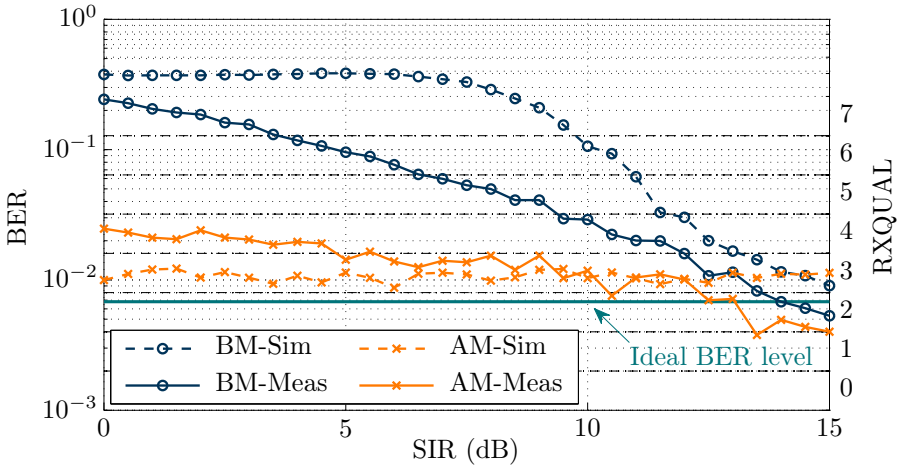


Figure 7.17: BER performance vs. SIR for the desired channel before (BM) and after mitigation (AM).

used for processing the measurement data. Thus, inaccuracies are compensated, that have not been considered in the reference model, such as memory effects and I/Q imbalance.

It has been shown that non-linear distortions can be the predominant interference phenomena, causing serious problems in scenarios with strong neighbouring channels. Due to the presence of non-linear distortions in almost every receiver, this algorithm can be applied in many scenarios. The effectiveness of this approach has been demonstrated here on real communication signals.

In general, the performance is limited by the basic assumptions for DSP-based mitigation processing [GSH+12a]. Assuming high input signal power, the receiver amplifier and mixing stages might be overloaded or even clipping might occur in the ADC. In such conditions, the generated distortions are almost irreparable. Moreover, detecting and isolating the interferer might need an additional RF path, making the receiver architecture more complex. In addition, the effect of the radio channel on the mitigation processing need to be looked at closely.

7.2.3 GSM-based Passive Radar

Beside GSM communication, similar problems with receiver non-linearities exist in GSM passive radar. A GSM based passive radar system, depicted in Figure 7.18, is a bistatic system with no dedicated transmitter [ZNW09; ZDB+12]. It uses GSM BTSs as illuminators of opportunity to detect and track moving targets. At the receiver, the direct signal of the

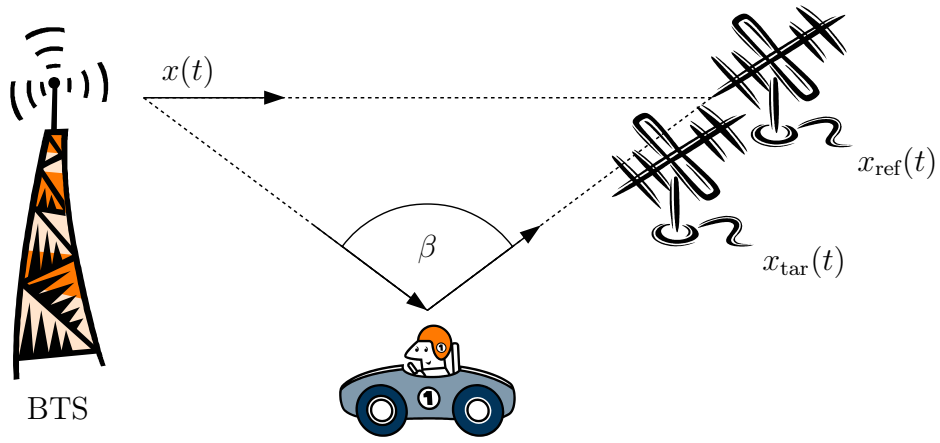


Figure 7.18: Sketch of the passive radar system.

BTS and the reflections from the target are obtained. Finally, the target is located by computing the cross-correlation of both signals, yielding the so called range-doppler map.

The performance of such a system depends critically on the used waveform, which is not under the control of the radar designer. A limiting factor in using GSM as an illumination waveform is given by the low bandwidth which results in a fairly low range resolution (i. e. the ability of separating two targets in range). To overcome this inconvenience, fusion of multiple bistatic configurations (several illuminators) is required. That is, simultaneous reception of all potential illuminations (BTSs) is necessary.

A multi-channel receiver system with an uniform linear antenna array is used to extract the direct path signal (reference path) and target echos (surveillance path), by taking advantage of spatial and time filtering methods. The challenge in passive radar is to detect the weak target echos in continuous presence of the strong direct path signal. Any possible small distortion or neighbouring channel cross-talk will not only affect the extraction of the direct signal, but also mask the weak target echos. The performance of clutter suppression and weak target detection depends, therefore, critically on a clean desired channel.

Next, a simulation is conducted to demonstrate that weak target echos are masked due to clutter induced by non-linearly induced interference. The signal scenario of interest is illustrated in Figure 7.19. It depicts the 8192-point FFT spectrum (with Hann window) averaged over 100 bursts, for the ideal and distorted signal, respectively. Two strong GSM carriers are constantly present at $-0.8\,\text{MHz}$ and $-1.6\,\text{MHz}$, causing third-order IMD around DC and $-2.4\,\text{MHz}$. In the DC zone, there is a target echo 70 dB below the blocker, having a Doppler shift $f_d = 0.2\,\text{kHz}$ and a time shift $t_{\text{shift}} = 40\,\text{ms}$ (360 samples). Thus, it is a delayed version (reflection) of the desired GSM carrier. In addition, there is a clutter signal being 60 dB below blocker level with $f_d = 1.7 \times 10^{-3}\,\text{kHz}$ and $t_{\text{shift}} = 4\,\text{ms}$. The desired

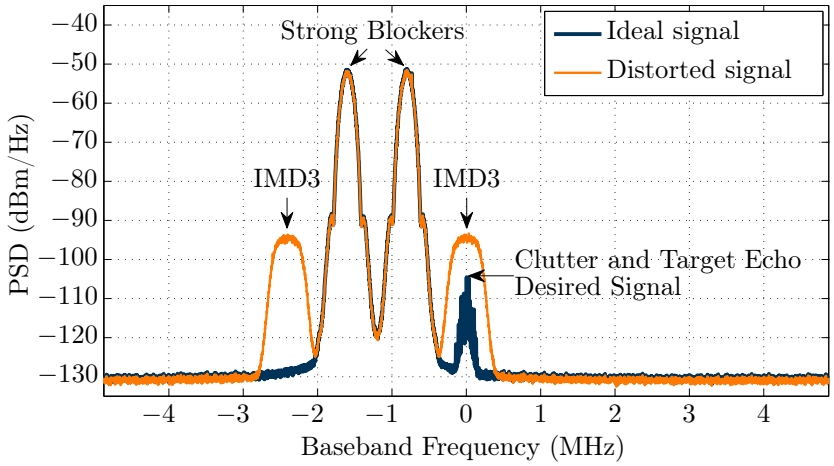


Figure 7.19: BB spectrum of a GSM-based passive radar scenario with a weak desired carrier and two strong blockers.

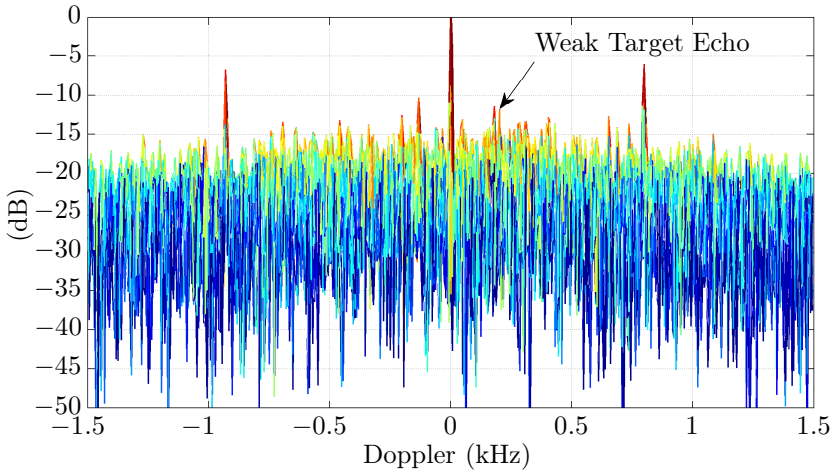
signal is masked by IMD due to RF non-linearity with a third-order coefficient of -12 that corresponds to an IIP3 = 17.4 dBm following (2.14). After filtering the desired channel by a 200 kHz lowpass, the target is located in range and Doppler by assuming an ideal direct (reference) signal.

The diagrams in Figure 7.20(a) and Figure 7.20(b) illustrate Doppler-cuts of the Range-Doppler map before and after applying the feedforward mitigation algorithm. In Figure 7.20(a), the weak target echo at $f_d = 0.2$ kHz is covered by an increased noise floor due to the RF distortion. The target cannot be distinguished from the nearby clutter due to the receiver non-linearity. After applying the algorithm, all interference on top of the desired channel has been mitigated, thereby making that weak target echo visible again.

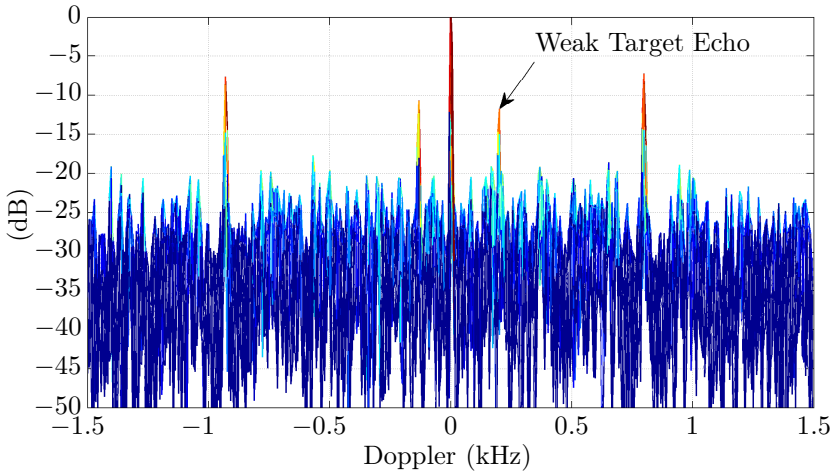
More information on GSM-based passive radar for medium range surveillance can be found in [ZNW09; ZDB+12] and references therein.

7.3 Summary of the Considered Application Scenarios

This chapter has presented three use cases that show the benefit of mitigating receiver non-linearity. In CR spectrum sensing, the false alarm rate due to non-linearly induced interference has been reduced significantly after applying the algorithm. Second, cellular mobile communication based on GSM has been investigated as an example for a real-world communication system. It has been shown through simulations and measurements, how the



(a) Before mitigation



(b) After mitigation

Figure 7.20: Doppler-cuts of the 3D-shaped Range-Doppler maps.

BER performance of weak distorted GSM channels can be improved for reliable communication in GSM-R. Finally, GSM-based passive radar needs to handle weak target reflections and strong direct signals from the GSM BTSs. It has been shown how weak targets can be made visible after applying the feedforward algorithm. The presented use cases are just examples of the broad range of applications of the feedforward mitigation. Basically, the mitigation algorithm can be beneficial in all use cases with weak and strong signals being present at wideband receivers.

8 Conclusions

8.1 Methodology and Main Results of the Thesis

This thesis provides a closed representation of receiver non-linearities, BB modelling, digital mitigation, and practical aspects including implementation issues. Non-linear distortions, being a subset of RF impairments, may cause a significant degradation in performance of sensing, communication, and radar.

IMD in signal configurations of weak and strong signals are of major concern, especially in coexistence scenarios of heterogeneous wireless systems. Circuit non-linearity is even more crucial in wideband receivers with limited selectivity, such as those used in SDRs. Simple behavioural modelling by means of memoryless polynomials, found by system identification techniques, has been shown to be sufficient for the application of distortion mitigation. Thereby, the behavioural model should match the target receiver topology as close as possible.

A purely digital feedforward mitigation algorithm, among other "Dirty RF" techniques, has been followed to enhance the RF performance of the receiver, especially due to its easy integration and efficient mitigation capabilities at system level. It has been shown through extensive simulations and real-world RF measurements that a considerable mitigation of non-linearly induced interference in low-cost RF front-ends can be achieved by applying that algorithm. As a result, a significant increase of the effective linearity of the RF front-end can be obtained. Beside suppression of the newly generated frequency components, the BER of modulated distorted signals has been reduced substantially. RF measurements, including device characterisation and mitigation verification, have been carried out with a typical wideband SDR, the USRP [Ett].

Moreover, it has been found through extensive simulations that the algorithm is able to cope with dynamic radio channel conditions, manifesting severe small-scale and frequency-selective fading. Finally, a first FPGA prototype of the purely digital algorithm has been successfully implemented on the SDR platform USRP, thereby demonstrating the real-time operation of the algorithm on an SDR.

In conclusion, the thesis demonstrates the algorithm's potential of mitigating RF and/or BB distortion created by typical DCR architectures. System-level DSP-based algorithms to handle "Dirty RF" effects represent a promising alternative to conventional analogue design optimisation. A benefit of the digital solution is its versatile range of applications due to its independence to the actual waveform of the desired input, as only the strong blocker signal is processed. The proposed solution is a cost efficient technique that can be easily implemented in hardware, right in between the digitiser and the BB-specific processing. Finally, the pure digital implementation totally meets the concept of SDR. Even more, the mitigation algorithm is generally applicable for wideband DCRs and is not restricted to SDR adopted in the majority of the presented use cases.

8.2 Contributions to the State of the Art

In brief, various studies have been carried out that bridge the gap between theory an practical applicability of the digital feedforward mitigation algorithm (NONLIM), especially with the aim of integrating that algorithm into real devices.

First, an experimental characterisation of the SDR under test (USRP) has been conducted to obtain simple memoryless and memory polynomial models for the purpose of distortion mitigation. Beside the common approach of analysing CW signals, wideband modulated and real communication signals have been used as test signals.

Second, a novel cascaded model has been developed that mimics the DCR receiver topology as close as possible, by joint consideration of non-linearities at RF and BB, as well as I/Q imbalance effects. A feedforward mitigation architecture based on this model clearly outperforms state-of-the-art techniques in terms of the achievable mitigation gain. Add to this, the effectiveness of NONLIM has been verified by extensive simulations and offline-processing of measured RF front-end signals. Above all, limitations of the NONLIM algorithm are summarised together with proposed solutions.

Third, novel studies have been conducted on the impact of fading on the nature of non-linear distortions at the receiver. These results provide general useful information of the interference behaviour in real-world radio environments. The algorithm's performance has been verified through simulations including realistic AGC functionality. The results show a good mitigation and promising BER improvement.

Next, a first FPGA prototype of the NONLIM algorithm has been built to explore the challenges of a real-time implementation. The results show that the algorithm can be easily

integrated in a resource-efficient manner, introducing a negligible overall processing delay, which is of concern in TDMA-like communication systems.

Finally, the algorithm has been applied to spectrum sensing for CR and GSM for reliable communication and weak target detection in passive radar. These use cases have not been considered so far and clearly show the benefit of the "Dirty RF" signal processing technique in real-world scenarios.

8.3 Future Work

Future work will address in more detail, in which scenarios the algorithm will fail or behave counter-productive, especially detecting the need for bypassing the algorithm under high SIR conditions. Next, some limitations of the approach need to be looked at closely, e.g. the burden on the A/D interface. A trade-off between analogue distortion regeneration and cancellation vs. purely digital feed-forward algorithm is to be found. Obtaining ideal reference signals without in-band distortion, either by additional linear RF paths or by re-modulating the blockers, is also an attractive approach to further enhance the mitigation performance. The use of larger FPGAs would allow, for example, to implement the total cascaded model or to enable the mitigation processing at the full ADC sampling rate. Finally, the real-time implementation can be ported to DSP processors, in order to demonstrate the practical applicability of NONLIM on various real-world digital back-ends.

List of Abbreviations

A/D . . . . . analogue/digital

ACPR . . . . . adjacent channel power ratio

ADC . . . . . analogue-to-digital converter

AF . . . . . adaptive filter

AGC . . . . . automatic gain control

AM/AM . . . . amplitude-to-amplitude

AM/PM . . . . amplitude-to-phase

ANT . . . . . antenna

API . . . . . application programming interface

AWG . . . . . arbitrary waveform generator

AWGN . . . . . additive white Gaussian noise

BB . . . . . baseband

BCCH . . . . . broadcast control channel

BEMP . . . . . bit error mitigation performance

BER . . . . . bit error rate

BP . . . . . bandpass

BPSK . . . . . binary phase shift keying

BTS . . . . . base transceiver station

CD . . . . . clipping distortion

CIC . . . . . cascaded integrator comb

CIR . . . . . channel impulse response

CLB . . . . . configurable logic block

CORDIC . . . . coordinate rotation digital computer

CPLD . . . . . complex programmable logic device

CR . . . . . cognitive radio

CW . . . . . continuous wave

DA . . . . . distributed arithmetic

DAC . . . . . digital-to-analogue converter

DB . . . . . dummy burst

DC . . . . . direct current

DCR . . . . . direct-conversion receiver

DDC . . . . . direct-down conversion

DNL . . . . . differential non-linearity

DPD . . . . . digital predistortion

DSP . . . . . digital signal processing

DUT . . . . . device under test

ED . . . . . energy detector

EDA . . . . . electronic design automation

EVM . . . . . error vector magnitude

FCCH . . . . . frequency correction channel

FDD . . . . . frequency division duplexing

FDMA . . . . . frequency-division multiple access

FF . . . . . flip-flop

FFT . . . . . fast Fourier transform

FIR . . . . . finite impulse response

FPGA . . . . . field-programmable gate array

GigE . . . . . gigabit ethernet

GMSK . . . . . Gaussian minimum-shift keying

GPIB . . . . . general purpose interface bus

GSCM . . . . . geometry-based stochastic channel model

GSM . . . . . Global System for Mobile Communications

GSM-E . . . . . GSM extended band

GSM-R . . . . . GSM for railways

HD . . . . . harmonic distortion

HDL . . . . . hardware description language

I . . . . . in-phase component

I/Q . . . . . in-phase/quadrature

IF . . . . . intermediate frequency

IIP2 . . . . . input-referred second-order intercept point

IIP3 . . . . . input-referred third-order intercept point

IMD . . . . . intermodulation distortion

INL . . . . . integral non-linearity

IRR . . . . . image rejection ratio

LMS . . . . . least-mean square

LNA . . . . . low-noise amplifier

LO . . . . . local oscillator

LOS . . . . . line of sight

LP . . . . . lowpass

LS . . . . . least squares

LSB . . . . . least significant bit

LTE . . . . . Long Term Evolution

LTI . . . . . linear time invariant

LUT . . . . . look-up table

M-IMR . . . . . multi-tone intermodulation ratio

MAC . . . . . multiply-accumulate

MATLAB . . . MATrix LABoratory

MIMO . . . . . multiple input multiple output

ML . . . . . maximum likelihood

MLSE . . . . . minimum least square error

MS . . . . . mobile station

MSE . . . . . mean-square error

MUX . . . . . multiplexer

NB . . . . . normal burst

NCO . . . . . numerically controlled oscillator

NF . . . . . noise figure

NLMS . . . . . normalised least-mean square

NLOS . . . . . non-line of sight

NMSE . . . . . normalised mean square error

NONLIM . . . . non-linearly induced interference mitigation

NPR . . . . . noise power ratio

OFDM . . . . . orthogonal frequency-division multiplexing

P1dB . . . . . 1dB compression point

PA . . . . . power amplifier

PAPR . . . . . peak-to-average power ratio

PCB . . . . . printed circuit board

PDF . . . . . probability density function

PPS . . . . . pulse per second

PSD . . . . . power spectral density

PU . . . . . primary user

Q . . . . . quadrature component

RAM . . . . . random-access memory

RF . . . . . radio frequency

RMS . . . . . root mean square

ROC . . . . . receiver operating characteristic

RSSI . . . . . received signal strength indicator

RXQUAL . . . receiver quality

SAW . . . . . surface acoustic wave

SCH . . . . . synchronization channel

SDR . . . . . software defined radio

SFDR . . . . . spurious-free dynamic range

SINAD . . . . . signal-to-interference ratio including noise and distortion

SINR . . . . . signal-to-interference-plus-noise ratio

SIR . . . . . signal-to-interference ratio

SL . . . . . strictly-linear

SNR . . . . . signal-to-noise ratio

SOI . . . . . signal of interest

SPICE . . . . . simulation program with integrated circuit emphasis

SR . . . . . software radio

SU . . . . . secondary user

TCH . . . . . traffic channel

TDMA . . . . . time division multiple access

TDNN . . . . . time-delay neural network

TS . . . . . time slot

TT . . . . . two-tone

UHD . . . . . universal hardware driver

UMTS . . . . . Universal Mobile Telecommunications System

USRP . . . . . Universal Software Radio Peripheral

VGA . . . . . variable-gain amplifier

WCDMA . . . . wideband code division multiple access

WINNER . . . wireless world initiative for new radio

WL . . . . . widely-linear

WL-RC . . . . . reduced-complexity widely-linear

XMD . . . . . cross-modulation distortion

List of Symbols

$\underline{x}$ . . . . . column vector

$\mathbf{X}$ . . . . . matrix

$[\mathbf{X}]^{-1}$ . . . . . matrix inversion

$(.)^*$ . . . . . complex conjugate

$(.)^T$ . . . . . transpose

$(.)^H$ . . . . . Hermitian transpose (complex conjugate transpose)

a_1 . . . . . linear gain of the RF amplifier

a_2 . . . . . complex third-order coefficient of the RF amplifier

a_3 . . . . . linear gain of the BB amplifier

a_4 . . . . . complex third-order coefficient of the BB amplifier

$E\{.\}$ . . . . . expectation operator

J . . . . . mean square error

k_1 . . . . . complex mismatch coefficient for modelling I/Q imbalance

k_2 . . . . . complex mismatch coefficient for modelling I/Q imbalance

$x(t)$ . . . . . complex (transmitted) BB signal

$x_{\mathrm{I}}(t)$ . . . . . in-phase component of complex BB signal $x(t)$

$x_{\mathrm{Q}}(t)$ . . . . . quadrature component of complex BB signal $x(t)$

$x_{\mathrm{RF}}(t)$ . . . . . received antenna RF signal

$y'_{\mathrm{BB}}(t)$ . . . . . BB signal including RF and/or BB distortion

$y'_{\mathrm{RF}}(t)$ . . . . . RF signal including LNA distortion

$y_{\mathrm{RF}}(t)$ . . . . . BB representation of the distorted RF signal

List of Figures

List of Tables

List of Publications

[DGS+13] Diego Dupleich, Michael Grimm, Florian Schlembach, and Reiner S. Thomä. "Practical aspects of a digital feedforward approach for mitigating nonlinear distortions in receivers". In: *Proc. 11th International Conference on Telecommunication in Modern Satellite, Cable and Broadcasting Services (TELSIKS)*. Vol. 01. Nis, Serbia, Oct. 2013, pp. 170–177. DOI: 10.1109/TELSKS.2013.6704914.

[GAM+14] Michael Grimm, Markus Allén, Jaakko Marttila, Mikko Valkama, and Reiner Thomä. "Joint mitigation of nonlinear RF and baseband distortions in wideband direct-conversion receivers". In: *IEEE Transactions on Microwave Theory and Techniques* 62.1 (Jan. 2014), pp. 166–182. ISSN: 0018-9480. DOI: 10.1109/TMTT.2013.2292603.

[GKM+11a] Michael Grimm, Alexander Krah, Noman Murtaza, Rajesh Kumar Sharma, Markus Landmann, Reiner Thomä, Albert Heuberger, and Matthias Hein. "Performance evaluation of directional spectrum sensing using an over-the-air testbed". In: *Proc. 4th International Conference on Cognitive Radio and Advanced Spectrum Management (CogART '11)*. Barcelona, Spain, Oct. 2011, 17:1–17:5. ISBN: 978-1-4503-0912-7. DOI: 10.1145/2093256.2093273.

[GKM+11b] Michael Grimm, Alexander Krah, Noman Murtaza, Rajesh Kumar Sharma, Markus Landmann, Reiner Thomä, Albert Heuberger, and Matthias Hein. "Performance evaluation of directional spectrum sensing using an over-the-air testbed". In: *Proc. 2nd COST Action IC0902 Workshop*. Castelldefels/Barcelona, Spain, Oct. 2011.

[GSH+12a] Michael Grimm, Rajesh Kumar Sharma, Matthias Hein, and Reiner Thomä. "DSP-based mitigation of RF front-end non-linearity in cognitive wideband receivers". In: *FREQUENZ Journal of RF-Engineering and Telecommunications, Special Issue WSR2012* 66.9-10 (Sept. 2012), pp. 303–310. DOI: http://dx.doi.org/10.1515/freq-2012-0052.

[GSH+13] Michael Grimm, Rajesh Kumar Sharma, Matthias Hein, Reiner S. Thomä, and Reda Zemmari. "Improved BER performance in GSM by mitigating non-linear distortions in the receiver". In: *Proc. European Microwave Conference (EuMC)*. Nuremberg, Germany, Oct. 2013, pp. 565–568. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6686718.

[GSH+12b] Michael Grimm, Rajesh Kumar Sharma, Matthias Hein, and Reiner Thomä. "Mitigation of non-linearly induced interference in cognitive wideband receivers". In: *Proc. 7th Karlsruhe Workshop on Software Radios*. Karlsruhe, Germany, Mar. 2012, pp. 119–123.

[GSH+12c] Michael Grimm, Rajesh Kumar Sharma, Matthias Hein, and Reiner Thomä. "Non-linearly induced interference and its mitigation in cognitive wideband receivers". In: *Proc. of 18th European Wireless Conference*. Poznan, Poland, Apr. 2012, pp. 1–6. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6216854.

[KGM+11] Alexander Krah, Michael Grimm, Noman Murtaza, Wim Kotterman, Markus Landmann, Albert Heuberger, Reiner Thomä, and Matthias Hein. "Over-the-air test strategy and testbed for cognitive radio nodes". In: *Proc. XXXth URSI General Assembly and Scientific Symposium*. Istanbul, Turkey, Aug. 2011, pp. 1–4. DOI: 10.1109/URSIGASS.2011.6050532.

[MKG+11] Noman Murtaza, Alexander Krah, Michael Grimm, Albert Heuberger, Reiner Thomä, and Matthias Hein. "Multi-band direction-sensitive cognitive radio node". In: *Proc. IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications (APWC)*. Sept. 2011, pp. 251–254. DOI: 10.1109/APWC.2011.6046759.

[SGT13] Florian Schlembach, Michael Grimm, and Reiner Thomä. "Real-time implementation of a DSP-based algorithm on USRP for mitigating non-linear distortions in the receiver RF front-end". In: *Proc. 10th International Symposium on Wireless Communication Systems (ISWCS2013)*. Ilmenau, Germany, Aug. 2013, pp. 110–114. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6629706.

Bibliography

[ALA09] T. Adali, H. Li, and R. Aloysius. "On properties of the widely linear MSE filter and its LMS implementation". In: *Proc. 43rd Annu. Conf. Information Sciences and Systems (CISS2009)*. Baltimore, MD, Mar. 2009, pp. 876–881. DOI: http://dx.doi.org/10.1109/CISS.2009.5054840.

[AMV10a] M. Allen, J. Marttila, and M. Valkama. "Digitally-enhanced wideband analog-digital interfaces for future cognitive radio devices". In: *Proc. 8th IEEE International NEWCAS Conference*. June 2010, pp. 361–364. DOI: 10.1109/NEWCAS.2010.5604009.

[AMV+13] M. Allen, J. Marttila, M. Valkama, S. Makinen, M. Kosunen, and J. Ryynanen. "Digital linearization of direct-conversion spectrum sensing receiver". In: *Proc. IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. Austin, TX, Dec. 2013, pp. 1158–1161. DOI: 10.1109/GlobalSIP.2013.6737112.

[ALM+12] Markus Allén, Toni Levanen, Jaakko Marttila, and Mikko Valkama. "Iterative Signal Processing for Mitigation of Analog-to-Digital Converter Clipping Distortion in Multiband OFDMA Receivers". In: *Hindawi Journal of Electrical and Computer Engineering* 2012, Article ID 532560 (2012), p. 16. DOI: 10.1155/2012/532560.

[AMV10b] Markus Allén, Jaakko Marttila, and Mikko Valkama. "Modeling and mitigation of nonlinear distortion in wideband A/D converters for cognitive radio receivers". In: *International Journal of Microwave and Wireless Technologies* 2.02 (2010), pp. 183–192. DOI: 10.1017/S1759078710000292.

[Anr00] Anritsu. "Intermodulation distortion measurements using the 37300 series vector network analyzer". In: *Application Note 11410-00257* (Sept. 2000), pp. 1–11. URL: http://www.anritsu.com/en-US/Downloads/Application-Notes/Application-Note/DWL3.aspx.

[Anr12] Anritsu. "Understanding Passive Intermodulation (PIM)". In: *Application Note 11410-00629C* (Dec. 2012). URL: http://www.anritsu.com/en-US/Downloads/Application-Notes/Application-Note/DWL9227.aspx.

[AVR08] L. Anttila, M. Valkama, and M. Renfors. "Circularity-based I/Q imbalance compensation in wideband direct-conversion receivers". In: *IEEE Transactions on Vehicular Technology* 57.4 (July 2008), pp. 2099–2113. ISSN: 0018-9545. DOI: 10.1109/TVT.2007.909269.

[Ant11] Lauri Anttila. "Digital front-end signal processing with widely-linear signal models in radio devices". PhD thesis. Tampere, Finland: Dept. Commun. Eng., Tampere University of Technology, 2011. URL: http://urn.fi/URN:ISBN:978-952-15-2978-8.

[Ars07] Hüseyin Arslan, ed. *Cognitive radio, software defined radio, and adaptive wireless systems*. Dordrecht: Springer, 2007. URL: http://www.gbv.de/dms/ilmenau/toc/51782308X.PDF.

[BFM+10] Gianmarco Baldini, Igor Nai Fovino, Marcelo Masera, Marco Luise, Vincenzo Pellegrini, Enzo Bagagli, Giuseppe Rubino, Raffaele Malangone, Marcoccio Stefano, and Fabio Senesi. "An early warning system for detecting GSM-R wireless interference in the high-speed railway infrastructure". In: *International Journal of Critical Infrastructure Protection* 3.3-4 (2010), pp. 140–156. ISSN: 1874-5482. DOI: 10.1016/j.ijcip.2010.10.003.

[BR06] P. Balister and J. Reed. *USRP hardware and software description*. Tech. rep. Virginia Polytechnic Institute & State University, June 2006.

[Etsa] *Base Station radio transmission and reception (3GPP TS 36.104 version 11.8.2 Release 11)*. ETSI TS 136 104 V11.8.2 (2014-04).

[BESJ+05] Daniel S. Baum, Hassan El-Sallabi, Tommi Jämsä, et al. *WINNER final report on link level and system level channel models*. Tech. rep. Information Society Technologies, Nov. 2005. URL: http://www.ist-winner.org/DeliverableDocuments/D5.4.pdf.

[BB99] Sergio Benedetto and Ezio Biglieri. *Principles of digital transmission : with wireless applications*. New York: Kluwer, 1999. URL: http://zbmath.org/?q=an:0954.94004.

[BBC88] E. Biglieri, S. Barberis, and M. Catena. "Analysis and compensation of nonlinearities in digital transmission systems". In: *IEEE Journal on Selected Areas in Communications* 6.1 (Jan. 1988), pp. 42–51. ISSN: 0733-8716. DOI: 10.1109/49.192728.

[BCB+] Eric Blossom, Johnathan Corgan, Martin Braun, Matt Ettus, and Tom Rondeau. *The GNU Software Radio*. GNU project. URL: `http://gnuradio.org/redmine/projects/gnuradio/wiki`.

[BCG+10] M. Bruno, J. Cousseau, A.S.H. Ghadam, and M. Valkama. "On high linearity - high efficiency RF amplifier design". In: *Proc. Argentine School of Micro-Nanoelectronics Technology and Applications (EAMTA)*. Oct. 2010, pp. 97 –102.

[CB05] D. Cabric and R.W. Brodersen. "Physical layer design issues unique to cognitive radio systems". In: *Proc. IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. Vol. 2. Sept. 2005, 759–763 Vol. 2. DOI: `10.1109/PIMRC.2005.1651545`.

[CGG+10] H. Celebi, I. Guvenc, S. Gezici, and H. Arslan. "Cognitive-radio systems for spectrum, location, and environmental awareness". In: *IEEE Antennas and Propagation Magazine* 52.4 (Aug. 2010), pp. 41–61. ISSN: 1045-9243. DOI: `10.1109/MAP.2010.5638235`.

[Che11] Jesse E. Chen. *Modeling RF systems*. The Designer's Guide Community. Nov. 2011. URL: `http://www.cktsim.org/modeling/modeling-rf-systems.pdf`.

[CGH+11] Zhe Chen, Nan Guo, Zhen Hu, and R.C. Qiu. "Channel state prediction in cognitive radio, part I: Response delays in practical hardware platforms". In: *Proc. of IEEE Southeastcon*. Mar. 2011, pp. 45 –49. DOI: `10.1109/SECON.2011.5752903`.

[Chi11] Y. Chiu. "Equalization techniques for nonlinear analog circuits". In: *IEEE Communications Magazine* 49.4 (Apr. 2011), pp. 132–139. ISSN: 0163-6804. DOI: `10.1109/MCOM.2011.5741157`.

[DS07] C. Dehos and T. C. W. Schenk. "Digital compensation of amplifier nonlinearities in the receiver of a wireless System". In: *Proc. 14th IEEE Symposium on Communications and Vehicular Technology in the Benelux*. Nov. 2007, pp. 1–6. DOI: `10.1109/SCVT.2007.4436243`.

[DGM+06] Rahul Dhar, Gesly George, Amit Malani, and Peter Steenkiste. "Supporting integrated MAC and PHY software development for the USRP SDR". In: *Proc. IEEE Workshop on Networking Technologies for Software Defined Radio Networks*. Sept. 2006, pp. 68–77. DOI: `10.1109/SDR.2006.4286328`.

[DAS03] F.F. Digham, M.-S. Alouini, and M.K. Simon. "On the energy detection of unknown signals over fading channels". In: *Proc. IEEE International Conference on Communications (ICC)*. Vol. 5. May 2003, pp. 3575–3579. DOI: 10.1109/ICC.2003.1204119.

[Etsb] *Digital cellular telecommunications system (Phase 2+); Multiplexing and multiple access on the radio path (GSM 05.02 version 8.5.1 Release 1999)*. ETSI EN 300 908 V8.5.1 (2000-11).

[Etsc] *Digital cellular telecommunications system (Phase 2+); Radio subsystem link control (GSM 05.08 version 8.4.1 Release 1999)*. ETSI EN 300 911 V8.4.1 (2000-10).

[Etsd] *Digital cellular telecommunications system (Phase 2+); Radio transmission and reception (GSM 05.05 version 8.5.1 Release 1999)*. ETSI EN 300 910 V8.5.1 (2000-11).

[DSA09] R. Dinis, P. Silva, and T. Araujo. "Turbo equalization with cancelation of nonlinear distortion for CP-assisted and zero-padded MC-CDM schemes". In: *IEEE Transactions on Communications* 57.8 (2009), pp. 2185–2189. ISSN: 0090-6778. DOI: 10.1109/TCOMM.2009.08.070485.

[DF11] J. Dohl and G. Fettweis. "Blind estimation and mitigation of nonlinear channels". In: *Proc. IEEE 22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*. 2011, pp. 829–833. DOI: 10.1109/PIMRC.2011.6140083.

[DF12] J. Dohl and G. Fettweis. "Evaluation of estimation and mitigation algorithms for nonlinearly distorted OFDM signals on a SDR platform". In: *Proc. IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. 2012, pp. 535–539. DOI: 10.1109/SPAWC.2012.6292966.

[DKF10] Jan Dohl, Stefan Krone, and Gerhard Fettweis. "On the impact of non-linear amplifiers in single-carrier systems: an analytical approach". In: *Proc. IEEE Vehicular Technology Conference (VTC)*. 2010.

[Dup13] Diego Andrés Dupleich. "Mitigation of non-linearly induced interference in wideband receivers under fading conditions". MA thesis. Ilmenau University of Technology, Germany, May 2013. URL: http://www.gbv.de/dms/ilmenau/abs/749809817duple.txt.

[Dup12] Diego Andrés Dupleich. "Modelling and characterization of a non-linear RF front-end for SDR". Advanced Research Project. Ilmenau University of Technology, Germany, Sept. 2012.

[EVB+09] Jörg Eberspächer, Hans-Joerg Vögel, Christian Bettstetter, and Christian Hartmann. *GSM : architecture, protocols and services*. 3rd. Chichester: Wiley, 2009, IX, 326 S. URL: http://www.gbv.de/dms/ilmenau/toc/520372328.PDF.

[EM97] Arne Norre Ekstrøm and Jan Hvolgaard Mikkelsen. *GSMsim: a MATLAB implementation of a GSM simulation platform*. Technical Report R97-1004. Aalborg Universitetsforlag, 1997.

[Ett09] Matt Ettus. *Software Radio, GNU Radio, and the USRP Product Family - Open Hardware for Software Radio*. Mar. 2009.

[FB11] B. Fehri and S. Boumaiza. "Systematic estimation of memory effects parameters in power amplifiers' behavioral models". In: *Proc. IEEE MTT-S International Microwave Symposium Digest (MTT)*. 2011, pp. 1–4. DOI: 10.1109/MWSYM.2011.5972927.

[Fet09] Bruce Alan Fette, ed. *Cognitive radio technology*. Amsterdam: Elsevier, Acad. Press, 2009. URL: http://www.gbv.de/dms/ilmenau/toc/594461537.PDF.

[FLP+05] G. Fettweis, M. Lohning, D. Petrovic, M. Windisch, P. Zillmann, and W. Rave. "Dirty RF: a new paradigm". In: *Proc. IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. Vol. 4. Sept. 2005, pp. 2347–2355. DOI: 10.1109/PIMRC.2005.1651863.

[gha11] A. Shahed hagh ghadam. "Contributions to analysis and DSP-based mitigation of nonlinear distortion in radio transceivers". PhD thesis. Tampere, Finland: Dept. Commun. Eng., Tampere University of Technology, 2011. DOI: http://urn.fi/URN:ISBN:978-952-15-2794-4.

[GCB08] Vito Giannini, Jan Craninckx, and Andrea Baschirotto. *Baseband analog circuits for software defined radio*. Analog circuits and signal processing series. Dordrecht: Springer, 2008, XIII, 142 S. URL: http://www.gbv.de/dms/ilmenau/toc/534916120.PDF.

[Ham08] Firas Abbas Hamza. *The USRP under 1.5X Magnifying Lens*. June 2008. URL: http://gnuradio.org/redmine/attachments/129/USRP_Documentation.pdf.

[Hay02] S. Haykin. *Adaptive filter theory*. Prentice Hall, 2002. URL: http://www.gbv.de/dms/hebis-darmstadt/toc/110863747.pdf.

[Hay05] S. Haykin. "Cognitive radio: brain-empowered wireless communications". In: *IEEE Journal on Selected Areas in Communications* 23.2 (Feb. 2005), pp. 201–220. ISSN: 0733-8716. DOI: 10.1109/JSAC.2004.839380.

[HB08] François Horlin and André Bourdoux. *Digital compensation for analog front-ends : a new approach to wireless transceiver design*. Chichester: Wiley, 2008, VIII, 255 S. URL: http://www.gbv.de/dms/ilmenau/toc/562160094.PDF.

[HS11] Gernot Hueber and Robert Bogdan Staszewski, eds. *Multi-mode/multi-band RF transceivers for wireless communications : advanced techniques, architectures, and trends*. Piscataway, NJ: IEEE Press, 2011. URL: http://www.gbv.de/dms/ilmenau/toc/574815058.PDF.

[Kar01] James L. Karki. "Designing for low distortion with high-speed op amps". In: *Texas Instruments Analog Applicat. J.* (July 2001), pp. 25–33. URL: http://www.ti.com/lit/an/slyt133/slyt133.pdf.

[KBN+06] K. Kastell, S. Bug, A. Nazarov, and R. Jakoby. "Improvements in Railway Communication via GSM-R". In: *Proc. IEEE 63rd Vehicular Technology Conference*. Vol. 6. May 2006, pp. 3026–3030. DOI: 10.1109/VETECS.2006.1683424.

[KH08a] E. A. Keehr and A. Hajimiri. "Digitally-assisted linearization of wideband direct conversion receivers". In: *Proc. European Microwave Integrated Circuit Conference (EuMIC)*. Oct. 2008, pp. 159–162. DOI: 10.1109/EMICC.2008.4772253.

[KH09a] E.A. Keehr and A. Hajimiri. "Analysis of internally bandlimited multistage cubic-term generators for RF receivers". In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 56.8 (Aug. 2009), pp. 1758–1771. ISSN: 1549-8328. DOI: 10.1109/TCSI.2008.2008282.

[KH08b] E.A. Keehr and A. Hajimiri. "Equalization of third-order Intermodulation Products in Wideband Direct Conversion Receivers". In: *IEEE Journal of Solid-State Circuits* 43.12 (Dec. 2008), pp. 2853–2867. ISSN: 0018-9200. DOI: 10.1109/JSSC.2008.2005701.

[KH09b] Edward A. Keehr and Ali Hajimiri. "Digitally assisted equalization of third-order intermodulation products in wideband direct conversion receivers". In: *International Journal of Microwave and Wireless Technologies* 1.Special Issue 04 (2009), pp. 377–385. DOI: 10.1017/S1759078709990341.

[Ken00] P. B. Kenington. *High-linearity RF amplifier design*. Artech House, 2000. URL: http://www.gbv.de/dms/bowker/toc/9781580531436.pdf.

[KAV14] A Kiayani, L. Anttila, and M. Valkama. "Digital Suppression of Power Amplifier Spurious Emissions at Receiver Band in FDD Transceivers". In: *IEEE Signal Processing Letters* 21.1 (Jan. 2014), pp. 69–73. DOI: 10.1109/LSP.2013.2291032.

[KAV13] A Kiayani, L. Anttila, and M. Valkama. "Modeling and dynamic cancellation of TX-RX leakage in FDD transceivers". In: *Proc. IEEE 56th International Midwest Symposium on Circuits and Systems (MWSCAS)*. Columbus, OH, Aug. 2013, pp. 1089–1094. DOI: 10.1109/MWSCAS.2013.6674842.

[KK01] J. Kim and K. Konstantinou. "Digital predistortion of wideband signals based on power amplifier model with memory". In: *Electronics Letters* 37.23 (2001), pp. 1417–1418. ISSN: 0013-5194. DOI: 10.1049/el:20010940.

[Law10] George Lawton. "Cognitive Radio Adds Intelligence to Wireless Technology". In: *Computing Now* (2010). URL: http://www.computer.org/portal/web/computingnow/archive/news059.

[LW10] Michael J. Leferman and Alexander M. Wyglinski. "Taming software-defined radio: a graphical user interface for digital communication system prototyping". In: *DSP-FPGA.com - The Journal of Embedded Signal Processing* (Dec. 2010).

[Lef10] Michael Joseph Leferman. "Radio prototyping interface for software defined radio experimentation". MA thesis. Worcester Polytechnic Institute (WPI), Feb. 2010.

[LHG+10] M. Li, L. Hoover, K.G. Gard, and M.B. Steer. "Behavioural modelling and impact analysis of physical impairments in quadrature modulators". In: *IET Microwaves, Antennas & Propagation* 4.12 (2010), pp. 2144–2154. ISSN: 1751-8725. DOI: 10.1049/iet-map.2009.0278.

[MKH+10] D. H. Mahrof, E. A. M. Klumperink, J. C. Haartsen, and B. Nauta. "On the effect of spectral location of interferers on linearity requirements for wideband cognitive radio receivers". In: *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum*. Apr. 2010, pp. 1–9. DOI: 10.1109/DYSPAN.2010.5457850.

[MGC10] A. Mariani, A. Giorgetti, and M. Chiani. "Energy detector design for cognitive radio applications". In: *Proc. International Waveform Diversity and Design Conference (WDD)*. Aug. 2010, pp. 53–57. DOI: 10.1109/WDD.2010.5592343.

[Mar09] P.F. Marshall. "Cognitive radio as a mechanism to manage front-end linearity and dynamic range". In: *IEEE Communications Magazine* 47.3 (Mar. 2009), pp. 81–87. ISSN: 0163-6804. DOI: 10.1109/MCOM.2009.4804391.

[MRS91] R. Meyer, R. Reng, and K. Schwarz. "Convolution algorithms on DSP processors". In: *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*. Vol. 3. Toronto, Canada, Apr. 1991, pp. 2193–2196. DOI: 10.1109/ICASSP.1991.150849.

[MB04] Uwe Meyer-Bäse. *Digital signal processing with field programmable gate arrays*. Berlin: Springer, 2004. URL: http://zbmath.org/?q=an:1050.94005.

[Mit95] J. Mitola. "The software radio architecture". In: *IEEE Communications Magazine* 33.5 (1995), pp. 26–38. ISSN: 0163-6804. DOI: 10.1109/35.393001.

[MM99] J. Mitola and G.Q. Maguire. "Cognitive radio: making software radios more personal". In: *IEEE Personal Communications* 6.4 (Aug. 1999), pp. 13–18. ISSN: 1070-9916. DOI: 10.1109/98.788210.

[MMK+06] D. R. Morgan, Zhengxiang Ma, Jaehyeong Kim, M. G. Zierdt, and J. Pastalan. "A generalized memory polynomial model for digital predistortion of RF power amplifiers". In: *IEEE Transactions on Signal Processing* 54.10 (Oct. 2006), pp. 3852–3860. ISSN: 1053-587X. DOI: 10.1109/TSP.2006.879264.

[NNS10] F. G. A. Neto, V. H. Nascimento, and M. T. M. Silva. "Reduced-complexity widely linear adaptive estimation". In: *Proc. 7th Int. Symp. Wireless Communication Systems (ISWCS2010)*. York, United Kingdom, Sept. 2010, pp. 399–403. DOI: http://dx.doi.org/10.1109/ISWCS.2010.5624294.

[OKA06] H. Oba, M. Kim, and H. Arai. "FPGA implementation of LMS and N-LMS processor for adaptive array applications". In: *Proc. International Symposium on Intelligent Signal Processing and Communications (ISPACS)*. Dec. 2006, pp. 485–488. DOI: 10.1109/ISPACS.2006.364703.

[OAKS+10] M. S. Oude Alink, E. A. M. Klumperink, M. C. M. Soer, A. B. J. Kokkeler, and B. Nauta. "A 50MHz-to-1.5GHz cross-correlation CMOS spectrum analyzer for cognitive radio with 89dB SFDR in 1MHz RBW". In: *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum*. Apr. 2010, pp. 1–6. DOI: 10.1109/DYSPAN.2010.5457887.

[Xila] *Partial Reconfiguration User Guide*. Xilinx, Inc. Apr. 2013. URL: http://www.xilinx.com/support/documentation/sw_manuals/xilinx14_5/ug702.pdf.

[PC03] Jose Carlos Pedro and Nuno Borges Carvalho. *Intermodulation distortion in microwave and wireless circuits*. Boston: Artech House, 2003. URL: http://www.gbv.de/dms/bowker/toc/9781580533560.pdf.

[Pin01] Rik Pintelon. *System identification : a frequency domain approach*. Ed. by Johan Schoukens. Piscataway, NJ: IEEE Press, 2001. URL: http://www.gbv.de/dms/bowker/toc/9780780360006.pdf.

[PKM+11] A. Puschmann, A. Krah, N. Murtaza, M. Grimm, S. Khan, A. Mahdi, M. Kalil, and R. Sharma. "Poster: cognitive radio research activities at Ilmenau University of Technology". In: *First International Summer School on Cognitive Wireless Communications*. Florence, Italy, July 2011.

[QCP+08] Zhi Quan, Shuguang Cui, H. Poor, and A. Sayed. "Collaborative wideband sensing for cognitive radios". In: *IEEE Signal Processing Magazine* 25.6 (Nov. 2008), pp. 60–73. ISSN: 1053-5888. DOI: 10.1109/MSP.2008.929296.

[Etse] *Radio Frequency system scenarios (3GPP TR 36.942 version 11.0.0 Release 11)*. ETSI TR 136 942 V11.0.0 (2012-10).

[Raz09] B. Razavi. "Challenges in the design of cognitive radios". In: *Proc. IEEE Custom Integrated Circuits Conference (CICC)*. Sept. 2009, pp. 391–398. DOI: 10.1109/CICC.2009.5280806.

[Raz10] B. Razavi. "Cognitive radio design challenges and techniques". In: *IEEE Journal of Solid-State Circuits* 45.8 (2010), pp. 1542–1553. ISSN: 0018-9200. DOI: 10.1109/JSSC.2010.2049790.

[Raz97] B. Razavi. "Design considerations for direct-conversion receivers". In: *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing* 44.6 (June 1997), pp. 428–435. ISSN: 1057-7130. DOI: 10.1109/82.592569.

[RGV+13] E. Rebeiz, AS.H. Ghadam, M. Valkama, and D. Cabric. "Suppressing RF front-end nonlinearities in wideband spectrum sensing". In: *Proc. 8th International Conference on Cognitive Radio Oriented Wireless Networks (CROWNCOM)*. Washington D.C., July 2013, pp. 87–92. DOI: 10.1109/CROWNCom.2013.6636799.

[Rou09] Tony J. Rouphael. *RF and digital signal processing for software-defined radio : a multi-standard multi-mode approach*. Amsterdam: Elsevier, Newnes, 2009. URL: http://www.gbv.de/dms/weimar/toc/578099594_toc.pdf.

[Sch08] T. Schenk. *RF imperfections in high-rate wireless systems : impact and digital compensation*. Dordrecht: Springer, 2008. URL: http://www.gbv.de/dms/ilmenau/toc/547495994.PDF.

[Sch94] A. Schiml. "Interference analysis in the GSM-system". In: *Proc. IEEE 44th Vehicular Technology Conference*. Vol. 1. June 1994, pp. 686–690. DOI: 10.1109/VETEC.1994.345040.

[Sch13] Florian Schlembach. "Implementation of DSP-based algorithms on USRP for mitigating non-linear distortions in the receiver". MA thesis. Ilmenau University of Technology, Germany, May 2013. URL: http://www.db-thueringen.de/servlets/DocumentServlet?id=22477&lang=en.

[Sch09] Dominique Schreurs, ed. *RF power amplifier behavioral modeling*. Cambridge Univ. Press, 2009. URL: http://www.gbv.de/dms/ilmenau/toc/54883024X.PDF.

[Sha] Thomas Shafer. *Balanced Amplifier IP3 and P1dB Improvement due to 90° Hybrid Couplers*. URL: http://www.rfcafe.com/references/electrical/balanced-amplifier-ip3-improvement.htm.

[Sha10] Rajesh Kumar Sharma. "Algorithms, models and measurements for sensing and secure transmission in cognitive radio". PhD thesis. Jacobs University, Bremen, Germany, 2010.

[Spt] *Signal Processing Toolbox*. The MathWorks, Inc. 2013. URL: http://www.mathworks.de/products/signal/.

[Sim] *SimRF - Design and simulate RF systems*. The MathWorks, Inc. 2013. URL: http://www.mathworks.de/products/simrf/.

[Xilb] *Spartan-3A DSP FPGA Family Data Sheet*. Xilinx, Inc. Oct. 2010. URL: http://www.xilinx.com/support/documentation/data_sheets/ds610.pdf.

[Str09] Gilbert Strang. *Introduction to linear algebra*. 4. ed. Wellesley, Mass.: Wellesley-Cambridge Press, 2009, IX, 574 S. URL: http://www.gbv.de/dms/ilmenau/toc/592029204.PDF.

[Teh09] Ali Soltani Tehrani. "Behavioral modeling of radio frequency transmitters". MA thesis. Göteborg, Sweden: Chalmers University of Technology, 2009.

[TGT+96] R. S. Thomä, H. Groppe, U. Trautwein, and J. Sachs. "Statistics of input signals for frequency domain identification of weakly nonlinear systems in communications". In: *Proc. IEEE Instrumentation and Measurement Technology Conference (IMTC-96)*. Brussels, Belgium, June 1996, pp. 2–7. DOI: 10.1109/IMTC.1996.507189.

[Ett] *Universal Software Radio Peripheral*. Ettus Research, A National Instruments Company. URL: http://www.ettus.com/.

[Urk67] H. Urkowitz. "Energy detection of unknown deterministic signals". In: *Proc. IEEE* 55.4 (Apr. 1967), pp. 523–531. ISSN: 0018-9219. DOI: 10.1109/PROC.1967.5573.

[VSHGA+06] M. Valkama, A. Shahed Hagh Ghadam, L. Anttila, and M. Renfors. "Advanced digital signal processing techniques for compensation of nonlinear distortion in wideband multicarrier radio receivers". In: *IEEE Transactions on Microwave Theory and Techniques* 54.6 (June 2006), pp. 2356 –2366. ISSN: 0018-9480. DOI: 10.1109/TMTT.2006.875274.

[VSH10] M. Valkama, A. Springer, and G. Hueber. "Digital signal processing for reducing the effects of RF imperfections in radio devices - An overview". In: *Proc. of IEEE International Symposium on Circuits and Systems (ISCAS)*. 2010, pp. 813–816. DOI: 10.1109/ISCAS.2010.5537444.

[VRW+05] J. Verspecht, D.E. Root, J. Wood, and Alex Cognata. "Broad-band, multi-harmonic frequency domain behavioral models from automated large-signal vectorial network measurements". In: *Proc. IEEE MTT-S International Microwave Symposium Digest*. 2005, pp. 1975–1978. DOI: 10.1109/MWSYM.2005.1517130.

[VR03] Joel Vuolevi and Timo Rahkonen. *Distortion in RF power amplifiers*. Boston, Mass.: Artech House, 2003. URL: http://www.gbv.de/dms/bowker/toc/9781580535397.pdf.

[Vye10] David Vye, ed. *X-Parameters - Fundamentally Changing Nonlinear Microwave Design*. Vol. 53. No. 3. Microwave Journal, Mar. 2010, pp. 22–44. URL: http://cp.literature.agilent.com/litweb/pdf/5990-5594EN.pdf.

[WGC11] Jianfeng Wang, M. Ghosh, and K. Challapali. "Emerging cognitive radio applications: a survey". In: *IEEE Communications Magazine* 49.3 (Mar. 2011), pp. 74 –81. ISSN: 0163-6804. DOI: 10.1109/MCOM.2011.5723803.

[Wbx] *WBX 50-2200 MHz RF daughterboard for USRP*. Ettus Research, A National Instruments Company. URL: http://code.ettus.com/redmine/ettus/attachments/download/211/wbx.pdf.

[WM10] Matthias Wellens and Petri Mähönen. "Intelligente Funker - Funkmodule passen sich dynamisch an ihre Umgebung an". In: *Heise c't Magazin für Computertechnik* 2 (2010), pp. 146–149.

[Wen09] Song Wenmiao. "Configure Cognitive Radio using GNU Radio and USRP". In: *3rd IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications*. Oct. 2009, pp. 1123–1126. DOI: 10.1109/MAPE.2009.5355934.

[WSW+10] T.D. Werth, C. Schmits, R. Wunderlich, and S. Heinen. "An active feedback interference cancellation technique for blocker filtering in RF receiver front-Ends". In: *IEEE Journal of Solid-State Circuits* 45.5 (May 2010), pp. 989–997. ISSN: 0018-9200. DOI: 10.1109/JSSC.2010.2041405.

[WLR+06] J. Wood, M. LeFevre, D. Runton, J.-C. Nanan, B.H. Noori, and P.H. Aaen. "Envelope-domain time series (ET) behavioral model of a Doherty RF power amplifier for system design". In: *IEEE Transactions on Microwave Theory and Techniques* 54.8 (2006), pp. 3163–3172. ISSN: 0018-9480. DOI: 10.1109/TMTT.2006.879134.

[YMC+08] Zhi Yan, Zhangchao Ma, Hanwen Cao, Gang Li, and Wenbo Wang. "Spectrum sensing, access and coexistence testbed for cognitive radio using USRP". In: *Proc. 4th IEEE International Conference on Circuits and Systems for Communications (ICCSC)*. May 2008, pp. 270–274. DOI: 10.1109/ICCSC.2008.63.

[ZDB+12] R. Zemmari, M. Daun, G. Battistello, and U. Nickel. "Target estimation improvement of GSM passive coherent location system". In: *Proc. International conference on radar systems Radar2012*. Oct. 2012.

[ZNW09] R. Zemmari, U. Nickel, and W.-D. Wirth. "GSM passive radar for medium range surveillance". In: *Proc. European Radar Conference (EuRAD)*. Oct. 2009, pp. 49–52.

[ZMB12] Per Zetterberg, Nima Moghadam, and Emil Björnsson. "On RF impairments in MIMO, interference alignment (IA) and coordinated multipoint (CoMP)". In: *Proc. 5th COST IC1004 MC and Scientific Meeting*. Bristol, UK, Sept. 2012.

[ZWC10] Hanxin Zhou, Guojin Wan, and Limin Chen. "A nonlinear memory power amplifier behavior modeling and identification based on memory polynomial model in soft-defined shortwave transmitter". In: *Proc. 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*. Sept. 2010, pp. 1–4. DOI: 10.1109/WICOM.2010.5600848.

[ZMS09] Q. Zou, M. Mikhemar, and A. H. Sayed. "Digital compensation of cross-modulation distortion in software-defined radios". In: *IEEE Journal of Selected Topics in Signal Processing* 3 (June 2009), pp. 348–361. ISSN: 1932-4553. DOI: 10.1109/JSTSP.2009.2020266.

[Zuc13] Giorgia Zucchelli. *Design and Verify RF Transceivers for Wireless Communication Systems*. SimRF Webinar. The MathWorks, Inc. July 2013. URL: https://www.mathworks.de/company/events/webinars/wbnr77705.html.

A Mathematical Derivations

A.1 Complex Memoryless Polynomial

The common polynomial equation (2.7) stemming from a Taylor series expansion holds only for real-valued input $\tilde{x}(t)$ and suits particularly well for modelling RF distortion. However, a notation with complex input $x(t)$ can be derived out of it.

Starting from (2.7)

$$y(t) = \sum_{n=1}^{N} c_n \tilde{x}^n(t) = c_1\tilde{x}(t) + c_2\tilde{x}^2(t) + c_3\tilde{x}^3(t) + \ldots + c_N\tilde{x}^N(t),$$

and writing the real part of $\tilde{x}(t) = \mathrm{Re}\left[x(t)\right]$ as

$$\tilde{x}(t) = \frac{1}{2}\left[x(t)e^{j2\pi f_c t} + x^*(t)e^{-j2\pi f_c t}\right] \tag{A.1}$$

yields

$$\begin{aligned} y(t) &= \sum_{n=1}^{N} c_n \frac{1}{2^n}\left[x(t)e^{j2\pi f_c t} + x^*(t)e^{-j2\pi f_c t}\right]^n \\ &= \sum_{n=1}^{N} c_n \frac{1}{2^n} e^{-j2\pi n f_c t} \underbrace{\left[x^*(t) + x(t)e^{j2\pi 2 f_c t}\right]^n}_{\sum_{k=0}^{n}\binom{n}{k}[x^*(t)]^{n-k}x^k(t)e^{j2\pi 2k f_c t}} \\ &= \sum_{n=1}^{N} c_n \frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k} x^k(t)\left[x^*(t)\right]^{n-k} e^{j2\pi(2k-n)f_c t}, \end{aligned} \tag{A.2}$$

by using the binomial identity. Next, an important assumption is made to simplify (A.2). It is assumed that $y(t)$ passes a zonal filter within the receiver line-up, thus, all frequency components other than those centred at f_c will be removed [MMK+06; BB99; KK01; Teh09]. Hence, only terms with $2k-n = \pm 1 \Leftrightarrow k = (n\pm 1)/2$ remain and only *odd* n will provide non-zero output. Therefore, the binomial terms in (A.2) reduce to two and yield the simplified

output

$$
\begin{aligned}
y(t) &= \sum_{n=1}^{N} c_n \frac{1}{2^n} 2 \binom{n}{\frac{n+1}{2}} x^{\frac{n+1}{2}}(t) \left[x^*(t)\right]^{n-\frac{n+1}{2}} \\
&= \sum_{n=1}^{N} c_n \frac{1}{2^n} \frac{1}{2^{-1}} \binom{n}{\frac{n+1}{2}} \underbrace{\left[x(t)x^*(t)\right]}_{|x(t)|^2}^{\frac{n-1}{2}} x^1(t) \\
&= \sum_{n=1}^{N} \underbrace{c_n \frac{1}{2^{n-1}} \binom{n}{\frac{n+1}{2}}}_{\tilde{c}_n} x(t) \left|x(t)\right|^{n-1} \\
&= \sum_{n=1}^{N} \tilde{c}_n x(t) \left|x(t)\right|^{n-1}. \qquad \text{(A.3)}
\end{aligned}
$$

Thus, (2.8) and (2.9) have been proven.

B Universal Software Radio Peripheral

B.1 Overview

The USRP is an inexpensive and versatile platform for building SDRs, developed by Ettus Research LLC, California [Ett]. It is a frequency-agile RF front-end that can transmit and receive arbitrary signals on a reconfigurable frequency. With its completely open design, it has become a very popular experimental platform for education, research projects, and industrial applications. It consists of a motherboard and a variety of daughter boards and needs a host computer for the main signal processing and visualisation of the results. USRP was developed from the open-source software project "GNU Radio" by Matt Ettus. The current version USRP N210 is on sale since February 2012 and is on main focus in the following sections.

B.2 USRP Motherboard

The main USRP board acts as an interface between the host PC and the actual RF front-ends. A block scheme of the whole USRP system is illustrated in figure B.1. Figure B.2 shows a picture of the USRP motherboard with an accentuation of its most important circuit components. Its task is the digital/analogue and analogue/digital conversion to and from the daughter boards, signal preprocessing, and data transmission to the PC. The printed circuit board (PCB) provides two channels, one receiver and one transmitter. It holds a dual-channel ADC with 14 bit and 100 Msps and a dual-channel digital-to-analogue converter (DAC) with 16 bit and 100 Msps for the I and Q component of the receiver and transmitter path [Ett].

Because the major part of the signal processing is performed on the host PC, high data rates have to be achieved by the interface between USRP and the host. This is a well-known bottleneck of the USRP system. Compared to the former Bus-based USRP, the networked USRP N210 is connected to the PC via a faster GigE instead of universal series bus 2.0 by an extra Ethernet controller chip. However, the data rate still needs to be decreased by

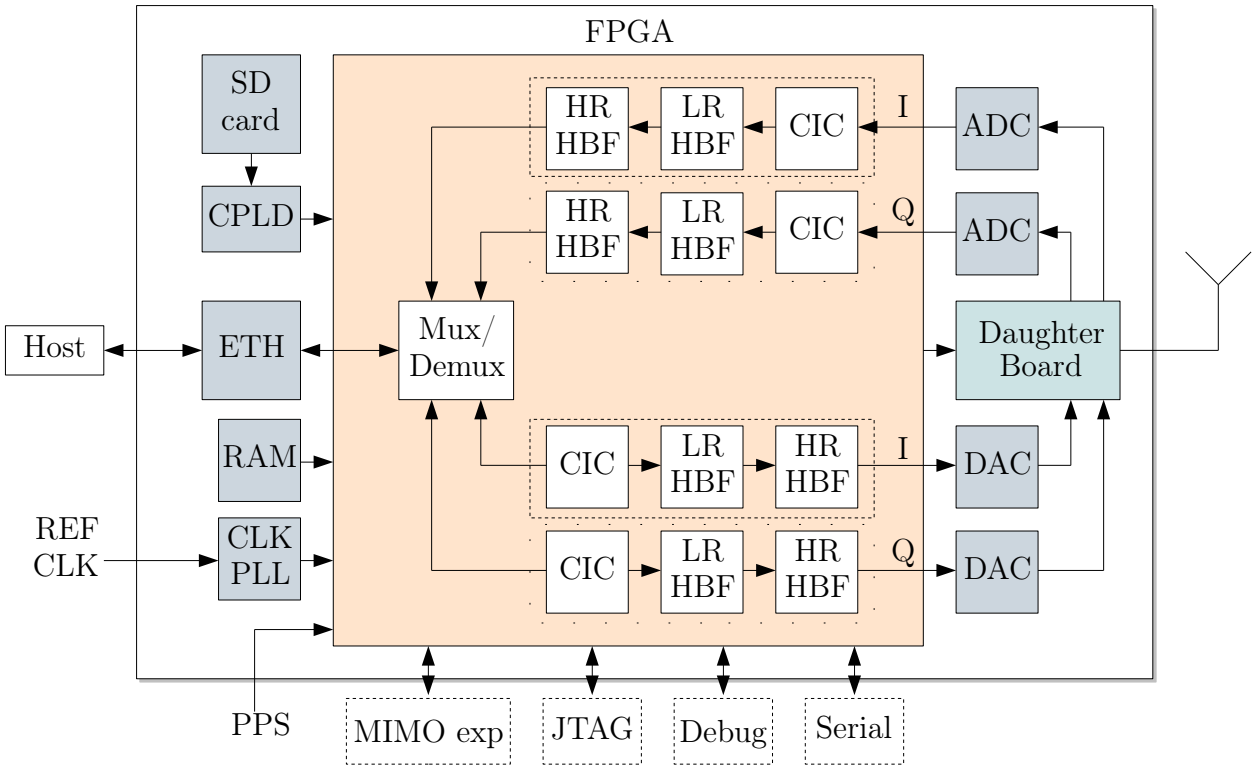


Figure B.1: Block scheme of the complete USRP SDR platform.

decimation and interpolation, respectively. The maximum transferable RF bandwidth for USRP N210 is 25 MHz in each direction, i.e. a total of 50 MHz in full-duplex mode.

The heart of the PCB is an FPGA, the Xilinx Spartan 3A DSP 3400 (XC3SD3400A), which performs high sample rate preprocessing [Ett]. Here, the digital up- and down-conversion to and from an IF stage is done by an NCO and CIC filters [BR06]. The USRP system has a low-IF receiver architecture, i.e. the up- and down-conversion of the BB to RF and vice versa is done in two steps. First, there is a coarse tuning with the programmable synthesizer at the RF daughter board. Then, a digital fine tuning is executed, if the analogue LO frequency is apart from the desired centre frequency. There is a decimation/interpolation filter chain of each I and Q component: a low-rate half-band filter, the CIC and a high-rate half-band filter. The rate of decimation/interpolation can be changed in a range $[4, 512]$, although a factor as multiple of 4 is recommended to enabled both filters. For the transmission of the sampled/generated data via the Ethernet connection, the data is multiplexed and demultiplexed, respectively. There are free logic resources to shift further steps of signal processing directly onto the device by developing an own FPGA design [Lef10; SGT13]. This would allow a higher data rate due to further decimation or a standalone mode of the USRP. The FPGA design uses a lot of open cores, e.g. the Wishbone bus, interrupt controller, GPIO, SPI,

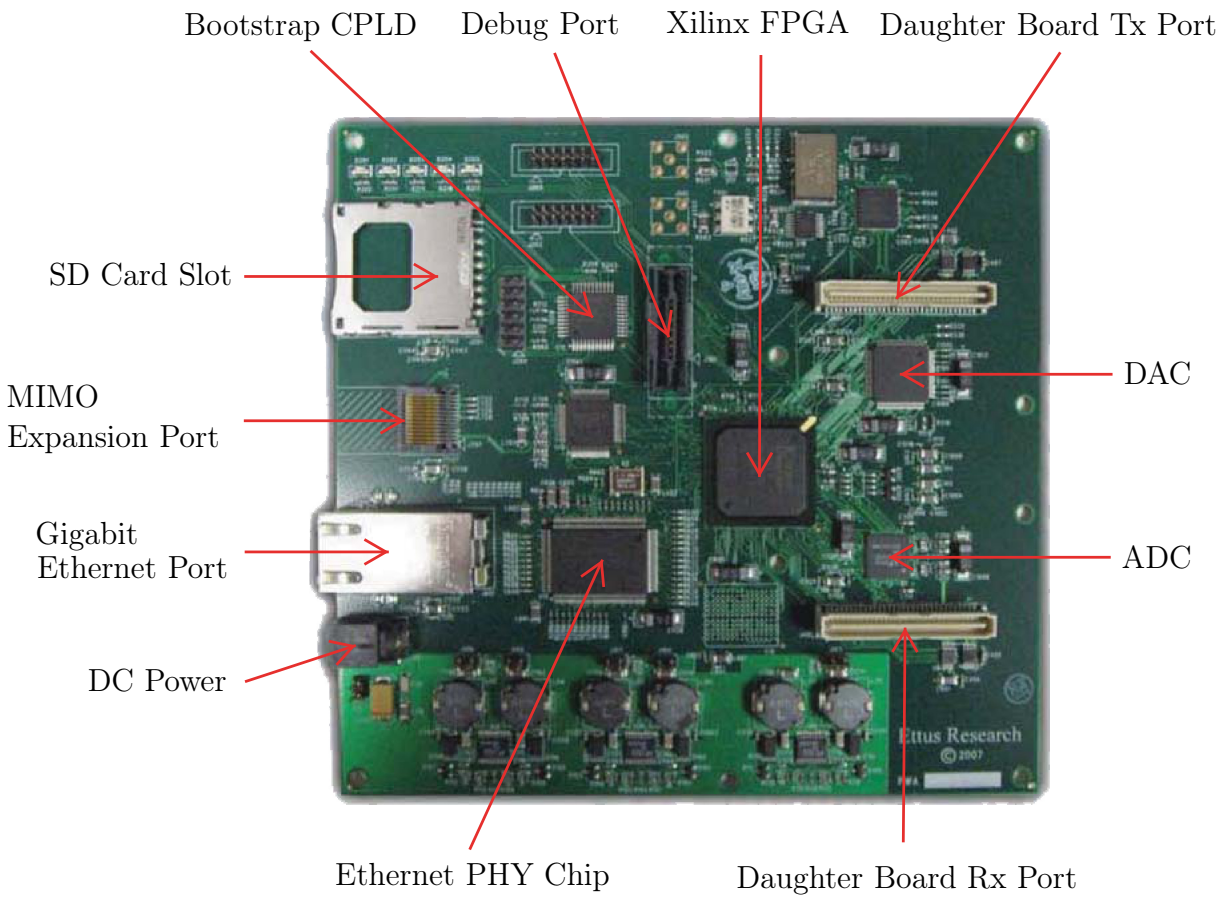


Figure B.2: USRP2 motherboard, Rev. 1.0 (2007), cp. [Ett]

I2C, UART, Gigabit Ethernet MAC and the complex programmable logic device (CPLD) bootstrap [Ett09]. Furthermore, it holds a 32 bit 50 MHz RISC soft-core micro processor, the "aeMB", for control of the daughter boards, Ethernet etc. The FPGA bitfile is loaded from a CPLD after power-up to the actual FPGA.

There a lot of further parts and features on the USRP, but most of them are currently not supported by the prebuild design and need a custom design. There are

- 1 MB of extra SRAM onboard,
- A multiple input multiple output (MIMO) expansion to run multiple USRPs fully coherent in a multi antenna system [LW10],
- A reference clock input for synchronisation,
- A 1 PPS input for precise timing applications,

- Auxiliary input and output channels (auxiliary ADCs and DACs) for received signal strength indicator (RSSI) level measurement, temperature, bias levels [Ham08],
- I/O control interface for daughter boards (e.g. selection of RF port, synthesizer lock detection [Ham08]),
- Debug interface for common logic analyzers,
- High-speed serial interface, and a
- JTAG interface for reprogramming the CPLD or FPGA debugging purposes using the Xilinx ChipScope tool.

B.3 RF Front-end Daughter Boards

Ettus Research LLC offer a variety of RF front-ends for designated frequency ranges from DC up to 6 GHz. The transmitter and receiver are either in single daughter boards or combined in one transceiver board. The daughter boards hold an analogue circuitry to up-convert the signal to passband frequency or to down-convert it via synthesizers and I/Q mixers, which are software-controllable [BR06]. The daughter board is the direct interface for attaching antennas. Furthermore, it provides amplification or attenuation of the IF signal. All parameters can be controlled from software or directly through the FPGA [Ett]. In addition, most of the daughter cards have a built-in transmit/receive switching to configure either one RF port as Tx/Rx or to have separated Tx and Rx on two RF ports for a full-duplex operation. They also hold an analogue RSSI measurement and a digital I/O to control external devices like antenna switches. Because of the USRP's open design, a development of an own custom daughter board is possible to fulfil specific design requirements.

The daughter boards used at Ilmenau University of Technology are [Ett]:

- RFX2400, full-duplex TRx for Wi-Fi frequency from 2.3 GHz ... 2.9 GHz and
- WBX, full-duplex wideband TRx for 50 MHz ... 2.2 GHz.

The RFX2400 has a transmit power of 50 mW (17 dBm) and the WBX up to 100 mW (20 dBm), which is frequency-dependent. Figure B.3 and figure B.4 show a picture of these PCBs. The TRx boards have independent LOs/RF synthesizers for the Tx and Rx path

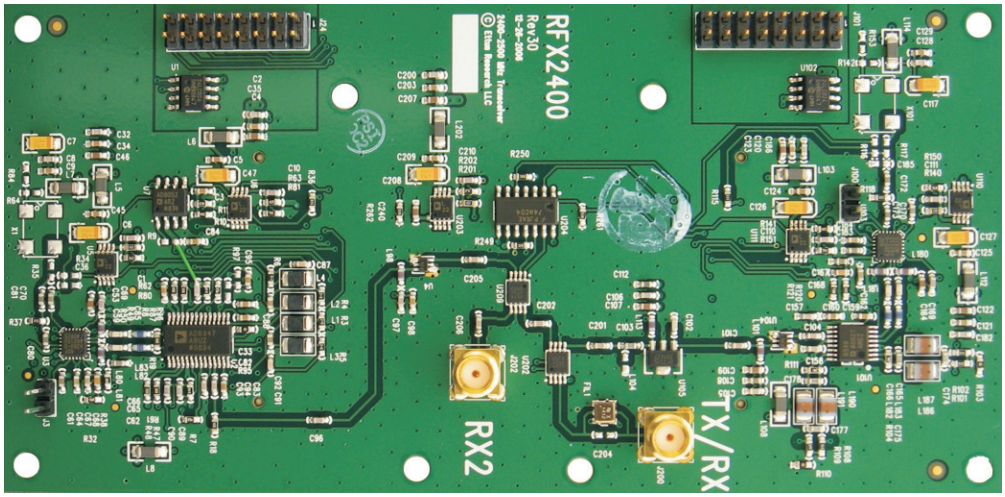


Figure B.3: Wi-Fi transceiver board RFX2400 [Ett]

to allow a split-frequency operation [Ham08]. Moreover, each daughter board has an EEPROM, which identifies the board with its parameters to the system (e.g. range of carrier frequency and gain) and stores calibration data like DC offsets or I/Q imbalances.

B.4 Host Computer Processing and Data Acquisition

The main part of the BB signal processing takes place at the host computer. However, the main idea of SDR is to perform all the waveform-specific processing (modulation and demodulation) in software to be most flexible [Ham08]. There are some existing software frameworks for own SDR implementations like GNU Radio, IRIS, and OSSIE [Ett]. Furthermore, some groups achieved to connect USRP with MATLAB/Simulink and LabView. The most famous SDR framework is GNU Radio and the origin for the idea to develop USRP.

GNU Radio is an open-source toolkit and started in early 2000 by Eric Blossom [DGM+06]. It includes drivers for USRP and a lot of pre-assembled libraries for modulators, filters, channel coders, and equalizers. With the GNU Radio libraries, a flow graph (written in Python) with sources and sinks and a connection of signal processing blocks (written in C++) can be realised. The flow graph is a direct realisation of radio block diagrams and allows signal processing of continuous data streams. Furthermore, there is a control interface in Python for the software-defined parameters, e.g. centre frequency, gain/attenuation, interpolation/decimation factor, Tx and Rx multiplex options, whereas the range for certain parameters depend on the attached daughter board [Wen09]. One disadvantageous aspect of GNU Radio is its continuous transmission of data streams without concepts for packets or

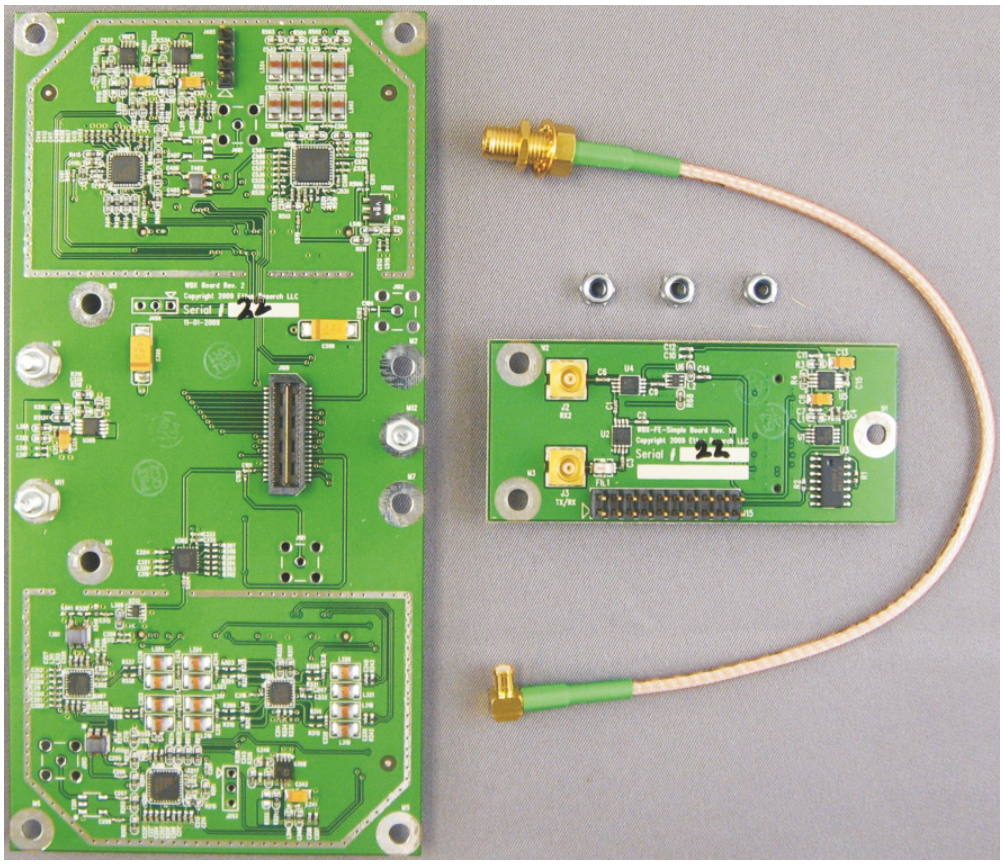


Figure B.4: Wideband transceiver board WBX [Ett]

frames as well as timers [DGM+06]. Therefore, a direct MAC implementation is difficult. A control for the MIMO expansion, external clocks and antenna selection for USRP is currently not supported by GNU Radio and needs a custom FPGA and firmware design [LW10].

After USRP is powered on, the onboard CPLD uploads the bit file to the FPGA and the firmware for the soft-core micro processor. In the transmit case, the host generates samples and sends it to the USRP device. Here, the data is demultiplexed, passed through filter chains and converted to analogue by the DAC. After a further analogue filtering, the signal is up-converted to RF frequency at the RF daughter card. The higher the interpolation rate is, the fewer samples have to be generated by the host.

The signal chain for the receive case is completely inverse. First, a down-conversion of the received signal from RF to BB is done a the RF daughter card. Then, the signal is converted to digital by the ADC, filtered and decimated, multiplexed and transmitted to the host PC. According to the transmit case, the higher the decimation rate is, the lower the data rate for the transmission to the host PC.

C Source Code

C.1 NONLIM Algorithm

The NONLIM simulator is an object-oriented implementation of the DSP-based digital feedforward mitigation algorithm in MATLAB. It comprises a group of objects, in the following referred to as "blocks", that are connected with each other to form a flowgraph that realises the feedforward mitigation processing. This flowgraph is illustrated in Figure C.1 and is similar to a MATLAB Simulink or GNURadio environment [BCB+]. The main idea for this

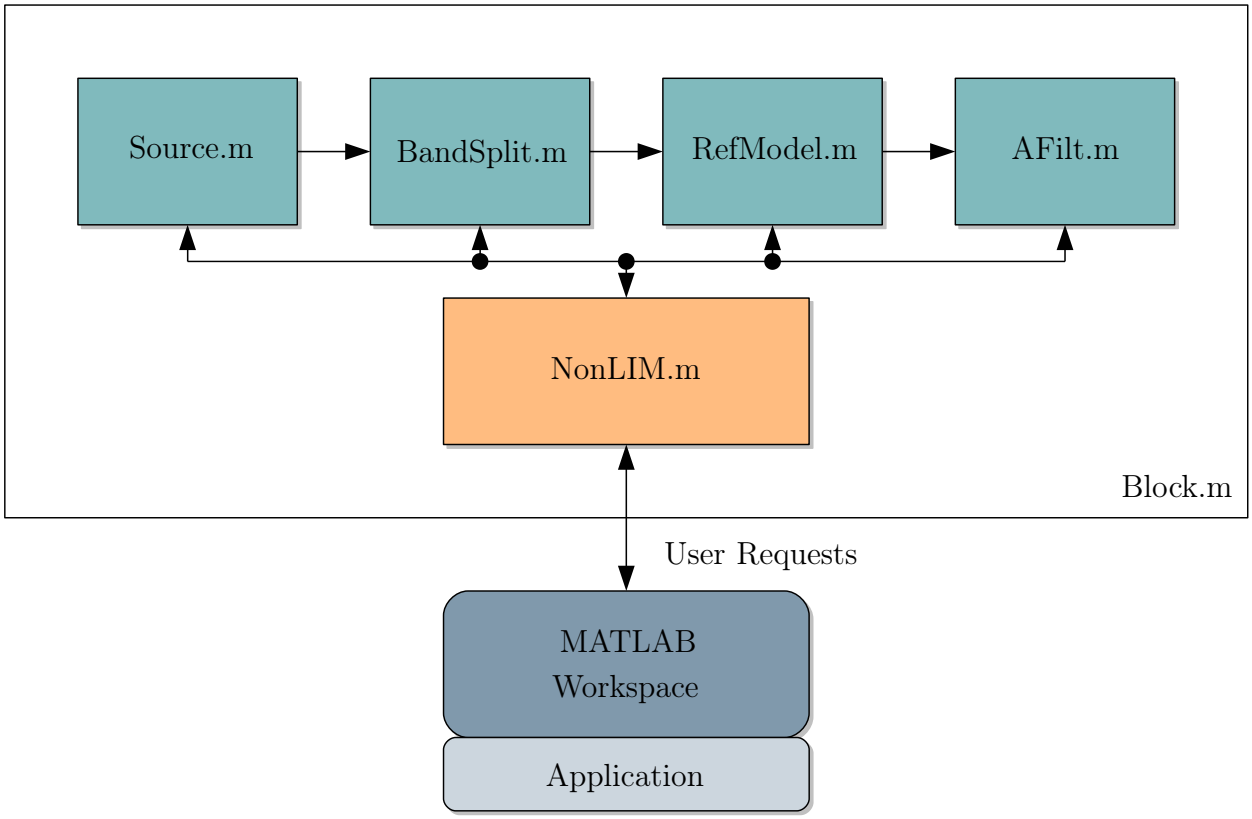


Figure C.1: Flowgraph of the MATLAB simulator NONLIM.

unitised structure is to represent the block diagram Figure 4.1 and to assign all block-related properties and methods in an own object. All blocks are connected with their input and

output matrices and a few control signals. Thereby, all individual settings of each block can be easily accessed and controlled in the MATLAB workspace by invoking the related handles. Another great advantage of NONLIM is that it can be easily extended by new methods. That was a great benefit during the development and research, especially for investigations with the different reference models and AF variants including parameter variations. Finally, the algorithm can be applied to very different kind of data, either synthetic simulated samples or samples from real-world RF measurements.

Next, the functionality of each block is briefly described.

NonLIM.m is the top class that instantiates other blocks and connects them into the flowgraph. It holds some methods for plotting and calculation of figure of merits, such as the mitigation gain.

BandSplit.m does the detection of the spectral location of the strong blocker signal and splits the input into a desired and a reference signal.

RefModel.m computes the distortion estimates by using a specific reference signal and performs the subsequent filtering to provide only the added distortion products in the band of interest to the AF, without the actual blocker signal.

AFilt.m holds the AF and adjusts the distortion estimates in amplitude and phase to those in the desired signal. It is the actual core of the algorithm and allows for choosing different implementations, such as LMS, NLMS, and recursive least square [Hay02].

Block.m is a parent class for all blocks and holds very general properties, such as the FFT size and axis limits for plotting. All other blocks are derived from this (inheritance).

C.2 USRP MATLAB Interface

This USRP MATLAB interface was developed at the department "Electronic Measurement Research Lab" and is used for all offline measurements presented in this thesis. It is composed of a MATLAB class and a console application (arg-interface), as illustrated in Figure C.2. The interface follows the idea of a "USRP recording service", i.e. it receives a finite number of samples from the USRP and stores them on the host PC together with some other files

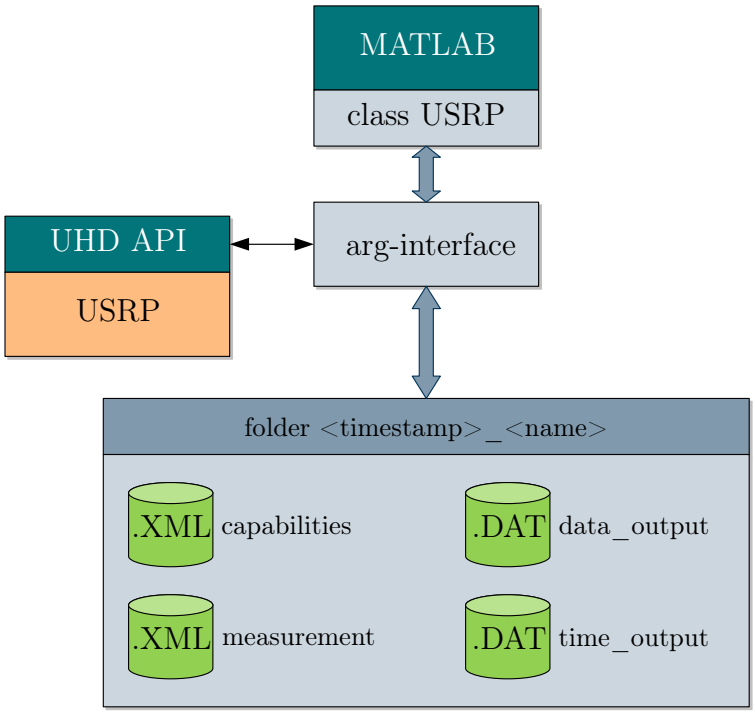


Figure C.2: Software framework of the USRP MATLAB interface.

containing meta information about the measurement. Thereby, all device parameters can be directly set from MATLAB. Eventually, the measurement data including the device's configuration during measurement time can be read back to the MATLAB workspace at any time for further processing of the data.

The arg-interface mainly calls methods of the universal hardware driver (UHD), which denotes an application programming interface (API) for USRP devices [Ett]. They are both written in C++. The arg-interface implements all low-level functions of the interface and can be parametrised by command line options. On the application level, MATLAB employs an abstract class that generates these command line parameters and calls the arg-interface. All intelligence, especially the error handling, is moved into this class. At user level, all parameters of the USRP can be simply invoked by the object properties.

During data acquisition, the arg-interface stores a set of files to the hard disk. First, there are two binary .dat files containing the I/Q samples (data_output.dat) and the timestamp information provided by the UHD API (time_output.dat). Second, there are two .xml files holding meta information about the general device parameters (capabilities.xml) and the measurement configuration (measurement.xml).

The software project has been continuously developed and successfully applied also in other physical layer-related projects that base on USRP devices.

Theses

1. Modern communication systems impose tough requirements on the transceiver's electronics that reach far beyond the state-of-the-art technology.
2. The analogue stages of a radio's RF front-end are inherently imperfect. Achieving a sufficient linearity, among other RF impairments, is a challenging issue in the practical receiver design.
3. Circuit non-linearity is even more crucial in wideband receivers with limited selectivity, such as those used as software defined radios (SDRs).
4. Coexisting heterogeneous wireless systems lead to scenarios with weak desired and strong unwanted (blocking) signals. These blocking signals can easily enter front-end amplification and mixing stages.
5. Receiver non-linearity generates additional frequency components that may fall into free frequency bands or hit weak desired signals. Thereby, odd-order intermodulation distortions is of major concern.
6. Receiver non-linearity can be handled by design optimisation, cancelling distortion-producing signals, or by pre- or post-correction (mitigation) techniques.
7. System-level "Dirty RF" signal processing in the digital domain is a popular alternative to alleviate non-linear distortions in the receiver.
8. Mitigating non-linear distortions at the *receiver* is more challenging than at the *transmitter*, especially due to the presence of multiple unknown signals with different power levels and dynamics due to the radio channel.
9. A digital feedforward mitigation algorithm is employed that splits the received signal in a desired and a reference signal, re-generates the distortion products, and adaptively subtracts them from the desired signal.
10. Feedforward mitigation decreases the bit error rate (BER) of distorted signals, and minimizes the energy of non-linearly induced interference. Finally, the effective linearity of the front-end and the dynamic range is increased.

11. Behavioural modelling of non-linear RF front-ends is preferred to physical modelling. However, behavioural models need to match the physical front-end architecture and are to be verified with different data sets and measurements.

12. A cascaded non-linear model is developed that fully matches to the topology of a direct-conversion receiver and addresses all generated distortions at the RF and baseband (BB) stages.

13. No detailed knowledge of the non-linearity profile and no exact polynomial coefficients are necessary. The adaptive filter (AF) stage adjusts the reference distortions in amplitude and phase, in order to finally subtract them from the desired signal.

14. The static reference non-linearity model and the adaptive filtering stage form a kind of adaptive Hammerstein model that takes also linear memory effects into account.

15. A typical low-cost SDR under test, the Universal Software Radio Peripheral (USRP), is modelled by simple memoryless polynomials derived by two-tone measurements. This type of modelling is sufficient for distortion mitigation application.

16. Measurements with wideband WCDMA signals have revealed that the USRP manifests little memory effects of order $M = 2$, thus, same AF length is utilised in the distortion mitigation architecture.

17. The feedforward mitigation principle works independently of the desired signal waveform and processes only the strong blocker signal. From theory, the mitigation architecture works for arbitrary blocker signals.

18. Extensive simulations and real-world RF measurements demonstrate the effectiveness of the algorithm based upon two-tone and modulated BPSK signals.

19. The algorithm has been proven to work even under frequency-selective and fast fading conditions imposed by wave propagation in a real radio environment.

20. The purely digital algorithm can be easily integrated on an FPGA with enhanced DSP features. Thereby, the receiver self-regulates its non-linearity in real-time and immediately provides the corrected I/Q samples for the following BB processing.

21. The feedforward technique has been successfully applied in cognitive radio (CR) for improving the reliability of spectrum sensing, in GSM communication for improving the BER performance of distorted GSM channels, and in GSM-based passive radar for enabling weak target detection.