I/Q Imbalance Compensation in Quadrature Modulators

Bhardwaj, Aparna Bhisham Chander

http://hdl.handle.net/11023/3358
master thesis

University of Calgary graduate students retain copyright ownership and moral rights for their thesis. You may use this material in any way that is permitted by the Copyright Act or through licensing that has been assigned to the document. For uses that are not allowable under copyright legislation or licensing, you are required to seek permission.

Downloaded from PRISM: https://prism.ucalgary.ca
Abstract

Transmitter architectures in radio and wireless communications are highly susceptible to gain and phase imbalances triggered by the analog quadrature modulator. An efficient yet, simple adaptive approach for estimating these impairments has been proposed comparing the instantaneous powers of the output and input signal without interrupting the transmission mode using a batch-gradient descent algorithm. The proposed methodology is tested for different values of I/Q imbalances and different batch-size in both simulations and on hardware platform. The technique shows a significant suppression of image with an image to signal ratio as low as 50 dB for a batch-size as low as 1000-1500 with the inclusion of noise (Signal-to-noise ratio = 60 dB). The simulation and hardware results at a carrier frequency of 2.14 GHz validate the effectiveness of the proposed approach. The results have been compared with a state-of-the-art method in literature and show improved results, robustness with lesser computational complexity.
Acknowledgements

I take this opportunity to thank my supervisor, Dr. Ghannouchi for placing his belief in me and having given me this chance to pursue the Master’s degree in the first place. His willingness to provide all the available and appropriate resources helped me immensely in concluding my work timely. The iRadio Lab has been the starting point of my learning curve here in Canada and is very special. I can’t thank Dr. Mohammed Helaoui enough for providing me the timely support on countless occasions and pointed feedback which gave a beautiful shape to the conclusion of my thesis. My mentor, Ph.D. candidate, Abubakr Abdelhafiz has been extremely motivating and positive all through the period which kept me going even when I found myself at junctures I thought I could progress no further. Thanks to Dr. Andrew Kwan for helping me carry out the experiments in the lab smoothly. My lab colleagues, Dr. Mayda and Ph.D. candidates, Mohsin Aziz Baig and Abul Hassan who always came forward to help when I was conceptually in doubt.

Parents and siblings can’t be missed out on the moments who motivate you when you miss them the most, away from home. I found my new family in my Fiancé Kartik who helped me in the toughest periods of stress by providing me with all the emotional strength and motivated me to bounce back with force. My Uncle, Ashok to whom I would end up giving long frantic calls when I needed to hear the most positive of words and for whom I can’t express my gratitude in any form. I have to thank my guide and supervisor from the previous firm, Dr. Gopal Patro for grooming me in the design of high power amplifiers. Also, I express my heartfelt gratitude to my close friend Dimple who made it possible for me to have a comfortable stay in Canada. I can’t thank each and everyone enough, who made this possible for me. Cheers and many thanks to all my friends in Calgary for being the cheerful and regular stress busters in between the stressful periods of research.
To

My parents and the Almighty who have always taken care of me despite all of my doubts
Table of Contents

Abstract ........................................................................................................... ii
Acknowledgements ......................................................................................... iii
Table of Contents .......................................................................................... v
List of Tables .................................................................................................. vii
List of Figures and Illustrations ........................................................................ viii
List of Abbreviations and Nomenclature ........................................................ x
List of Symbols ............................................................................................. xi

CHAPTER ONE: INTRODUCTION ................................................................. 1
1.1 Motivation ................................................................................................ 1
1.2 Objective and Outline of thesis .............................................................. 3
1.3 I/Q Baseband Signal and Bandpass transmission .................................... 4
1.4 Transmitter Fundamentals .................................................................... 5
1.4.1 Types of Transmitter Architectures .................................................. 6
1.4.2 Effect of Imbalances on Tx Architectures: ......................................... 8
1.4.3 Figures of merit .................................................................................. 11

CHAPTER TWO: REVIEW OF EXISTING TECHNIQUES ............................. 16
2.1 Methods: ............................................................................................... 16
2.1.1 Receiver feedback based methods: .................................................... 16
2.1.2 Envelope based methods: ................................................................. 16
2.2 Comparison and Observations with an existing method: ....................... 19
2.3 Conclusion ............................................................................................. 23

CHAPTER THREE: CLOSED-LOOP IMPAIRMENT COMPENSATION USING BATCH
GRADIENT DESCENT ON THE REAL ENVELOPES OF THE SIGNAL .... 24
3.1 Introduction ............................................................................................. 24
3.1.1 Stochastic Gradient Descent ............................................................. 24
3.1.2 Gradient Descent ............................................................................. 25
3.2 Pre-compensation Block and Imbalance Block ...................................... 25
3.3 Cost-function and mathematical analysis: ............................................. 27
3.4 Conclusion ............................................................................................. 33

CHAPTER FOUR: SIMULATION AND IMPLEMENTATION RESULTS .......... 34
4.1 Simulation Results .................................................................................. 34
4.1.1 Performance analysis for different number of samples: ..................... 35
4.1.2 Comparison with a state-of-the-art technique: .................................... 39
4.2 Measurement Results: .......................................................................... 40
4.3 Conclusion: ........................................................................................... 46

CHAPTER FIVE: CONCLUSION AND FUTURE SCOPE ............................... 47
5.1 Conclusion and Summary ....................................................................... 47
5.2 Contributions: ...................................................................................... 48
5.3 Future Scope: ....................................................................................... 48
List of Tables

Table 4.1: Summary of performance for different number of samples ........................................ 39
Table 4.2: Summary of performance comparison between two techniques ................................. 40
Table 4.3: Performance for different imbalances for 100K samples ......................................... 44
Table 4.4: Summary of performance for different number of samples (in measurement); \( g = 0.5 \text{ dB} \ (1.0592), \varphi = -5^\circ \) ......................................................................................................................... 46
List of Figures and Illustrations

Fig 1.1: Proposed Adaptation Block .................................................................................. 3
Fig 1.2: Transmitter Chain .................................................................................................. 6
Fig 1.3: Direct Conversion Modulator Chain ...................................................................... 7
Fig 1.4: Super-heterodyne Modulator Chain ..................................................................... 8
Fig 1.5: DC Offset in a modulated spectrum ..................................................................... 9
Fig 1.6: Spectrum showing ISR for different values of imbalances .................................. 10
Fig 1.7: Phase noise ........................................................................................................... 11
Fig 1.8: EVM Diagram ..................................................................................................... 12
Fig 1.9: Image Interference for (a) Low-IF Tx and (b) DC Tx ........................................... 14
Fig 2.1: Convexity Plot for (a) 100K (b) 10K and (c) 1K samples ..................................... 21
Fig 2.2: Convexity Plot of gain for Fourth Order Moment for (a) 100K and (b) 10K samples [21] ...................................................................................................................... 22
Fig 2.3: Convexity Plot of phase for Fourth Order Moment for 100K samples [21] .......... 22
Figure 3.1: Pre-compensation and Imbalance Block ......................................................... 26
Figure 3.2: Convergence plots of various values of $\mu_\beta$ for a given value of $\mu_\alpha$ .... 31
Figure 3.3: Convergence plots of various values of $\mu_\alpha$ for a given value of $\mu_\beta$ ....... 31
Figure 4.1: Estimation Error of $\alpha$ .................................................................................. 35
Figure 4.2: Estimation Error of $\beta$ .................................................................................. 36
Figure 4.3: Imbalance parameter estimates ...................................................................... 36
Figure 4.4: Convergence of Cost-Function $J(n)$ ............................................................. 37
Figure 4.5: ISR and NMSE .............................................................................................. 37
Figure 4.6: CDF plots of ISR for different number of samples .......................................... 38
Figure 4.7: CDF Plot of ISR compared with [21] .............................................................. 40
Figure 4.8: Post-estimation Block ..................................................................................... 41
Figure 4.9: Post-estimation and pre-compensation Procedure ......................................................... 42

Figure 4.10: (a) Imbalance Output (b) Compensated Output for 2 values of imbalances ........ 44

Figure 4.11: ISR Spectrum Plots for different number of samples; $g = 0.5 \, dB \, (1.0592), \phi = -5^\circ$ ..................................................................................................................... 45
**List of Abbreviations and Nomenclature**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWGN</td>
<td>Additive-White Gaussian Noise</td>
</tr>
<tr>
<td>BPF</td>
<td>Band-Pass Filter</td>
</tr>
<tr>
<td>DAC</td>
<td>Digital-Analog Converter</td>
</tr>
<tr>
<td>DC</td>
<td>Direct-conversion</td>
</tr>
<tr>
<td>DSP</td>
<td>Digital Signal Processing</td>
</tr>
<tr>
<td>EVM</td>
<td>Error Vector Magnitude</td>
</tr>
<tr>
<td>FPGA</td>
<td>Field-Programmable Grid Array</td>
</tr>
<tr>
<td>IC</td>
<td>Integrated Circuit</td>
</tr>
<tr>
<td>IF</td>
<td>Intermediate Frequency</td>
</tr>
<tr>
<td>I/Q</td>
<td>In-phase/Quadrature</td>
</tr>
<tr>
<td>ISR</td>
<td>Image to Signal Ratio</td>
</tr>
<tr>
<td>LMS</td>
<td>Least-Mean Squares</td>
</tr>
<tr>
<td>LPF</td>
<td>Low-Pass Filter</td>
</tr>
<tr>
<td>LTE</td>
<td>Long-term Evolution</td>
</tr>
<tr>
<td>NMSE</td>
<td>Normalized Mean Square Error</td>
</tr>
<tr>
<td>OFDM</td>
<td>Orthogonal Frequency-Division Multiplexing</td>
</tr>
<tr>
<td>PA</td>
<td>Power Amplifier</td>
</tr>
<tr>
<td>PCB</td>
<td>Printed Circuit Board</td>
</tr>
<tr>
<td>QAM</td>
<td>Quadrature Amplitude Modulation</td>
</tr>
<tr>
<td>RF</td>
<td>Radio Frequency</td>
</tr>
<tr>
<td>RLS</td>
<td>Recursive-Least Squares</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>α, β</td>
<td>Imbalance coefficients</td>
</tr>
<tr>
<td>ñ, ß</td>
<td>Estimate of imbalance coefficients</td>
</tr>
<tr>
<td>g</td>
<td>Gain imbalance</td>
</tr>
<tr>
<td>φ</td>
<td>Phase imbalance</td>
</tr>
</tbody>
</table>
CHAPTER ONE: INTRODUCTION

In the past few years, radio and digital wireless communication has seen an unimaginable growth, thanks to the availability of highly integrated digital transceiver integrated circuits (ICs) with built-in digital signal processing (DSP) engines, it has resulted into the advent of Software Defined Radio (SDR), in the form of multi-technology smart phones available with almost everybody. SDR, as the name suggests is an advanced architecture scheme where software defines/chooses controls blocks of hardware/firmware software inside the SDR entity. With the ever-growing consumers in the wireless market, the need for spectrally efficient standards is on the rise. However, with the complex modulation schemes and wide band signals comes the challenge of eliminating the vulnerability of the transceiver systems utilizing these standards and effectively retrieving the original signal without any infidelity.

1.1 Motivation

Up-conversion architectures like: direct conversion (DC), super-heterodyne and low intermediate-frequency (IF) in the transmitter chain are utilized for transmitting the signal for various signal standards [1]. Direct conversion chain comes with the challenge of handling the local oscillator (LO) imbalances introduced by a quadrature modulator. The analog compensation techniques to eliminate these imbalances, give us the capability of controlling the changes anywhere in the system through hardware means. However, hardware changes bring along with them the investment of cost, particularly if the hardware is very application and frequency specific. The costly analog compensation techniques have taken a step back with the availability of efficient digital compensation techniques which are cost-effective.
The expanse of literature to mitigate the frequency dependent and independent defects in the transmitter chain is immense [2,5,11-21]. The thesis here confines itself to the frequency independent case since we reasonably assume that the in-phase/quadrature-phase (I/Q) impairments pre-dominantly contribute to the imbalances in a modulator and are constant over the frequency-bandwidth [2],[3]. It becomes extremely important to take into account if adaptively, the digital pre-compensation techniques can accomplish the desired results in the minimum processing time with fewer samples to avoid the computation costs. A disadvantage of going with higher order moments is that they require longer data lengths when finding the accurate estimates of time-averaged signals [4]. The thesis considers a batch of the incoming data to calculate the averaged cost-function in our algorithm. The idea behind going for a receiver-less feedback chain for digital compensation and a power dependent scheme is to eliminate the need for a quadrature demodulator or a mixer and hence, the associated LO related imbalances which add to the existing coefficients for estimation. Additionally, the problem of phase-shift on the demodulated signal gets completely eliminated [2].

The modulator imbalances of gain and phase manifest themselves as an image signal overriding the original signal in case of direct conversion transmitters and as a signal appearing at ‘f_{RF} − f_{IF}’ if the original signal appears at ‘f_{RF} + f_{IF}’ in low IF transmitter architectures, where f_{RF} is the modulating carrier frequency and f_{IF} is the IF shift. The appearance of this image distorts the quality of the signal which can be quantified in terms of the image to signal ratio (ISR), the constellation diagram, error vector magnitude (EVM), normalized mean square error (NMSE), etc. It is important to carry out the adaptation of gain and phase simultaneously and iteratively as, an optimum value is reached by adjusting each parameter simultaneously until the image cannot be further suppressed.
1.2 Objective and Outline of thesis

This thesis aims at adaptively solving the $I/Q$ imbalance problems in a quadrature modulator by minimizing the mean square of the instantaneous power difference between output and input signal as shown in Fig. 1.1. This thesis will aim to reduce any envelope fluctuations that may exist as a result of the present gain and phase imbalances using a gradient based approach by justifying that a unique solution exists. Essentially, envelope based techniques (e.g. the use of diode detectors) can work with both, DC and low-IF architectures without the need for any mixer or quadrature demodulator based circuits [2]. The algorithm jointly compensates for the gain and phase imbalance as a single parameter with two coefficients using a single cost-function [5].

The focus of research work in this thesis has been to propose a direct-learning of inverse model approach through the use of a simple algorithm. Chapter 1 talks about the fundamentals of the transmitter architectures, their associated impairments, and the performance metrics. Some performance metrics have been introduced to quantify the quality of the transmitted signal. Chapter 2 gives a brief literature review of the available techniques on receiver and envelope based methods.

Fig 1.1: Proposed Adaptation Block
for digital $I/Q$ compensation. Chapter 3 succinctly discusses the theory of the two widely used algorithms followed by a mathematical analysis of the power-dependent cost-function which is the target criterion to be minimized for achieving the best ISR utilizing a batch-type gradient descent technique. Chapter 4 comprehensively justifies the chosen technique for a varied number of samples using a standard Long Term Evolution (LTE) 5 MHz wideband signal with its sub-carriers utilizing 64 quadrature amplitude modulation (QAM) technique. In the last part of the chapter, simulation results for 100K samples have been shown to be closely coherent with the experiments using a makeshift in-built transmitter-receiver (Tx-Rx) system of vector signal generator (ESG 4438C). Chapter 5 discusses and concludes the results obtained in the thesis along with the contributions of the research work and the possible extension and scope for the future work.

1.3 $I/Q$ Baseband Signal and Bandpass transmission

The representation of complex two-dimensional signals in communications has been a natural choice to consider for ease of signal processing. These complex signals consist of nothing but a pair of two real-valued signals and projection on the perpendicular axis representing the pair value. We can consider them as two-dimensional or quadrature signals. In the frequency domain, the real-valued signals occupy twice the signal bandwidth as can be seen by the symmetry in their spectra while the complex signals on the other hand utilize the bandwidth efficiently without holding the symmetry [6]. The baseband signal is up-converted to the carrier frequency, $f_{RF}$ for transmission. A high carrier frequency reduces the antenna length for transmission. This translation from baseband to a pass-band signal in frequency domain appears as spectral components at $' - f_{RF}'$ and $' + f_{RF}'$ which essentially contain the entire information about the input baseband signal. On the receiver side, the radio-frequency (RF) signal is down-converted to baseband to allow the
sample frequency to be greatly reduced for processing. Say, $x(t)$ is a complex-baseband representation of the input signal, up-conversion of $x(t)$ to an RF passband signal centered around $f_{RF}$. $y(t)$ at the transmitter output can be represented as:

$$y(t) = \text{Re}[x(t)e^{j2\pi f_{RF}t}]$$

The two components of the input baseband signals, $x_I(t)$ and $x_Q(t)$, are retrieved from the RF signal, $y(t)$ at the receiver output by multiplying it with $\cos(2\pi f_{RF}t)$ and $\sin(2\pi f_{RF}t)$ for I and Q components separately, represented as $x_{B,I}(t)$ and $x_{B,Q}(t)$ respectively, and then passing through the low-pass filters (LPFs). The baseband components obtained at the receiver output from $y(t)$ can be represented separately for I and Q components as:

$$x_{B,I}(t) = (x_I(t) \cos(2\pi f_{RF}t) - x_Q(t) \sin(2\pi f_{RF}t)) \cos(2\pi f_{RF}t)$$

$$= (x_I(t) \cos(2\pi f_{RF}t) - x_Q(t) \sin(2\pi f_{RF}t)) \cos(2\pi f_{RF}t)$$

$$= \frac{1+\cos(4\pi f_{RF}t)}{2} x_I(t) - \frac{\sin(4\pi f_{RF}t)}{2} x_Q(t) \quad (1.1a)$$

On passing through the LPFs, $\cos(4\pi f_{RF}t)$ and $\sin(4\pi f_{RF}t)$ terms get filtered. Similarly, for the baseband Q-component on multiplication with $\sin(2\pi f_{RF}t)$, the baseband equivalent is:

$$x_{B,Q}(t) = \frac{\sin(4\pi f_{RF}t)}{2} x_I(t) - \frac{1 - \cos(4\pi f_{RF}t)}{2} x_Q(t) \quad (1.1b)$$

This transformation becomes important because the up-conversion and down-conversion are the essences of a transceiver system.

1.4 Transmitter Fundamentals

A transmitter consists of a digital and analog block, shown in Fig. 1.2. Digital block consists of an interpolation filter for raising the sampling frequency. It may also consist of a digital up-conversion
block for translating the baseband signal to digital IF. It is followed by an $I/Q$ digital-to-analog converter (DAC). The analog outputs are passed through reconstruction LPFs in order that the Nyquist images be removed.

**Fig 1.2: Transmitter Chain**

The analog $I/Q$ modulator in the transmitter chain up-converts this baseband signal for transmission on the carrier by multiplying the quadrature components of the split LO with the baseband signal. The RF signal can either be directly converted from baseband (DC transmitters) or from an intermediate frequency (low-IF Transmitters). The output after up-conversion must ideally replicate the baseband signal without any infidelity in order for the receiver to sustain the information contained in the original signal. However, non-ideal gain and phase mismatches between the in-phase and quadrature phase branches give rise to the frequency-independent imbalances which manifest themselves as either a signal overriding the fundamental signal (dc transmitters) or appear as an additional image for low-IF transmitters. The DACs and LPFs in the $I$ and $Q$ branches contribute towards the frequency-dependent impairments due to the difference in impulse responses of these components of the two branches.

1.4.1 Types of Transmitter Architectures

1.4.1.1 Direct Conversion

Direct up-conversion transmitters are the favored choice when looking for simplicity and low-cost implementation in comparison with a super-heterodyne architecture. The quadrature up-converter
generates the waveform at RF suitable for transmission. This compact architecture comes with its share of challenges like I/Q mismatches between the in-phase and quadrature phase branches, LO injection pulling and LO leakage. The unequal gains and the deviation of the 90-degree difference between the LOs of the I and Q paths depend on the operating temperature. The power amplifier (PA) output in the transmitter chain contains a lot of spectral components due to non-linearity around the LO frequency which leaks back to the voltage-controlled oscillator causing pulling [7]. The LO leakage results due to the difference in dc bias levels between the in-phase and quadrature phase components of the baseband signal at the input of the quadrature modulator leading to in-band interference which cannot be filtered. Despite the aforementioned challenges, direct-conversion transmitters (Fig 1.3) are the favored choice due to their simple architecture as they obviate the IF components unlike super-heterodyne [8].

![I/Q Baseband Signal Processing](image)

**Fig 1.3: Direct Conversion Modulator Chain**

1.4.1.2 Super-Heterodyne Architecture

This dual up-conversion architecture has an advantage of no LO pulling and it is relatively easier to design since controlling of the signals can be done at an IF level by filtering the undesired
frequency components. The presence of IF band-pass filter (BPF) and RF surface-acoustic wave (SAW) filters at the output of the first IF mixer and second RF mixer respectively removes the unwanted frequency products from the output of these mixers (Fig 1.4). However, this architecture requires high quality-factor components, more complexity, and effort in careful frequency planning, which makes this architecture economically unviable.

![Fig 1.4: Super-heterodyne Modulator Chain](image)

1.4.2 Effect of Imbalances on Tx Architectures:

1.4.2.1 DC Offset

A modulated signal results at the output of the modulator when the baseband quadrature signals multiply with the LO signal. The presence of dc offsets in these in-phase and quadrature-phase arms, cause the unmodulated LO signals to appear at the output of the modulator at the center of the desired bandwidth (Fig 1.5). A finite isolation between the RF and IF ports of a mixer in addition to the printed circuit board (PCB) component tolerances like- parasitic capacitances and bond- wire to bond-wire coupling add to the contribution of the net LO feedthrough [1].
1.4.2.2 Gain and Phase Imbalance

The gain and phase imbalances in a modulator occur due to amplitude and phase mismatch respectively between the modulating signals of the I and Q baseband data. The extent of the impact of gain and phase mismatches in the two arms of the quadrature modulator depend on the type of the modulation signal being used and the bandwidth of the signal. Higher complexity modulation schemes such as 16-QAM, 64-QAM, 128-QAM consisting of a number of constellation points in close proximity are more susceptible to imbalances due to incorrect symbols received. In the presence of the imbalances, the output of modulator becomes:

\[ y(t) = x_i(t) \cos(\omega_{RF}t) - g x_q(t) \sin(\omega_{RF}t + \phi) \]

where \( g \) and \( \phi \) are gain and phase imbalance respectively. As mentioned before, the effect of these imbalances is clearly visible in the form of a signal appearing at an adjacent channel (low-IF) as shown in Figure 1.6 for different levels of imbalances.
1.4.2.3 Phase Noise

Phase noise in a local oscillator is caused by an unstable LO due to the presence of various components of frequencies. An LO is essentially a carrier with noise oscillating around it. A carrier is a single harmonic wave but noise consists of several frequencies. The output signal along with multiple frequencies having a random phase with respect to the LO gets amplified at the output of the transmitter due to multiple stages of amplification and manifests as a raised noise floor in receivers (fig 1.6) [9].
1.4.3 Figures of merit

To evaluate the performance of a transmitted/corrected signal, it is important to introduce widely used performance metrics present in literature. These include NMSE, adjacent channel power ratio (ACPR), EVM and ISR. NMSE and EVM are well accepted figures of merit for evaluating the performance of methods that aim to alleviate various transceiver imperfection such as PA nonlinearity, dc offset, phase noise, etc. However, for the purpose of evaluating the performance of methods aiming to deal with I/Q imbalance for Low IF transmitters, ISR is a very useful figure of merit, since image suppression can be easily seen by plotting the spectrum of the corrected signal. For instance, for the method used in [5], the authors use ISR as their target cost function to mitigate the resulting imbalances for low IF topology.

1.4.3.1 Error Vector Magnitude

The impairments in a transmitter cause the constellation points to shift from their ideal locations as a result of which the symbols are distorted after demodulation, as shown in Fig 1.8. An EVM essentially measures the deviation of how far these points get dislocated upon passing through a
transceiver due to the phase and the magnitude error between the transmitted and received symbols. The sources of degradation of a digitally modulated signal can be easily traced with the help of an error vector between the measured and the ideal vector in the I/Q plane. It can be either expressed in decibels (dB) or percentage (%).

\[
EVM_{RMS}(\%) = \frac{\sum_{n \in N}[(I_{T,n} - I_{R,n})^2 + (Q_{T,n} - Q_{R,n})^2]}{\sum_{n \in N}[I_{R,n}^2 + Q_{R,n}^2]} \times 100
\]

(1.2)

\(I_{T,n} + jQ_{T,n}\): Transmitted 'nth' Symbol Vector

\(\hat{I}_{R,n} + j\hat{Q}_{R,n}\): Received 'nth' Symbol Vector

**Fig 1.8: EVM Diagram**

For N symbols, root mean square (RMS) or normalized EVM is defined as the square root of the ratio of mean error vector power between measured and ideal symbol to the mean of the reference symbol power [10].
1.4.3.2 ISR

The effectiveness of the proposed approach to eliminate the impairments is demonstrated using an input IF-shifted signal at 5 MHz in simulations and hardware. For this purpose, ISR has been used throughout the thesis to quantify the imbalances in a transmitter. The presence of impairments in a non-ideal transceiver lead to a finite side-band image rejection ratio and an unwanted side-band appearance for low-IF transmitters and a coinciding image with the fundamental signal in case of direct-conversion transmitters (Fig 1.9).

For an input digital baseband signal, $x(t)$, the baseband equivalent modulator output $y(t)$ can be modelled at the baseband as [5]:

$$y(t) = K_1 x(t) + K_2 x^*(t)$$  \hspace{1cm} (1.3)

Where $K_1$ and $K_2$ are the imbalance coefficients defined as:

$$K_1 = \frac{1}{2} \left( 1 + ge^{j\phi} \right) \text{ and } K_2 = \frac{1}{2} \left( 1 - ge^{-j\phi} \right);$$

$g$: gain imbalance between the in-phase and quadrature phase branch of the modulator, $\phi$: phase imbalance between the LO arms; $x(t)$: desired signal, $x^*(t)$: image signal.
Fig 1.9: Image Interference for (a) Low-IF Tx and (b) DC Tx

$K_2$ contributes to the image component in the desired band which leads to a finite rejection between the $x(t)$ and $x^*(t)$ defined by:

$$ISR(dB) = 10 \log_{10} \left( \frac{|K_2|^2}{|K_1|^2} \right)$$  \hspace{1cm} (1.4)

On substitution of the values in equation (1.4) from (1.3),

$$ISR(dB) = 10 \log_{10} \left( \frac{1 + g^2 - 2g \cos \phi}{1 + g^2 + 2g \cos \phi} \right)$$  \hspace{1cm} (1.5)

In the frequency domain, we define it as the ratio of integrated power in the useful bandwidth of the image to that of the signal. We can see from Fig 1.10 that different gain and phase imbalances lead to different levels of image rejection.
1.4.3.3 **NMSE**

An alternate way of defining a signal quality is by comparing the output signal, $y(n)$, with respect to the input signal, $x(n)$. It is defined as:

$$NMSE \ (dB) = 10 \times \log_{10} \left( \frac{\sum_{n=1}^{N} |y(n) - x(n)|^2}{\sum_{n=1}^{N} |x(n)|^2} \right)$$  \hspace{1cm} (1.6)

where, $y(n) = Ax(n)e^{j\Delta \theta}$; $A$: gain of the system, $\Delta \theta$: phase shift between $y(n)$ and $x(n)$ due to time delay in the measurement circuit.
CHAPTER TWO: REVIEW OF EXISTING TECHNIQUES

Many techniques have been reported for mitigating frequency-dependent impairments in a quadrature modulator [2, 11, 13]. Thesis focusses on catering to the frequency-independent case which reasonably assumes the LO signal to be a continuous wave exhibiting a non-modulated behavior.

2.1 Methods:

2.1.1 Receiver feedback based methods:


2.1.2 Envelope based methods:

Special test signals for assessing the estimate values have been used by various authors like Authors in [13] use special test-signals to generate intermodulation products to determine the frequency-independent impairments. Further, a time-division method is used for transmitting the in-phase and quadrature-phase LO distinctly mixed with I and Q signals which are used for the self-demodulation of I and Q branch separately to determine the frequency-dependent impairments. Reference [14] uses amplitude of the various frequency components formed by
feeding carefully selected sinusoidal test signals to the DC transmitter arm. Through simple mathematical calculations on the envelope detector output, the amplitude terms for each component are separated to predict the imbalance parameters. A similar process is performed for extracting the third-order intercept of the PA with the same test-signals as used for determining the linear impairments in a modulator but, at higher power. The author in [15] estimates the values of modulator imbalances using mathematical models and equations. The techniques in [16] and [17] use least squares on the detector output to extract the estimate of the imbalance values. The inverse operation for estimation is computationally heavy on the system especially when the number of samples is very large. Various adaptive techniques have been quoted for estimating and compensating the impairments. An RLS technique has been utilized in [18] for equalizing the actual envelope output with the imbalance modeled envelope output. The cost-function again being sample based and being dependent on the input signal statistics requires an accurate estimation of the system gain. Stochastic gradient has garnered a lot of attention owing to the simplicity of its computation and online implementation. Gradient Descent on the other hand, involves the use of the entire training dataset for the calculation of weight coefficients. Stochastic gradient involves a careful choice of step-size owing to peak changes due to the random motion of cost-function with practical systems containing noise. This places a limitation on the minimum a cost-function can reach thereby leading to a continuous oscillation even after reaching the optimum point. A convenient trade-off between the stochastic-gradient and gradient-descent, a batch gradient descent compensates for the large fluctuations in standard least mean squares (LMS) cost-function and the higher computation complexity of the Gradient Descent. Instead of adapting with a ‘true’ gradient, it is a reasonable assumption to calibrate the system with a segment of data to time-average the cost-function for a certain batch and consequently update the coefficients until the next
batch appears. The time-averaging greatly reduces the impact of noise when working with highly noisy systems. The stochastic approach can be a cumbersome job to closely track the variations in the signal characteristics due to the presence of imbalance and noise. Several proposed and implemented methods have gone by the conventional and simple LMS methods. Reference [19] mentions several theoretical methods for estimating the imbalances with different cost-functions for gain and phase using LMS along with an improved proposed technique, however, practical implementation results have not been presented to demonstrate the strength of the proposed algorithm. The cost-functions in [18] and [19] require an accurate estimate of a gain of the system which can be easily overcome when using gradient descent by normalizing the cost-functions. The author in [20] does a sample based LMS adaptation for estimation of imbalance coefficients which essentially being samples based, can be more prone to a noisy gradient (as seen by the impact of noise in the proposed approach when batch-size is reduced) due to the non-averaging of the cost-function. However, there is a certain ambiguity in terms of the performance of the proposed approach in the above sample-based algorithms in the presence of noise in simulations and hardware [18, 20]. Reference [2] comprehensively solves for all the set of impairments with promising results by separately solving for gain and phase imbalances. The gain imbalance, which the authors have incorporated in the filter response itself involves computations by iteratively adapting the filter response (containing memory) and then inverting the coefficients in each iteration to determine the compensation filter coefficients which makes it a complex approach. It can be seen that all the authors in [2, 18, 20] make use of the forward model, comparing the actual envelope output with the forward modeled imbalance envelope output and inverting the imbalance coefficients to determine the inverse coefficients. This forward-modelling makes the computation complex when incorporating the memory effects and, a direct-determination of the inverse
coefficients is a favorable choice. The research work in the thesis makes use of a direct learning architecture of the inverse model comparing the envelope of the output signal with the error-free input envelope as the desired error criterion.

2.2 Comparison and Observations with an existing method:

The method in [21] uses a gradient descent based on a second-order moment to estimate the dc offsets and a fourth-order moment of the quadrature modulator output to estimate gain and phase imbalances which has shown to deliver a good performance for a large batch-size but, the proposed second-order moment approach outperforms this fourth-order moment method in terms of the best-attained ISR for the same batch-size, simplicity in terms of the number of samples needed for processing and the robustness in terms of the stability of the performance for any batch of incoming data (The observations have been explained under the simulation results section of Chapter 4). Convexity plots in Figs 2.1 and 2.2 summarize the observations on the performance for a second-order moment and fourth order moment respectively for a gain imbalance and phase imbalance of 0.5 $dB$ and $-5^\circ$ respectively. As it is clear from the figure 2.2 that, the convexity of the function highly degrades the moment the sample numbers is reduced resulting into many local minima which manifests as a residual estimation error. Convergence of phase imbalance requires the gain imbalance to have been estimated before [21]. It is clearly seen that for the proposed method, a simultaneous adaptation of both the parameters can easily estimate the values, shown in Fig 2.1. A second order cost-function begins to show a degradation for a batch containing around 1000-1500 samples. From the figures 2.1 and 2.2, it is evident that, a sample-based adaptation will be highly susceptible to the noisy fluctuations in the signal leading to an unstable convergence. Hence, the thesis has focused on a ‘batch’ gradient adaptation.
Surf Plot of Cost Function for 100K Samples

Surf Plot of Cost Function for 10K Samples

Cost Function: \( J(\alpha, \beta) \)
Fig 2.1: Convexity Plot for (a) 100K (b) 10K and (c) 1K samples
Fig 2.2: Convexity Plot of gain for Fourth Order Moment for (a) 100K and (b) 10K samples [21]

Fig 2.3: Convexity Plot of phase for Fourth Order Moment for 100K samples [21]
2.3 Conclusion

The first part of this chapter lists the various prior research works done in the field of digital compensation of I/Q imbalances. The mentioned references describe their proposed approach and the differences between those existing approaches and the proposed method in this research are highlighted. A fair comparison using preliminary surf plot runs for different batch sizes indicate the possible degradation in performance that can be more prevalent for a fourth-order moment approach when compared with a second-order moment approach. The observations have been made in the presence of noise to determine any estimation error that may exist as the batch size is reduced.
CHAPTER THREE: CLOSED-LOOP IMPAIRMENT COMPENSATION USING BATCH GRADIENT DESCENT ON THE REAL ENVELOPES OF THE SIGNAL

3.1 Introduction

The main hypothesis of this research is to first validate that the proposed solution to the existing problem in a quadrature modulator can be modelled as a convex surface. It has been analyzed through the use of first-order and second-order differential equations. The proposed method can remove the impairments without interrupting the transmission mode. The method in this thesis has been established by first post-estimating the compensation coefficients and then pre-compensating the input signal. The various adaptive techniques and numerous variations to the same have been presented in the literature as explained in chapter two of this thesis. Stochastic gradient is a special case of gradient descent monitoring the gradient of each sample in real-time unlike the gradient of the entire training dataset (Gradient Descent) which makes the computation expensive for large size data segment. However, there are challenges to executing an LMS in the presence of a noisy environment. The theory behind each is explained below:

3.1.1 Stochastic Gradient Descent

The objective behind a stochastic approximation of gradient descent is to closely monitor the ‘true’ gradient every time a new sample is encountered into the system. The weights are updated in accordance with the cost-function. As the objective function reaches a minimum value, gradient theoretically reaches a ‘zero’ value. Hence, in Stochastic Gradient, weights are updated continuously in the direction of the negative gradient. Due to a large variance in the input data, choice of step-size is slightly difficult and generally smaller than that of gradient descent. The sample for calculation of weight coefficients needs to be shuffled as it can bias the gradient. The learning rate necessarily requires a heuristic approach with a different learning rate for each weight.
to ensure the convergence [22]. Weight update, \( \hat{w}(n) \) for each sample \( x(n) \) to minimize the objective function, \( J(w, x(n)) \) looks like:

\[
\hat{w}(n) = \hat{w}(n - 1) - \mu_w \nabla J(w(n - 1), x(n)) \tag{3a}
\]

### 3.1.2 Gradient Descent

Unlike the ‘sample-based’ LMS, gradient descent computes the cost-function by ensemble averaging over a batch of data. This averaging over the entire dataset or a segment of dataset eliminates the impact of noise.

\[
\hat{w}(n) = \hat{w}(n - 1) - \mu_w \nabla \hat{J}(w(n - 1)) \tag{3b}
\]

Where the gradient, \( \nabla \hat{J}(w(n - 1)) \) is now calculated by averaging the cost-function over the entire dataset.

It is intuitively expensive for processing in the case of large dataset especially when the data is redundant and also way slower than the stochastic gradient. However, the use of an approximation of true gradient guarantees a convergence to the global minimum for smoothly convex surfaces. An alternative to a gradient descent can be to iterate over a certain batch of the dataset, defined as a ‘batch gradient descent’.

### 3.2 Pre-compensation Block and Imbalance Block

The standard system block follows from [5] with the imbalances being incorporated in the quadrature branch as shown in Fig.3.1, and considering the cross-talk of quadrature branch in the in-phase branch which is a standard model block for most of the modulators. The intention is to minimize the I/Q imbalances which cause the error differences in the envelope between the output and input of the modulator. The modulator imbalances have been conjoined as a single parameter to enable a joint optimization. For gain imbalance \( g \) and phase imbalance \( \phi \), we define the imbalance parameters as: \( \beta = g \sin \phi \) and \( \alpha = g \cos \phi \).
The compensation coefficients are equal to the imbalance values such that \( \hat{\beta} = \beta \) and \( \hat{\alpha} = \alpha \). The coefficients in equation (3b) can be represented as vectors such that:

\[
\hat{\Omega}(n) = \begin{bmatrix} \hat{\alpha}(n) \\ \hat{\beta}(n) \end{bmatrix}, \quad \hat{\Omega}(n-1) = \begin{bmatrix} \hat{\alpha}(n-1) \\ \hat{\beta}(n-1) \end{bmatrix}
\]

and

\[
\mu_w \nabla J_\hat{w}(w) = \begin{bmatrix} \mu \alpha \\ \mu \beta \end{bmatrix} \cdot \begin{bmatrix} \delta J \\ \delta \alpha \\ \delta \beta \end{bmatrix}.
\]

The estimate coefficients \( \hat{\alpha} \) and \( \hat{\beta} \), which will be adapted to minimize \( J(\alpha, \beta) \) until \( \nabla J(\hat{\alpha}, \hat{\beta}) \) goes to zero. From the model we see that, the original signal is retrieved when imbalance and compensation coefficients are equal. This makes the pre-compensation block an inverse of the imbalance block. The same target is kept in mind for joint estimation of gain and phase. The actual envelope output of the modulator is equalized with the envelope of the original input by jointly adapting the gain and phase estimates until the envelope difference is minimized.

**Figure 3.1: Pre-compensation and Imbalance Block**
3.3 Cost-function and mathematical analysis:

The baseband complex input, \( x_{in}(n) = x_i(n) + jx_q(n) \) gives the complete compensated complex imbalance output incorporating the compensation coefficients, \( \hat{\alpha} \) and \( \hat{\beta} \) and imbalance coefficients, \( \alpha \) and \( \beta \) as:

\[
y_{out}(n) = \left\{ x_i(n) + \frac{\hat{\beta} - \beta}{\hat{\alpha}} x_q(n) \right\} + j \left\{ \frac{\alpha}{\hat{\alpha}} x_q(n) \right\}
\]  

(3.1)

For mathematical analysis, we neglect the effect of noise because time-averaging the samples increases the signal-to-noise ratio as, SNR increases by the square root of the number of samples taken [23].

The envelope or instantaneous powers can be obtained by simply squaring the complex input and output instantaneous envelope given by \( P_{in}(n) = \|x_{in}(n)\|^2 \) and \( P_{out}(n) = \|y_{out}(n)\|^2 \) respectively.

(a) Error-Function:

The instantaneous error is then defined by simply substituting the terms \( x(n) \) and \( y(n) \) in \( P_{in}(n) \) and \( P_{out}(n) \) respectively.

\[
e(n) = |P_{in}(n) - P_{out}(n)|
\]  

Substituting the values from (3.1) we get,

\[
e(n) = \left| 2 \left( \frac{\hat{\beta} - \beta}{\hat{\alpha}} \right) x_q(n)x_i(n) + \left( \frac{\hat{\beta} - \beta}{\hat{\alpha}} \right)^2 + \frac{\alpha^2}{\hat{\alpha}^2} - 1 \right| x_q^2(n)
\]  

(3.2)

Cost-function, \( J(n) \) at each instant ‘\( n \)’ is obtained by ensemble averaging the mean square of the instantaneous errors over any batch of dataset.

\[
J(n) = MSE = \langle e^2(n) \rangle = \langle \left[ 2 \left( \frac{\hat{\beta} - \beta}{\hat{\alpha}} \right) x_q(n)x_i(n) + \left( \frac{\hat{\beta} - \beta}{\hat{\alpha}} \right)^2 + \frac{\alpha^2}{\hat{\alpha}^2} - 1 \right] x_q^2(n) \rangle^2
\]  

(3.3)
Equivalently denoting expectation term by ‘E’, \( J(n) \) can be defined as,

\[
J(n) = E[e^2(n)]
\]

(3.4)

\[
J(n) = E \left[ 2 \left( \frac{\hat{\beta} - \beta}{\hat{\alpha}} \right) x_1(n) x_2(n) + \left( \frac{\hat{\beta} - \beta}{\hat{\alpha}} \right)^2 + \frac{\alpha^2}{\hat{\alpha}^2} - 1 \right] x_q^2(n)
\]

(3.5)

\[
= E \left[ \left( \frac{\hat{\beta} - \beta}{\hat{\alpha}} \right)^4 + \left( \frac{\alpha}{\hat{\alpha}} \right)^4 + 1 + 2 \left( \frac{\hat{\beta} - \beta}{\hat{\alpha}} \right)^2 \left( \frac{\alpha}{\hat{\alpha}} \right)^2 - 2 \left( \frac{\alpha}{\hat{\alpha}} \right)^2 - 2 \left( \frac{\hat{\beta} - \beta}{\hat{\alpha}} \right)^2 \right] x_q^2(n) + 4 \left( \frac{\hat{\beta} - \beta}{\hat{\alpha}} \right)^2 x_1^2(n) x_q^2(n) + 4 \left( \frac{\hat{\beta} - \beta}{\hat{\alpha}} \right) x_1(n) x_q^3(n) \left( \frac{\hat{\beta} - \beta}{\hat{\alpha}} \right)^2 + \left( \frac{\alpha}{\hat{\alpha}} \right)^2 - 1
\]

(3.6)

It is seen that \( J(n) \) becomes equal to zero for \( \hat{\alpha} = \alpha \) and \( \hat{\beta} = \beta \).

(b) First-Order Differentials:

The cost-function, \( J(n) \) will vary simultaneously with respect to the estimate parameters \( \hat{\beta} \) and \( \hat{\alpha} \) and will reach a stationary point when \( \hat{\alpha} = \alpha \) and \( \hat{\beta} = \beta \).

Differentiation is a linear operation i.e differentiation of expectation is the expectation of differentiation. Taking expectation of the in-phase and quadrature phase signals and differentiating \( J(n) \) with respect to \( \hat{\beta} \) (keeping \( \hat{\alpha} \) constant) we obtain:

\[
\frac{\partial J(n)}{\partial \hat{\beta}} = \frac{4(\hat{\beta} - \beta)^2}{\hat{\alpha}^2} \left[ \left( \frac{\hat{\beta} - \beta}{\hat{\alpha}} \right)^2 + \frac{\alpha^2}{\hat{\alpha}^2} - 1 \right] x_q^4(n) + 2x_q^2(n)x_1^2(n) + \frac{4}{\hat{\alpha}^2} \left[ 3 \left( \frac{\hat{\beta} - \beta}{\hat{\alpha}} \right)^2 \frac{x_q^4(n)}{x_1^2(n)} + \left( \frac{\alpha}{\hat{\alpha}} \right)^2 - 1 \right] x_1(n) x_q^3(n)
\]

(3.7)

Since, \( x_1(n) \) and \( x_q(n) \) are a pair of random variables, we can use the property of the product of Gaussian random variables which states that for say, 4 random gaussian variables, \( X_1, X_2, X_3, X_4 \), joint expectation is defined as [24]:

\[
E[X_1X_2X_3X_4] = E[X_1X_2]E[X_3X_4] + E[X_1X_3]E[X_2X_4] + E[X_1X_4]E[X_2X_3]
\]

(3.8)

Also, for two random and independent variables, \( E[X_1X_2] = E[X_1]E[X_2] \).

(3.9)

Since, \( x_1(n) \) and \( x_q(n) \) are zero mean, \( x_1(n) x_q(n) = 0 \).

(3.10)
Using eqns (3.8), (3.9) and (3.10) for \( x_i(n)x_q^3(n) \) and \( x_q(n)x_i^3(n) \) yields \[ x_i(n)x_q^3(n) = x_q(n)x_i^3(n) = 0 \] giving:

\[
\frac{\partial f(n)}{\partial \beta} = \frac{4(\bar{\beta} - \beta)}{\bar{\alpha}^2} \left\{ \left( \frac{\bar{\beta} - \beta}{\bar{\alpha}^2} \right)^2 + \frac{\alpha^2}{\bar{\alpha}^2} - 1 \right\} x_q^4(n) + 2x_q^2(n)x_i^2(n) \tag{3.12}
\]

Similarly, with respect to differentiating \( \hat{\alpha} \) and neglecting the terms \( x_i(n)x_q^3(n) \) and \( x_i(n)x_q(n) \),

\[
\frac{\partial f(n)}{\partial \hat{\alpha}} = \frac{2}{\bar{\alpha}^2} \left\{ \left( \frac{\hat{\alpha} - \beta}{\bar{\alpha}^2} \right)^4 - \frac{\alpha^2}{\bar{\alpha}^2} - 4 \left( \frac{\hat{\alpha} - \beta}{\bar{\alpha}^2} \right)^2 + \alpha + \frac{2}{\bar{\alpha}} (\hat{\beta} - \beta)^2 \right\} x_q^4(n) - \frac{4}{\bar{\alpha}} (\hat{\beta} - \beta)^2 x_q^2(n)x_i^2(n) \tag{3.13}
\]

It can be seen from equations (3.12) and (3.13) that,

\[
\frac{\partial f(n)}{\partial \hat{\alpha}} \to 0 \text{ and } \frac{\partial f(n)}{\partial \beta} \to 0 \text{ for } \hat{\alpha} \to \alpha \text{ and } \hat{\beta} \to \beta.
\]

Hence, the gradient goes to zero at optimum points.

(c) Second-Order Differentials:

Differentiating equation (3.12) with respect to \( \hat{\beta} \) again and using (3.11) we get,

\[
\frac{\partial^2 f(n)}{\partial \hat{\beta}^2} = 12 \left( \frac{\hat{\beta} - \beta}{\bar{\alpha}^4} \right)^2 + 4 \frac{\alpha^2}{\bar{\alpha}^4} - \frac{4}{\bar{\alpha}^2} \left\{ x_q^4(n) + \frac{8}{\bar{\alpha}^2} x_q^2(n)x_i^2(n) \right\} \tag{3.14}
\]

\[
\frac{\partial^2 f(n)}{\partial \hat{\alpha}^2} = \left\{ -4 \left( \hat{\beta} - \beta \right)^4 - 4\alpha^4 - 8 \left( \hat{\beta} - \beta \right)^2 \frac{\alpha^2}{\bar{\alpha}^4} \right\} \left( \frac{5}{\bar{\alpha}^5} \right) + \left\{ 4\alpha^2 + 4 \left( \hat{\beta} - \beta \right)^2 \right\} \left( \frac{-3}{\bar{\alpha}^3} \right) x_q^4(n) +
\]

\[
8 \left( \frac{\hat{\beta} - \beta}{\bar{\alpha}^2} \right) x_q^2(n)x_i^2(n) \tag{3.15}
\]

When \( \hat{\beta} \to \beta \) and \( \hat{\alpha} \to \alpha \),

\[
\frac{\partial^2 f(n)}{\partial \hat{\beta}^2} = \frac{8}{\bar{\alpha}^2} x_q^2(n)x_i^2(n) \text{ and } \frac{\partial^2 f(n)}{\partial \hat{\alpha}^2} = 8x_q^4(n)
\]
i.e. twice-differentiable function $J(\alpha, \beta)$ is positive definite at the optimum points showing that the surface is convex!

(d) **Adaptive Gradient Descent Equation:**

A parallel optimization is carried out for both the estimates. The gradient is evaluated using the standard secant method on the normalized cost-function.

\[
\dot{\beta}(n) = \dot{\beta}(n-1) - \mu_\beta \left( \frac{\partial J(\hat{\alpha}, \hat{\beta})}{\partial \hat{\beta}} \right) \tag{3.16}
\]

\[
\dot{\alpha}(n) = \dot{\alpha}(n-1) - \mu_\alpha \left( \frac{\partial J(\hat{\alpha}, \hat{\beta})}{\partial \hat{\alpha}} \right) \tag{3.17}
\]

The above update requires the initialization of two values for each $\hat{\alpha}$ and $\hat{\beta}$. Learning-rates $\mu_\alpha$ and $\mu_\beta$ have been initialized with reasonable values.

The initialized values have been plotted against various values to determine an upper bound to the values for both the parameters.

It can be seen in figures 3.2 and 3.3 that a single step-size requires the knowledge of an upper bound to prevent divergence.
Figure 3.2: Convergence plots of various values of $\mu_\beta$ for a given value of $\mu_\alpha$

Figure 3.3: Convergence plots of various values of $\mu_\alpha$ for a given value of $\mu_\beta$
A large step-size and the presence of noise do not lead to a smooth convergence. Hence, we need to be careful with the choice of step-size. The performance becomes sensitive to the step-size when the size of the batch goes small due to a higher sensitivity of noise on the cost-function. The choice of a batch-gradient in this thesis is influenced by the presence of noise especially when the processing is carried out on a very small section of data. The same has been analyzed for various batch-sizes which lead this thesis to conclude a good performance of the proposed approach for a batch size as low as 1000 samples.

It is observed that the algorithm converges in the initial few iterations, however, the sensitivity of the algorithm to various factors is seen when the rate of change of the parameter estimates is slow in comparison to the rate of change of cost-function. In other words, if the function changes slowly with respect to the change in estimates (or it is not deeply convex), it causes the algorithm to fluctuate and converge again. This is controlled by monitoring the condition on the minimum value of the cost-function consecutively for a certain number of iterations. This cost-function directly relates with the best ISR to be attained. The algorithm is then controlled beyond this point to further stop convergence. If this condition is not imposed, the random fluctuations would be more prominent for a smaller sized batch. Also, we observe that doing a simultaneous adaptation of both parameters is a big advantage of the algorithm as one parameter needn’t wait for the other parameter to converge. The only disadvantage of going with the simultaneous adaptation can occur when a momentary fluctuation on the convergence of one parameter jerks off the algorithm convergence momentarily. However, this divergence is not observed for a batch as low as 1000 samples.

We observe from the Figs 3.2 and 3.3 that we attain the best stability in the algorithm for an intermediate value of \( \mu_\alpha = 0.125 \) and \( \mu_\beta = 0.125 \) once the upper bounds are determined.
3.4 Conclusion

Two basic algorithms have been introduced in first part of the chapter. The pre-compensation block and imbalance block have been modelled to formulate the cost-function problem. The convergence for the two coefficients has been analyzed for the formulated cost-function using first order and second-order differential equations to determine an optimum solution. Optimum step-sizes have been determined by varying different values of step-sizes for each coefficient, keeping the other fixed to determine an upper-bound for the step-size. The stability of the algorithm is discussed in the presence of noise for a different batch-size of data.
CHAPTER FOUR: SIMULATION AND IMPLEMENTATION RESULTS

4.1 Simulation Results

Simulations have been carried out using a cascade of a pre-compensation block and an imbalance block as shown in Fig.3.1. A different batch of incoming data is processed during each iteration in adaptation to closely represent any changes in the envelope if they occur due to variations in the input data. Real envelopes of the output and input signal are generated in MATLAB. Simulation and measurement results validation have been carried out using an LTE 5 MHz wideband signal sampled at 92.16 MHz with a 64 QAM scheme. For the purpose of observing the images, we shift the signal at 5 MHz IF. A high sampling frequency reduces the requirements on the specifications of reconstruction LPFs to eliminate Nyquist images from the output of DAC.

We have confined to a single optimum step-size for analyzing the performance for all number of samples.

We consider the case of a reasonable imbalance of \( g = 0.5 \text{dB (1.0592)} \), \( \varphi = -5^\circ \). Additive White Gaussian noise (AWGN) of 60 dB SNR is included in the output feedback signal.

The joint imbalance parameters are then defined by \( \alpha = g \cos \varphi = 1.055 \), \( \beta = g \sin \varphi = -0.0922 \).

The gradient is calculated using a standard secant method which requires the initialization of two values for each variable.

For any parameter, the general update equation for a weight coefficient in the direction of negative gradient looks like:

\[
\hat{w}(n) = \hat{w}(n - 1) - \mu_w \left( \frac{J(n-2) - J(n-3)}{\hat{w}(n-2) - \hat{w}(n-3)} \right) \quad (4.1)
\]

Standard values of \( \hat{\alpha} \) and \( \hat{\beta} \) are chosen close to 1 and 0 respectively:
\[ \hat{\alpha}_1 = 1.0, \hat{\alpha}_2 = 1.0005 \]
\[ \hat{\beta}_1 = 0, \hat{\beta}_2 = 0.05 \]

4.1.1 Performance analysis for different number of samples:

We first summarize the observations for the above-chosen parameters for a case of 100 K samples and then compare the performance for a different batch-size.

![Estimation Error in α](image)

**Figure 4.1: Estimation Error of \( \hat{\alpha} \)**
Figure 4.2: Estimation Error of $\hat{\beta}$

Figure 4.3: Imbalance parameter estimates
Figure 4.4: Convergence of Cost-Function $J(n)$

Figure 4.5: ISR and NMSE
To prove the stability of the algorithm, we do a run of 50 simulations and average it over 50 runs. It can be observed from Fig 4.6 that performance of the algorithms in terms of ISR degrades for a lower-sized batch from $-61$ dB to $-55$ dB with larger variations for a case of 1K samples. However, the proposed technique still performs reasonably well. This insignificant degradation has been highlighted to emphasize the advantage of lower order moments.

**Figure 4.6: CDF plots of ISR for different number of samples**

It is observed from Fig 4.6 that in the case of 10K samples, for 99.99% of the runs the ISR obtained is better than $-58.6$ dB. In the case of 1K samples, for 99.9% of the runs, the ISR is better than $-54$ dB.
Table 4.1: Summary of performance for different number of samples

<table>
<thead>
<tr>
<th>No of samples</th>
<th>Performance metric</th>
<th>Average ISR (dB)</th>
<th>NMSE (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100K</td>
<td></td>
<td>−61.034</td>
<td>−57.45</td>
</tr>
<tr>
<td>10K</td>
<td></td>
<td>−59.28</td>
<td>−55.3</td>
</tr>
<tr>
<td>1K</td>
<td></td>
<td>−55.4</td>
<td>−49.6</td>
</tr>
</tbody>
</table>

It can be observed from Table 4.1 that the average ISR degrades from −61.034 dB for 100K samples to −55.4 dB for 1K samples. Similarly, NMSE degrades from −57.45 dB for 100K samples to −49.6 dB for 1K samples. The table summarizes the degradation in performance of the algorithm in terms of ISR and NMSE as, lower batch-size tends to give a less accurate estimate of moments, eventually reflecting on the performance of the algorithm.

4.1.2 Comparison with a state-of-the-art technique:

Similar conditions of the algorithm have been utilized for [21] for a case of 100K samples with an LTE signal of 64-QAM for the performance analysis. The gradient descent in [21] utilizes a fourth-order moment defined by:

\[ J(g, \varphi) = \frac{M_4(g, \varphi)}{M_2^2(g, \varphi)}; M_4(g, \varphi) = E[|y_{out}(n)|^4] \text{ and } M_2(g, \varphi) = E[|y_{out}(n)|^2] \]

where \( y_{out}(n) \) is the output signal from the quadrature modulator.

It can be seen from Fig. 4.7 that, in the proposed approach for the case of 100 K samples, 99.99 % of the runs exhibit an ISR of better than -61 dB. A fourth order moment, on the other hand, exhibits a varied range of ISR values with 99.99 % of the runs demonstrating an ISR of around ~ -51 dB. This clearly demonstrates an improved performance of a second-order moment over the fourth-order moment. It is evident from Figures 4.6 and 4.7 that, the processing of 1K samples using a second-order moment converging in 50 iterations shows a performance on par with the
performance of a fourth-order moment approach using 100K samples for processing, converging in about 20 iterations. The fourth-order moment approach from Fig 2.2.b has proven to show an unstable performance for samples < 100K making the convergence a difficult task. This indeed makes the proposed technique computationally a less complex approach. Table 4.2 summarizes the performance of the two techniques in the same simulation environment.

Table 4.2: Summary of performance comparison between two techniques

<table>
<thead>
<tr>
<th>For 100K Samples</th>
<th>Mean ISR (dB)</th>
<th>No. of iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>[21]</td>
<td>−54.5</td>
<td>26</td>
</tr>
<tr>
<td>Proposed Technique</td>
<td>−61</td>
<td>20</td>
</tr>
</tbody>
</table>

Figure 4.7: CDF Plot of ISR compared with [21]

4.2 Measurement Results:

The test-bench to carry out an on-air adaptive process is not within the scope of this research project. However, to validate the chosen approach and the algorithm, the post-estimation and then
pre-compensation is carried out on a make-shift experimental set-up in an ESG 4438C at a carrier frequency of 2.14 GHz. The block diagram of the set-up is shown in Fig 4.9. The pre-compensation block and the imbalance block is essentially a linear system which makes it possible in our case to reverse the blocks (Fig. 4.8) and still be able to carry out the estimation and compensation. The baseband signal is generated in the signal-processor in MATLAB. The signal is then downloaded in the ESG 4438C Vector Signal Generator and the modulated spectrum is then captured on Spectrum Analyzer E4440A using Vector Signal Analyzer (VSA) software. The sampling frequency is chosen to be 92.16 MSPS. However, the problem of phase-shift needs to now be considered in case of post-estimation as the input signal, \( x'(n) \) entering the compensation block would be phase shifted and it is this delayed signal which eventually leads to an errored comparison of the output envelope against the envelope of the input signal. The phase has thus been adjusted to align with the input signal before carrying out the post-estimation and then pre-compensation such that the feedback circuit is defined as \([21], \bar{y}(n) = y(n)e^{j\theta} \) where \( \theta \) is the phase-shift between the input and output signal due to delay in the measurement circuit.

Figure 4.8: Post-estimation Block
The entire procedure for carrying out the measurement and processing is shown in Fig. 4.9.
The measurement results have been obtained for three values of imbalances and for a different number of samples for a given value of imbalance.

Case 1: 2 different imbalances:
Figure 4.10: (a) Imbalance Output (b) Compensated Output for 2 values of imbalances

Table 4.3: Performance for different imbalances for 100K samples

<table>
<thead>
<tr>
<th>ISR (dB)</th>
<th>$g = 0.2 , dB, \varphi = -2^\circ$</th>
<th>$g = 0.8 , dB, \varphi = -5^\circ$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-54.7</td>
<td></td>
<td>-56</td>
</tr>
</tbody>
</table>

Table 4.3 summarizes the performance for both cases and it is evident that the performance of the algorithm is independent of the value of imbalances a modulator presents to the transmitted signal.
Case 2: Different Number of samples

Figure 4.11: ISR Spectrum Plots for different number of samples; $g = 0.5 \text{ dB (1.0592)}, \varphi = -5^\circ$.

Table 4.4 shows a summary of the results obtained for all number of samples. The results are essentially at their best for a higher sized batch, however, the discrepancy in the results between the simulations and measurement exist due to the constraint of the test-bench problem which induces the sensitivity of results due to inherent phase-shift in the instrument.
Table 4.4: Summary of performance for different number of samples (in measurement); $g = 0.5 \, dB \, (1.0592) \, , \, \phi = -5^\circ$

<table>
<thead>
<tr>
<th>Number of samples</th>
<th>ISR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 K</td>
<td>−58</td>
</tr>
<tr>
<td>10K</td>
<td>−44.1</td>
</tr>
<tr>
<td>1K</td>
<td>−40.5</td>
</tr>
</tbody>
</table>

4.3 Conclusion:
The first section of this chapter analyzes the performance in simulations for a different batch-size with 50 runs to determine the stability of the approach by comparing with each size of data. The case of 50 runs is compared in section 2 of this chapter for a case of 100K samples for the proposed and fourth-order moment approach to compare the achieved steady state ISR and NMSE and the number of iterations required for each case. The third part validates the approach in measurements by comparing the performance for different values of imbalances to show that the approach is independent of the quadrature modulator imbalances. Validation is also carried out for different number of samples.
CHAPTER FIVE: CONCLUSION AND FUTURE SCOPE

5.1 Conclusion and Summary

In this thesis, a simplified adaptive approach for solving the frequency independent I/Q impairments has been established and demonstrated based on a standard asymmetric imbalance model. A comprehensive mathematical analysis validates the chosen approach behind going for a simple and computation-friendly algorithm of a batch Gradient Descent. The convergence of the convex problem is justified and analyzed through the use of first-order and second-order differential equations. This algorithm has proven to provide efficient results in simulations for a varied number of samples using a standard LTE signal 5 MHz band-wide with practical values of imbalances and the inclusion of noise in the feedback signal. The algorithm converges in about 20 iterations. The degradation in performance is insignificant until 1000 samples. The algorithm jointly compensates for both gain and phase using a simultaneous adaptation. ISR was improved from 24 dB to 61.034 dB for a full-length data and to -55.4 dB for 1000 samples. NMSE has been obtained to be -57.4 dB for 100K samples and around -49.6 dB for 1000 samples.

The simulation results of the proposed compensation algorithm are in accordance with the results obtained after validation on hardware for 100K samples and meets the spectral mask requirements. Although the proposed approach will work efficiently for all kinds of feedback circuits [2], the approach will deliver an efficient performance for lower number of samples in case of post-estimation only when the phase-shift between the feedback signal and the input signal will be accurately known or else a significant residual image will show up as a result of the existing problem of phase-shift in the instrument which with a real-time setup will be completely eliminated. The results are also validated on hardware for two values of imbalances.
5.2 Contributions:

1. A simple yet computationally efficient algorithm jointly estimates and compensates for the I/Q impairments. A simultaneous approach for the two coefficients makes it a robust adaptive recipe without having to depend on the convergence of one parameter for the other even upon the occurrence of momentary fluctuations as it doesn’t affect the convergence behavior. The combination of this convenient established joint scheme along with an envelope based approach makes it computationally and economically viable.

2. A direct learning approach has been proposed to determine the pre-distortion coefficients directly by comparing the output signal envelope with the input signal envelope.

3. A comprehensive study of the existing techniques for elimination of I/Q imbalances and comparison with an established state-of-the-art method suggests the I/Q mitigation proposed approach to be promising.

5.3 Future Scope:

1. Eventually, instead of processing the signal to get the envelope in software, a diode detector will be used in the feedback circuit. The diode in such a case will be operated in a square-law region where the output voltage is proportional to the input power. Hence, modeling the diode is a potential problem to consider especially for high power signals (when the signal exceeds -20 dBm up to which the square law region of the diode exists) i.e. to model the deviation from the square-law due to which higher order terms would come into picture. The upper limit of the dynamic range is greatly limited by this deviation of the device from the square-law region and the lower end is constrained by noise which limits the precision of measurements [25].
2. DC offset has currently not been included as a part of the adaptive problem as alternatively, it can be removed by subtracting the mean of the signal from the original signal before feeding to the system. An adaptive approach, however, is yet another important area to cater to the time-varying changes in dc-offset.

3. An on-air adaptive pre-estimation should essentially eliminate the phase-shift problem which arises in the case of post-estimation in our case. The proposed approach can be tested on the relevant test-bench to eliminate this additional coefficient which leads to an errored estimation when we validate the proposed approach using post-estimation in the available make-shift arrangement in ESG 4438C.
REFERENCES


