Efficient Baseband Design and Implementation for High-Throughput Transmitters

by

Zhipeng Li

S.B., Physics, Massachusetts Institute of Technology (2009)
S.B., Electrical Science and Engineering, MIT (2009)
M.Eng., Electrical Engineering and Computer Science, MIT (2010)
Electrical Engineer, MIT (2011)

Submitted to the
Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2015

© Massachusetts Institute of Technology 2015. All rights reserved.
Efficient Baseband Design and Implementation for
High-Throughput Transmitters

by

Zhipeng Li

Submitted to the Department of Electrical Engineering and Computer Science
on August 28, 2015, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

Wireless communications are accelerating into the realm of higher data rates from hundreds of megabits to tens of gigabits per second. Increase in data rate requires higher throughput and higher utilization of spectral bandwidth. At the same time, we are seeing a demand for smaller chipsets with lower power budgets. Digital basebands with increased energy-efficiency are needed while fitting within tight area constraints. High spectral efficiency demands modulation schemes with high peak to average power ratio, increasing the precision requirements on the digital baseband circuitry.

To enable a new class of energy-efficient millimeter wave communication systems based on outphasing power amplifiers (PAs), we have explored ways to implement high-throughput outphasing baseband functions with the smallest energy and area footprints. Aware of the limitations of field-programmable gate arrays (FPGA) in throughput and energy-efficiency, we have chosen to implement our digital baseband in application-specific integrated circuits to allow a truly integrated energy-efficient transmitter. By utilizing the changes in micro-architecture (parallelism and pipelining) and aggressive back-end power optimization techniques (noncritical path Vt replacement and sizing reductions), we achieve a record energy-efficiency and throughput for asymmetric-multilevel-outphasing (AMO) signal component separator (SCS) of 32pJ/sample at 0.6V supply voltage and 400Msamples/s, with an area of 0.41mm². For high-throughput area-constrained applications, our static random-access memories based AMO SCS design achieves 2× area reduction over the register-based design at the same throughput to allow more parallelism to meet the stringent throughput requirements.

To compensate for system nonlinearity and memory effects, we implement a zero-avoidance shaping filter in place of the traditional shaping filter to improve convergence in model iterations of an outphasing transmitter, and design an energy- and area-efficient digital predistorter (DPD). We use this DPD architecture to compensate for nonideal phase modulation, preamplifier saturation, and many transmitter nonidealities. Applying this developed methodology in spice-level simulation, we improve adjacent-channel-power-ratio (ACPR) of the outphasing Q-band
(45GHz) transmitter with 1.1Gsamples/s throughput from -30.6dB to -44.0dB and reduced error vector magnitude (EVM) from 4.5% down to 1.0% with 64-Quadrature-Amplitude-Modulation (64QAM) and real-time zero avoidance. The energy efficiency of this predistorter at a throughput of 1.1Gsamples/s (3.3Gbps data rate with 64QAM modulation and oversampling ratio of 2) is 1.5nJ/sample.

To illustrate the wide applicability of this proposed linearization methodology, we applied it to compensate for distortion in a radio-frequency PA. We apply the off-line iterative compensation method to a PA with 1.97GHz carrier frequency and 737Mbps data throughput with 64QAM. We map the designed DPD structure onto FPGAs with a utilization of 144 DSP slices and an energy efficiency of 1.7nJ/sample. To meet an ACPR constraint of -48dB, the uncompensated PA has to back-off the input power by 12dB with 3.3% power efficiency. The compensated PA has to back-off by only 6dB with 9.2% overall transmitter power efficiency which includes the DPD power, almost 3× the efficiency of uncompensated PA.

Thesis Supervisor: Vladimir M. Stojanović
Title: Associate Professor of Electrical Engineering and Computer Science
Acknowledgments

I would like to thank my advisor Professor Vladimir Stojanović for his guidance and patience throughout these years. I would like to thank my thesis committee, Professor Alexander Megretski and Professor Duane Boning, for their generous helps in writing this thesis.

I would like to thank the members of the Efficient Linearized All-Silicon Transmitters ICs (ELASTx) project team at MIT, Carnegie Mellon University, and North Carolina State University: Yan Li, Zhen Li, Zohaib Mahmood, Wei Tai, Yingying Tang, Omer Tanovic, Mark Tobenkin, Junfeng Xu, Gilad Yahalom, Yi-Shin Yeh, DongNi Zhang, in the alphabetical order of their last names; we learned from each other and built a seamless team. I would like to thank the members of the MIT Integrated System Group: Wei An, Fred Chen, Sun Chen, Hossein Fariborzi, Michael Georgas, Byungsub Kim, Jonathan Leu, Fabian Lim, Benjamin Moss, Sanquan Song, Ranko Sredojevic, Oguzhan Uyar; without this wonderful team, I would not have known where to get started. I would like to thank many other colleagues at the Microsystems Technology Laboratories and the Research Laboratory of Electronics: Nigel Drego, Ali Farahanchi, Hayden Taylor, Daniel Truque, Xiaolin Xie, of Professor Duane Boning’s group; Reja Amatya, Harry L. T. Lee, Kevin Shao-Kwan Lee, Peter Mayer, Jason Orcutt, Tauhid Zaman, of Professor Ram Rajeev’s group; Taylor Barton, Jose Bohorquez, Sungwon Chung, Philip Godoy, Tania Khanna, Muyiwa Ogungbemi, William Sanchez, of Professor Joel Dawson’s group.

I would also like to thank friends and family for their support. My wife has read my thesis over and over again meticulously and helped me with grammar, style and flow, not only turning my text into beautiful poetry but also touching on my figures and tables into beautiful canvases. Besides myself, she has read this piece probably more times than anyone else. Without her diligence, this piece would never be as it stands today.

I would like to thank my parents. They are my inspiration. Their struggle with adversities in the nineties (I was too young to remember anything in the eighties,
and after nineties I left Wuhan and was there rarely) always remind me that no matter how powerless we small figures are, we run and we keep going.

I would also like to thank my grandparents, though regrettably none of them have lived to see me graduate. But with the availability of universal language translators in the after-life, I hope they enjoy my thesis wherever they may be at rest.
## Contents

1 Introduction .................................................. 19  
   1.1 Wireless Transmitters with Higher Data Rate but Lower Power Budget Requirements .............................................. 19  
   1.2 Versatile Digital Baseband Needs to be Energy-Efficient ................................................................. 20  
   1.3 Thesis Contributions ........................................... 21  
      1.3.1 Signal Component Separator for Outphasing Transmitters .......................................................... 21  
      1.3.2 Efficient Digital Compensator for Power Amplifier ................................................................. 22  
   1.4 Thesis Overview .................................................. 23  

2 Energy and Area Efficiency of High-Throughput Digital Baseband .................................................. 25  
   2.1 Overview of Outphasing Transmitters .............................. 25  
      2.1.1 Digital Baseband for Outphasing Transmitters ................................................................. 26  
      2.1.2 Nonideal Phase Modulation ................................................................. 30  
   2.2 Motivation for ASIC Implementation .................................. 37  
   2.3 Techniques to Achieve Higher Energy Efficiency ............... 40  
      2.3.1 Optimized Piece-Wise Linear Approximation ................................................................. 41  
      2.3.2 Backend Optimization ................................................................. 43  
      2.3.3 Micro-Architecture ................................................................. 46  
      2.3.4 Area Reduction for Parallelism ................................................................. 52  
   2.4 Architecture Overhead: AMO versus LINC ....................... 57  
   2.5 Impact on Overall System Efficiency by the Energy Footprint of Digital Baseband ........................................ 58
3 Efficient Digital Compensation of Transmitter Nonlinearity

3.1 Overview of Digital Compensation

3.1.1 Linearity Metrics

3.1.2 Popular Digital Compensation Techniques

3.1.3 Nonidealities in Outphasing Transmitters

3.2 Digital Compensator for Outphasing Transmitters

3.2.1 Design and Implementation of a Zero-Avoidance Shaping Filter for Outphasing Transmitters

3.2.2 Implementation of Digital Compensator for Outphasing Transmitters

3.2.3 Limitations

3.3 Digital Compensation of RF Power Amplifiers

3.3.1 Experimental Setup

3.3.2 Phase Modulator Calibration

3.3.3 Off-line Iterative Sequence-based Compensation

3.3.4 Real-time Compensator Model

3.3.5 Implementation of Digital Compensator for RF PA

3.4 Constellation Design

3.4.1 Alternative 64-Symbol Constellations to 64QAM

3.4.2 More Bits Per Symbol

3.4.3 Limitations

4 Conclusion and Future Research Directions

4.1 Conclusion

4.2 Future Research
List of Figures

2-1 Vector decompositions in different outphasing architectures assuming four restricted power supply levels indicated by blue solid or black dashed arc lines. ................................................. 26
2-2 Overview of AMO PA system. .................................................. 27
2-3 The block diagrams of AMO SCS and LINC SCS. ......................... 29
2-4 Power efficiency degradation by a 800Msamples/s-throughput AMO SCS of 106pJ/sample energy efficiency [29]. ............................. 31
2-5 Representation of phase modulation. ........................................... 32
2-6 Weight factors of triangular approximation in Equation (2.6). ...... 34
2-7 An illustration of the factor weights $c_i$ and $c_q$ for phase modulation and the corresponding normalization constraints. ...................... 35
2-8 Weight factors of triangular approximation in Equation (2.7). ...... 36
2-9 Mapping an FIR filter onto DSP48E1 slices on an FPGA. The two clock cycles of input delay generated inside the DSP48E1 slice and the dedicated cascade PCIN/PCOUT and ACIN/ACOUT connections internal to the column are utilized to achieve maximum performance irrespective of the number of coefficients [35]. .............................. 39
2-10 Micro-architecture of the OPWL approximation in hardware. ...... 42
2-11 The ultra-high threshold voltage standard cells consume only 1/10 of the leakage power of the regular threshold voltage stand cells, at the expense of reduced maximum performance. .............................. 44
2-12 The ASIC design flow with leakage reduction stages highlighted in red. 45
2-13 A digital system with various amount of pipelining. .................... 49
2-14 The hardware block diagram of the depipelined AMO SCS design. . . 50
2-15 Micro-architecture trade-offs for OPWL AMO SCS design. . . . . . 51
2-16 The throughput and energy with supply scaling for the AMO SCS chip in our first OPWL paper [22], multi-Vt AMO SCS (assumed 2-way parallel), and SRAM-based AMO SCS (assumed 4-way parallel). . . . 55
2-17 Deserialization of 2-way parallel system is absorbed into the 2× up-sampling and shaping filter and has minimum power impact, while serialization remains and consumes power. . . . . . . . . . . . . . . . 56
2-18 Power breakdown of the AMO SCS design and the LINC SCS design. 59
2-19 The measured throughput and energy with supply scaling for multi-Vt depipelined AMO SCS, LINC SCS chip. . . . . . . . . . . . . . . . 59
2-20 Power efficiency degradation by a 800Msamples/s-throughput AMO SCS of the improved 32pJ/sample energy efficiency. . . . . . 60
3-1 EVM is the rms magnitude of error vectors, with the peak symbol amplitude normalized to be 1. . . . . . . . . . . . . . . . . . . . . . . 65
3-2 An illustration of the ACPR definition. . . . . . . . . . . . . . . . . 65
3-3 Tradeoff between efficiency and ACPR of a single PA. . . . . . . . 66
3-4 When the inverse and the system are concatenated together, the output appears as a linear amplification of the input. . . . . . . . . . 67
3-5 Common nonlinear dynamical system structures. . . . . . . . . . . 68
3-6 Illustration of the transmitter system under compensation, with blocks highlighted in the red box simulated in Spectre and other blocks in MATLAB. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 71
3-7 Simulation setup to quantify the ACPR degradation. . . . . . . . . 72
3-8 The simulated ACPR numbers along the signal chain with different permutations of SPICE-level and ideal models. Note that the ACPR numbers with Wilkinson combiner and ideal combiner are similar. . . 73
3-9 EVM of the Q-band (45GHz) LINC transmitter under study, uncompensated versus off-line compensated with zero-avoidance input sequence. Ideal symbols are in red and the output symbols are in blue.  77

3-10 Uncompensated output versus compensated output ACPR of the LINC system, with real-time zero-avoidance input sequence.  77

3-11 A zero-avoidance shaping filter must be incorporated for the predistorter to be effective.  78

3-12 Illustration of the zero-avoidance zone in the IQ plot [24].  79

3-13 Illustration of the zero-avoidance shaping filter algorithm [24].  80

3-14 The functional diagram of a zero-avoidance shaping filter.  80

3-15 The ACPR using the zero-avoidance shaping filter, the maximum amplitude of shaped samples, and the average amplitude as percentages of $R_z$ of samples with amplitude below $R_z$.  82

3-16 The hardware implementation of the zero-avoidance shaping filter.  83

3-17 Optimized coefficients of $h_s[n]$ and $h_c[n]$ at $m_s = 59$ and $m_c = 61$.  83

3-18 The filtered samples without and with zero-avoidance.  84

3-19 An $m_s$-tap shaping filter requires only $(m_s+1)/2$ adders and $(m_s+1)/4$ multipliers, because of its no-ISI and symmetric properties and because the center tap is normalized $h_s[0] = 1$, as illustrated above with $m_s = 7$.  85

3-20 An $m_c$-tap correcting filter requires only $(m_c-1)/2$ adders and $(m_c-1)/4$ multipliers, because of its no-ISI and symmetric properties and that the center tap is normalized $h_c[0] = 1$, as illustrated above with $m_c = 9$.  86

3-21 The block diagram of the predistorter hardware implementation, configured in the LINC mode.  87

3-22 The block diagram of the digital baseband with zero-avoidance shaping filter, SCS and nonlinear predistorter.  91
3-23 Die photograph of transmitter chip \textit{nlcom6} with digital baseband including dynamic nonlinear predistorter with zero-avoidance shaping filter and integrated Q-band (45GHz) AMO PA and W-band (93GHz) LINC PA. ................................................. 92

3-24 Gate area breakdown of digital baseband of transmitter \textit{nlcom6}. .... 93

3-25 The power breakdown of the digital baseband of transmitter \textit{nlcom6}, based on post-layout analysis at 1.1GHz clock frequency. ............... 94

3-26 Power efficiency degradation by an 800Msamples/s-throughput AMO predistorter of the 1.5nJ/sample energy efficiency. Note that the uncompensated PA operates at same input power and has worse ACPR than the PA with DPD. For same ACPR, the uncompensated PA would have to operate with IBO which leads to lower overall PAE than with our DPD. ...................................................... 95

3-27 To quantify the nonlinearity of the oscilloscope in our measurement setup, an input test sinusoid is filtered heavily to reduce the second and higher harmonics. ........................................ 98

3-28 Block diagram of the transmitter including ML605 evaluation kit, FM-COMMS1 module and ADL5606 evaluation board. ...................... 101

3-29 Experimental setup to characterize nonlinearity of the FMCOMMS1 module with the ADL5606 board. ................................. 102

3-30 Static nonidealities of phase modulator. ................................. 103

3-31 A static model of nonidealities of phase modulator and hardware implementation of its static compensation. ............................ 105

3-32 Static nonidealities of the phase modulator under test before calibration: the most significant being the 87.4 degrees between in-phase and quadrature instead of 90 degrees. ............................ 107

3-33 AM-AM and AM-PM relationships, with the amplitude of the largest input symbol normalized to be 1. ................................. 108

3-34 ACPR of the PA improves while power efficiency decreases at larger IBO. ................................. 109
3-35 ACPR of the PA improves over iterations of compensation.

3-36 Spectrum of uncompensated and compensated output of PA: at 6dB IBO, ACPR improves from -35.1dB to -59.2 after four iterations.

3-37 EVM plots of PA output after four iterations of compensation (with amplitude of the largest symbol at 0dB IBO normalized to be 1); ideal symbols are in red and the output symbols are in blue.

3-38 Complementary cumulative distributions of normalized power of samples (normalized with average power to be 0dB).

3-39 The Lipschitz constants of the iterations.

3-40 Compensator structure proposed in [25].

3-41 Block diagram of predistorter implemented on FPGA for RF PA.

3-42 Because of its symmetric properties and that the center tap vanishes, a (2m + 1)-tap FIR filter \( h_2[n] \) requires only \( m \) DSP48E1 slices, as illustrated above with \( m = 2 \).

3-43 ACPR of the PA improved by real-time predistorter with gains in power efficiency for the same ACPR over the uncompensated PA.

3-44 EVM of the PA improved by real-time predistorter with a small penalty in power efficiency.

3-45 Power efficiency degradation by a 246Msamples/s-throughput predistorter of the 1.7nJ/sample energy efficiency. Note that the uncompensated PA operates at same input power and has worse ACPR than the PA with DPD. To meet the linearity requirements, the uncompensated PA would have to operate with IBO which leads to lower overall PAE than with our DPD.
3-46 The 64QAM constellations on a square grid, and its modified variants. The peak amplitude, normalized to be 1, is shown by the outer green circle. The diameter of red circles is the minimum symbol distance $D_{\text{min}}$. If the designed constellations, given the peak amplitude, supports more than 64 symbols, the symbols with smallest amplitudes are chosen to be redundant symbols to minimize PAPR; redundant symbols are highlighted in yellow.

3-47 Constellations on concentric circles. The peak amplitude, normalized to be 1, is shown by the green circle. The diameter of red circles is $D_{\text{min}}$. Redundant symbols are highlighted in yellow.

3-48 More constellations on concentric circles. The peak amplitude, normalized to be 1, is shown by the green circle. The diameter of red circles is $D_{\text{min}}$. Redundant symbols are highlighted in yellow circles.

3-49 The constellations on a hexagonal grid. The peak amplitude, normalized to be 1, is shown by the green circle. The diameter of red circles is $D_{\text{min}}$. Redundant symbols are highlighted in yellow.

3-50 Different constellations require different receiver complexity: (a) quadrature receiver of 64QAM constellation requires seven thresholds on either in-phase or quadrature, with threshold interval being $\sqrt{2}/7 = 0.2020$; (b) quadrature receiver of 64HEX2 constellation requires 16 thresholds on in-phase and eight thresholds on quadrature, with threshold intervals being $5/42 = 0.1190$ and $5\sqrt{3}/42 = 0.2062$.

3-51 The most closely packed 64-symbol constellation. The peak amplitude, normalized to be 1, is shown by the green circle. The diameter of red circles is $D_{\text{min}}$.

3-52 Tradeoff between the minimum distance $D_{\text{min}}$ (with peak symbol amplitude of each constellation normalized to be 1) and PAPR of 64-symbol constellation design.

3-53 Alternative 256-symbol constellation designs.
3-54 $m$-DENSE constellation versus $m$-QAM constellation, with the peak amplitude of all considered constellations normalized to be 1. . . . . . 133
# List of Tables

2.1 LINC and AMO SCS Equations. ........................................ 28
2.2 Comparison of different weight factors for phase modulation. ....... 36
2.3 Advantages and disadvantages of ASICs and FPGAs. ................. 38
2.4 Comparison of performance between FPGA and ASIC implementa-
tions of a 404-tap FIR filter bank example. ............................ 40
2.5 Comparison of real-time read-only storage-cells: registers versus SRAM. 53
2.6 Area and minimum-energy comparison with other works on digital AMO SCS, with no area constraint. ......................... 53
2.7 Energy comparison with other works on digital AMO SCS, with all designs, except [29], consuming roughly same area (~ 0.8mm²). ... 54
2.8 Area and energy comparison with other works on digital LINC SCS. . 58
3.1 Recent literature on outphasing transmitters with DPD. ............... 69
3.2 Recent literature on 45GHz-carrier transmitters with DPD. ACPR numbers are not reported (NR) in some literatures. ................. 70
3.3 Choices of modeling, with different level of complexities, for each ana-
log part in an outphasing transmitter. .................................... 72
3.4 Comparison of Wilkinson combiner versus zero-degree combiner. ... 75
3.5 ACPR and EVM performance comparisons between using input se-
quence with and without zero-avoidance property for 45GHz LINC system [24]. ......................................................... 76
3.6 LUT configuration of nonlinear transformation part of predistorter for LINC and AMO operation. ........................................... 89
3.7 Size of SRAM instances in \textit{nlcom6} digital baseband. .............. 93
3.8 Third harmonic distortion introduced by the oscilloscope with different sampling rate and different acquisition bandwidths. .............. 99
3.9 Noise on voltage and time measurement by the oscilloscope with different sampling rates and different acquisition bandwidths. ............ 100
3.10 The synchronous setup of FMCOMMS1 module with the ADL5606 board. ................................................................. 101
3.11 ACPR and EVM before and after off-line iterative compensation at different input power back-off. ........................................ 112
3.12 The minimum symbol distance $D_{min}$ and PAPR (before being spec-trally shaped, and assuming that all symbols in a constellation are equally probable and redundant symbols are never used) of different constellation designs, with peak symbol amplitude of each constellation normalized to be 1. .................................................. 130
Chapter 1

Introduction

1.1 Wireless Transmitters with Higher Data Rate but Lower Power Budget Requirements

Nowadays, with constant demands from emerging applications and user experiences, wireless communications are accelerating into the realm of higher data rates at hundreds of megabits to tens of gigabits per second. To support higher and higher data rates, we are simultaneously looking for wider spectral allocation and more efficient usage of the given spectrum. As of today, we do not yet have an exact definition of the fifth generation (5G) wireless communication specifications, but to overcome the interference-limited communication channels and to meet the data rate requirement of gigabits per second it is generally agreed that millimeter wave (mm-wave) technologies will be a key enabler for 5G communication [1–3]. This wider bandwidth in the carrier frequencies of 30GHz to 300GHz requires a fast digital baseband to meet the data throughput demand. Additionally, spectral efficiency demands higher system linearity, deploying modulation schemes with higher peak-to-average power ratio (PAPR) of more than 10dB. With larger dynamic range in signals, the precision requirement on digital baseband is also higher. These trends push the digital baseband into the high-throughput and high-precision realm.

The other significant trend in wireless communication is smaller transmitter out-
put power, into the range of 5 watt to 0.5 watt or even smaller with possible 5G multiple-input and multiple-output (MIMO) arrays. With both better silicon devices and new innovative topologies, the power efficiency of power amplifiers (PAs) keeps improving as well, even at these lower output power levels. With smaller output power, the budget of the transmitter system is significantly reduced. However, to not degrade the energy efficiency of the overall transmitter system, the digital baseband must meet the high-throughput and the high-precision requirements in an energy-efficient manner. We endeavor to meet all such requirements in the following thesis.

1.2 Versatile Digital Baseband Needs to be Energy-Efficient

Being different in bias point and operation, the topology of a single PA is classically categorized into class A, class AB, class B, class D, class E, class F, class F$, etc., with class A being most linear while class E/F being least linear but most power efficient. To break the tradeoff between linearity and amplifier drain efficiency, i.e., to improve power efficiency with less sacrifice to linearity, polar PAs and envelope-tracking PAs can be constructed by adding a dynamic power supply [4]; combining two or more PAs together, Doherty PAs and outphasing PAs are also very popular. In the category of outphasing PAs, the basic configuration is linear amplification with nonlinear components (LINC) [5, 6]. As more complexity is allowed on amplitude control, multi-level LINC (ML-LINC) [7] improves average power efficiency. Additionally, asymmetric-multilevel-outphasing (AMO) [8–12] has also been proposed in recent years.

These different architectures of PAs require completely different input signals and need vastly different baseband functionalities to support them. Specific to outphasing transmitters, the digital baseband needs to perform the task of a signal component separator (SCS) [13–18]: decomposing each transmitted vector signal into two vector
signals with different phases and restricted amplitudes. For N-Way Doherty PAs [19–21], the digital baseband needs to perform a different task of SCS. For a hybrid architecture of outphasing and Doherty, the baseband task of component separation becomes even more complex.

Signal component separation is one of the many power hungry tasks to perform in the digital baseband. Regardless of which transmitter architecture is used, every transmitter needs to modulate the baseband signal onto the carrier frequency. Non-idealities of phase modulation usually have to be calibrated or compensated for in the digital baseband.

After all linearization efforts through new innovative transmitter architectures, the tradeoff between linearity and power efficiency is still central to all PA designs. While analog designers push for higher and higher power efficiency of amplifiers, digital designers have to design better-performing and more energy-efficient digital predistorters (DPDs) in baseband to compensate for the nonlinearities and memory effects of nonideal analog components.

All these tasks, either originating from the system architecture or from nonidealities of an analog component, demand a range of versatile and capable digital basebands that are both high-performance and energy-efficient.

1.3 Thesis Contributions

In this thesis we explore the efficient baseband implementation of two essential transmitter signal processing functions, outphasing and nonlinear compensation.

1.3.1 Signal Component Separator for Outphasing Transmitters

In this work, we explore the implementation design space of AMO SCS, based on the optimized piece-wise linear (OPWL) approximation approach, by utilizing the changes in micro-architecture (parallelism and pipelining), choice of storage elements
of static random-access memories (SRAMs) versus flops, and aggressive back-end power optimization techniques (non-critical path Vt and sizing reductions). The OPWL approximation approach, which will be reviewed in section 2.3.1, was a joint work first published in [22]. In the following work, the emphasis is on the further energy footprint reduction with changes in micro-architecture, choice of memory, and aggressive back-end power optimization techniques.

With a combination of these techniques, 2\times energy and 100\times area savings can be achieved over the traditional approach, resulting in a record energy-efficiency and throughput for AMO and LINC SCS of 32pJ/sample and 22pJ/sample at 0.6V supply voltage, at 400Msamples/s and areas of 0.41mm\(^2\) and 0.38mm\(^2\) in a 45nm silicon on insulator (SOI) process [23].

### 1.3.2 Efficient Digital Compensator for Power Amplifier

We set up a simulation testbench consisting of Q-band (45GHz) outphasing transmitters, perform system identification, and propose DPD architecture for the transmitter. To facilitate iteration convergence, we design a zero-avoidance shaping filter to replace the traditional shaping filter in the system. We extend the OPWL approximation algorithm to an optimized multi-dimension piece-wise quadratic approximation in the predistorter design, and make extensive use of energy-efficient design techniques to reduce the energy and area footprint of the zero-avoidance shaping filter and the DPD. We improve adjacent-channel-power-ratio (ACPR) of the outphasing transmitter from -30.6dB to -44.0dB and reduce error vector magnitude (EVM) from 4.5\% down to 1.0\% with real-time zero avoidance. The heuristics of the zero-avoidance algorithm was outlined by another team member and first published in [24]. In this work, the emphasis is on the efficient realization of the zero-avoidance shaping filter.

To illustrate the wide applicability of this proposed linearization methodology, we apply it to compensate for distortion in a radio-frequency (RF) PA. We apply the off-line iterative compensation method to a PA with 1.97GHz carrier frequency and 737Mbps data throughput with 64-Quadrature-Amplitude-Modulation (64QAM) scheme. We map the desired DPD structure onto FPGA and improve the transmitter
ACPR from -36dB to -48dB. The uncompensated PA has to back-off the input power by 12dB to meet an ACPR requirement of -48dB at a power efficiency of 3.3%, while the compensated PA has to back-off by only 6dB and even after having included the DPD power, the overall transmitter power efficiency is 9.2%, almost 3× the efficiency of uncompensated PA meeting the same -48dB ACPR requirement. The analytically motivated predistorter model structure was proposed by other team members and was published in [25]. In this work, the emphasis is on the experimental verification of this model structure and on the realization of the model onto hardware.

1.4 Thesis Overview

The thesis is organized as follows.

Chapter 2 is devoted to the energy and area efficiency of digital baseband design. It starts with a short introduction of the outphasing PA architecture and the task of SCS. After having explained the motivation behind implementing the separator in an ASIC rather than an FPGA, it dives into details on how to design SCS in an area- and energy-efficient way. After having briefly compared LINC SCS versus AMO SCS, the chapter concludes with an analysis of the energy penalty of SCS on the efficiency of the overall transmitter system.

Chapter 3 presents the predistorter design for the PAs. It starts with an overview of digital predistortion techniques and discusses the sources of nonlinearity in the system. It goes into the design of the zero-avoidance shaping filter, specifically in the context of compensation for the outphasing transmitter. It moves on to discuss the dynamic predistorter model for RF PAs and energy-efficient design of a predistorter on FPGAs. It ends with an exploration of the constellation design to reduce bit error rate and to increase transmission range without upgrading to a PA with larger peak output power.

Chapter 4 concludes the thesis and suggests several directions for further research.
Chapter 2

Energy and Area Efficiency of High-Throughput Digital Baseband

2.1 Overview of Outphasing Transmitters

Outphasing power amplifier architectures, combining the different outputs of two or possibly more PAs, have been used to improve power efficiency by utilizing more efficient nonlinear PAs. LINC [5, 6] improves peak power efficiency with less sacrifice in linearity than a single PA, while ML-LINC [7] and AMO [8–12] aggressively improve average power efficiency with more complexity on amplitude control. The mathematical relations between the input vector $I + jQ$ and the two decomposed vectors $a_1e^{j\varphi_1}$ and $a_2e^{j\varphi_2}$ are illustrated in Fig. 2-1.

As mm-wave communication bandwidths are rapidly increasing to multi-GHz with desire for increased spectral efficiency, and signaling rates in cellular bands increase with channel aggregation, the power consumption of digital back-ends involved in processing samples threatens to degrade the overall transmitter efficiency. The main digital baseband bottleneck for outphasing transmitter architectures like LINC, ML-LINC, and AMO is the signal component separator.

In the LINC architecture, an SCS decomposes each transmitted sample vector signal into two vector signals with different phases but same amplitude as shown in Fig. 2-1. Each of the two decomposed vector signals is first modulated by the phase
modulator and then amplified with the fixed same amplitude by the high-efficiency switching PAs. Finally the two phase-modulated and amplified vectors are combined to deliver the final output. In the ML-LINC architecture, the two component vector signals have different phases but the same amplitude chosen from a set of restricted amplitude choices as illustrated in Fig. 2-1. In the AMO architecture, the two vector signals have different amplitudes from a set of restricted choices, for example, four choices as illustrated in Fig. 2-1 with corresponding architecture as shown in Fig. 2-2. In either one of these outphasing architectures, an SCS involves complex functional computations as shown in Table 2.1 [22], which require large silicon area, and at high-throughputs can degrade the overall power-added efficiency.

Figure 2-1: Vector decompositions in different outphasing architectures assuming four restricted power supply levels indicated by blue solid or black dashed arc lines.

2.1.1 Digital Baseband for Outphasing Transmitters

The typical low-throughput LINC SCS and recent ML-LINC and AMO implementations [14, 26–29] usually involve the use of a coordinate rotational digital computer (CORDIC) [30] and lookup table (LUT) map for the nonlinear functions [27,31]. Recently, we proposed the optimized fixed-point piece-wise linear (OPWL) algorithm for SCS design [22], and a corresponding chip nlcom2 was implemented using this approach to compute the nonlinear functions. The implemented AMO SCS has a
Figure 2-2: Overview of AMO PA system.
block diagram as shown in Fig. 2-3b, in contrast with the block diagram of a LINC SCS in Fig. 2-3a. At high-throughputs and resolution requirements, this method is significantly more area- and energy-efficient than LUTs, CORDIC, or nonlinear polynomial filters. However, to enable a new class of wide-band wireless communication system designs with lower output power and high spectral efficiency, we feel that even further power and area reduction of the SCS is necessary.

Table 2.1: LINC and AMO SCS Equations.

<table>
<thead>
<tr>
<th>LINC Equations</th>
<th>AMO Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A = \sqrt{T^2 + Q^2} )</td>
<td>( A = \sqrt{T^2 + Q^2} ),</td>
</tr>
<tr>
<td>( \theta = \arctan \left( \frac{Q}{T} \right) ) (linc1)</td>
<td>( \theta = \arctan \left( \frac{Q}{T} \right) ) (amo1)</td>
</tr>
<tr>
<td>( \alpha = \arccos \left( \frac{A}{2a} \right) ) (linc2)</td>
<td>( \alpha_1 = \arccos \left( \frac{a_1^2 + A^2 - a_2^2}{2a a_1} \right) ) (amo2)</td>
</tr>
<tr>
<td>( \phi_1 = \theta + \alpha ), ( \phi_2 = \theta - \alpha ) (linc3)</td>
<td>( \phi_1 = \theta + \alpha_1 ), ( \phi_2 = \theta - \alpha_2 ) (amo3)</td>
</tr>
<tr>
<td>( f(\phi_1) = \frac{1}{1 + \tan(\phi_1)} ), ( f(\phi_2) = \frac{1}{1 + \tan(\phi_2)} ) (linc4)</td>
<td>( f(\phi_1) = \frac{1}{1 + \tan(\phi_1)} ), ( f(\phi_2) = \frac{1}{1 + \tan(\phi_2)} ) (amo4)</td>
</tr>
</tbody>
</table>

It is important to quantify the digital baseband efficiency in the context of overall transmitter efficiency. In Fig. 2-4, we plot the raw RF power-added efficiency (PAE) versus output saturation power (Psat) in blue crosses for PAs above 10GHz on silicon-based technologies (either bulk CMOS, SOI CMOS, or SiGe), from IEEE references in year 2006 to year 2012. Pushing the design point upwards towards higher RF power efficiency is the goal of every PA designer. However, if the digital baseband is not efficient, it will drag the design point downwards towards lower overall system power efficiency.

A previous state-of-the-art design of an AMO SCS, implemented in the 90nm CMOS technology with 8-bit phase resolution, achieved a throughput of 40Msamples/s as reported in [29], consuming 0.36mW of power. The traditional design does not scale well with higher phase precision and higher throughput requirements of emerging mm-wave technologies. If the traditional design is scaled to meet a 12-bit phase precision requirement still at a throughput of 40Msamples/s, it would consume...
Figure 2-3: The block diagrams of AMO SCS and LINC SCS.
0.36/(8 \times 2^8) \times (12 \times 2^{12}) = 8.64\text{mW} \text{ of power. If that design is migrated from the 90nm CMOS technology to a more advanced 45nm technology, its power consumption would be approximately halved to 4.32mW. To operate at a reasonably high throughput, say 800Msamples/s (to achieve a data rate of 2.4Gbps with 64QAM modulation and oversampling ratio of 2), a 20-way parallelized version of the traditional AMO SCS will consume at least 86mW without consideration of the serialization overhead in 20-way parallelization.}

The contour lines in Fig. 2-4 list the PAs where the overall system efficiency will be degraded by 0.1% or 1% respectively by the 86mW power of traditional SCS. For example, for any PAs with output power up to 100mW and higher than 12% power efficiency, the overall power-efficiency of the system will degrade by more than 1% by including the 86mW SCS power. To visualize the efficiency penalty due to the energy footprint of the digital baseband, power efficiency of the system after having included the 86mW AMO SCS power is also plotted (in red) on Fig. 2-4. In order to reduce the gap from the RF efficiency (blue) to the overall system efficiency (red), we have to implement the SCS in a more energy-efficient way.

### 2.1.2 Nonideal Phase Modulation

Let us look at the mathematics of phase modulation and typical circuit implementation. Equation (linc4) or (amo4) in Table 2.1 are nonlinear functions from the phases in radians to the inputs of the phase modulator: 

\[
f(\phi_1) = \frac{1}{1 + \tan(\phi_1)}, \quad f(\phi_2) = \frac{1}{1 + \tan(\phi_2)}.
\]

These nonlinear functions originate due to higher precision requirements of the system. These nonlinear functions are not native to the LINC or AMO architecture, but common to the circuit implementation of a type of phase modulator. A phase modulator takes the in-phase carrier component \(\sin(\omega t)\), the quadrature carrier component \(\cos(\omega t)\), and their respective weight factors \(c_i\) and \(c_q\) as inputs. Then, as shown in Fig. 2-5, it produces a modulated signal \(y(t)\):

\[
y(t) = c_i(\phi[nT]) \sin(\omega t) + c_q(\phi[nT]) \cos(\omega t)
\]

\[
= A(\phi[nT]) \sin(\omega t + \phi[nT] + e(\phi[nT])).
\]  

(2.1)
Ideally, the phase modulation would have a zero phase error $e(\varphi[nT]) = 0$ and produce an amplitude $A$ independent of the phase input $\varphi[nT]$. So, if we have the following weight factors given the phase $\varphi$:

$$c_i(\varphi), c_q(\varphi) = \cos \varphi, \sin \varphi \tag{2.2}$$

then the phase modulator outputs the following,

$$y(t) = \cos (\varphi[nT]) \sin (\omega t) + \sin (\varphi[nT]) \cos (\omega t)$$

$$= \sin (\omega t + \varphi[nT]). \tag{2.3}$$

which gives us a time-invariant constant amplitude and the desired phase $\varphi[nT]$.

However, considering the physical implementation of circuits, it is expensive to compute and inconvenient to use both weight factors $c_i(\varphi) = \cos \varphi$ and $c_q(\varphi) = \sin \varphi$. 
with the normalization constraint

\[ c_i(\varphi)^2 + c_q(\varphi)^2 = 1. \] (2.4)

To simplify the derivation of the weighting factors, these sinusoidal functions are approximated by triangular ones with the normalization constraint [32,33]

\[ |c_i(\varphi)| + |c_q(\varphi)| = 1, \] (2.5)

as illustrated in Fig. 2-6a, resulting in:

\[
\begin{align*}
c_i(\varphi), c_q(\varphi) &= \begin{cases} 
1 - \frac{2}{\pi} \varphi, & \frac{2}{\pi} \varphi \text{ for } 0 \leq \varphi < \frac{1}{2}\pi \\
1 - \frac{2}{\pi} \varphi, & 2 - \frac{2}{\pi} \varphi \text{ for } \frac{1}{2}\pi \leq \varphi < \pi \\
-3 + \frac{2}{\pi} \varphi, & 2 - \frac{2}{\pi} \varphi \text{ for } \pi \leq \varphi < \frac{3}{2}\pi \\
-3 + \frac{2}{\pi} \varphi, & -4 + \frac{2}{\pi} \varphi \text{ for } \frac{3}{2}\pi \leq \varphi < 2\pi
\end{cases}
\end{align*}
\] (2.6)

These approximated weight factors in Equation (2.6) can be easily computed without any computation of trigonometric functions. The simple normalization constraint
(2.5) can reduce the number of current digital-to-analog converter (DAC) in use from two to only one. With only one current DAC, area and power are nearly halved, and the number of current sources requiring strict matching are also halved [34]. Another advantage of this normalization (2.5), compared to (2.4), is that it keeps the common mode level of the phase modulator output constant.

However, as shown in Fig. 2-6b, the phase error introduced by the above approximations is significant. The maximum phase error is 4 degrees, so this approximation could be used for any system with precision requirement lower than 6-bit since 1 LSB is 5.625 degrees in 6-bit phase modulation. For applications which require precision in phases above 6 bits, alternative weight factors $c_i(\varphi)$ and $c_q(\varphi)$ have to be derived.

As shown in Fig. 2-7, to keep the convenient normalization constraint $|c_i(\varphi)| + |c_q(\varphi)| = 1$ without any sacrifice in phase precision, we use the following weight factors to achieve $e(\varphi) = 0$:

$$c_i(\varphi), c_q(\varphi) = \frac{1}{1 + \tan \varphi} \frac{\tan \varphi}{1 + \tan \varphi}.$$  \hspace{1cm} (2.7)

Because of this convenient normalization constraint, the weights in Equation (2.7) requires computation of only one nonlinear function, instead of two as in Equation (2.2). With the new weights, we still have the benefits of using one current DAC instead of two and of constant output common mode level, as with Equation (2.6). The phase modulator outputs the following,

$$y(t) = \frac{1}{1 + \tan \varphi[nT]} \sin(\omega t) + \frac{\tan \varphi[nT]}{1 + \tan \varphi[nT]} \cos(\omega t)$$

$$= \frac{1}{\cos \varphi[nT] + \sin \varphi[nT]} \sin(\omega t + \varphi[nT]),$$  \hspace{1cm} (2.8)

which has the desired phase $\varphi$; however, as seen in Fig. 2-8b, it still has a nonconstant amplitude $|A(\varphi)| = \frac{1}{\cos \varphi + \sin \varphi} = \frac{1}{\sqrt{2} \sin(\varphi + \pi/4)}$ which is phase-dependent. The amplitude is at minimum of $\frac{1}{\sqrt{2}}$ when $\varphi$ is $\pi/4$, $3\pi/4$, $5\pi/4$, or $7\pi/4$ radians, and is at maximum of 1 when $\varphi$ is 0, $\pi/2$, $\pi$, or $3\pi/2$ radians; the amplitude variation is $20 \log_{10} \frac{\sqrt{2}}{1} = 3$dB, the same as with weight factors in Equation (2.6).

In an outphasing amplifier system, we expect to have preamplifiers between the
(a) Weight factors of ideal phase modulation in Equation (2.2) versus its triangular approximation in Equation (2.6).

(b) Phase error $e(\varphi)$ and amplitude variation $A(\varphi)$, introduced by the triangular approximation in Equation (2.6).

Figure 2-6: Weight factors of triangular approximation in Equation (2.6).
Figure 2-7: An illustration of the factor weights $c_i$ and $c_q$ for phase modulation and the corresponding normalization constraints.

Phase modulators and the PAs. These preamplifiers are usually saturated, for example implemented as inverters driven into saturation to give rail-to-rail outputs, to serve as phase-preserving amplitude limiters and mitigate the phase-dependent amplitude variation. Therefore, we have the phase-to-weight mapping equations (linc4) without worrying too much about amplitude variation: $f(\varphi_1) = \frac{1}{1 + \tan(\varphi_1)}$, $f(\varphi_2) = \frac{1}{1 + \tan(\varphi_2)}$.

Different weight factors for phase modulation, Equation (2.2), (2.6), (2.7), are summarized and compared in Table 2.2. Whenever amplitude variation could be tolerated or mitigated in later stages of transmitter signal chain, the triangular approximation (2.6), which requires no computation of any trigonometric function, should be used for applications with 6 bits or less of phase precision; (2.7) should be used instead for any applications with more than 6 bits of phase precision.

In general, Equation (linc4) will differ depending on the actual implementation of the phase modulator. To make them flexible and compatible with different designs of phase modulators, Equation (linc4) should be implemented to be programmable rather than fixed $\frac{1}{1+\tan \varphi}$ functions. Additionally, in-phase versus quadrature gain...
Figure 2-8: Weight factors of triangular approximation in Equation (2.7).

Table 2.2: Comparison of different weight factors for phase modulation.

<table>
<thead>
<tr>
<th>weight factors</th>
<th>normalization</th>
<th>number of DACs</th>
<th>amplitude variation (dB)</th>
<th>phase error (degree)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ideal (2.2)</td>
<td>$c_i(\varphi)^2 + c_q(\varphi)^2 = 1$</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>triangular (2.6)</td>
<td>$</td>
<td>c_i(\varphi)</td>
<td>+</td>
<td>c_q(\varphi)</td>
</tr>
<tr>
<td>(2.7)</td>
<td>$</td>
<td>c_i(\varphi)</td>
<td>+</td>
<td>c_q(\varphi)</td>
</tr>
</tbody>
</table>
mismatch, offsets, and any static nonlinearity of phase modulator, could be also considered and easily taken care of in the programmable computations of these weight factors $c_i$ and $c_q$.

2.2 Motivation for ASIC Implementation

The digital baseband design requires careful optimization of energy to minimize degradation to overall system energy efficiency at high-throughput. Here we show the motivation for using application specific integrated circuits (ASICs) to implement the digital baseband, rather than using the popular alternative implementation of field programmable gate arrays (FPGAs).

Compared to ASICs, FPGAs provide different benefits to digital circuit designers and can often be a compelling alternative. Designers must carefully evaluate the engineering tradeoffs before choosing one over the other. Today’s FPGAs easily push the performance barrier at a few hundred megahertz, and provide an unprecedented increase in logic density. Looking at design cycle time and cost, FPGAs provide more favorable and flexible solutions. FPGA solutions require no manufacturing steps, while ASIC solutions have long turn-around times because of mask and manufacturing steps. FPGA solutions do not incur any upfront non-recurring expenses, while ASIC solutions have significant engineering design time and mask costs. However, for our system, the high-throughput performance requirement and stringent energy constraints make ASICs the compelling choice. We choose to implement in ASICs to minimize degradation to overall system energy efficiency. The gap between FPGAs and ASICs has narrowed over the years, but right now, ASICs are still the implementation of choice for a high-throughput system with very small energy budget. Nevertheless, FPGAs serve as an excellent alternative for functional verification. A summary of FPGA versus ASIC pros and cons can be seen in Table 2.3.

To better understand the throughput performance and energy efficiency gap between FPGAs and ASICs, we design a finite impulse response (FIR) filter as an example to compare these two implementations. The FIR filters are utilized in many
Table 2.3: Advantages and disadvantages of ASICs and FPGAs.

<table>
<thead>
<tr>
<th></th>
<th>FPGAs</th>
<th>ASICs</th>
</tr>
</thead>
<tbody>
<tr>
<td>design cycle time</td>
<td>short (weeks)</td>
<td>long (months to a year)</td>
</tr>
<tr>
<td>engineering cost</td>
<td>low</td>
<td>high (especially at deep sub-micron nodes)</td>
</tr>
<tr>
<td>performance</td>
<td>below gigahertz</td>
<td>up to a few gigahertz</td>
</tr>
<tr>
<td>power</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>consumption</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>reprogrammability</td>
<td>high</td>
<td>very low</td>
</tr>
</tbody>
</table>

places in the digital baseband design, including the spectrum shaping filter and the dynamic predistorter. The Z-transform of a \((m+1)\)-tap FIR filter \(h[n]\) is expressed as follows:

\[
H(z) = \sum_{n=0}^{m} h[n]z^{-n}.
\] (2.9)

This specific example is a bank of four FIR filters, each with 101 taps and 11-bit resolution. We look at the post-place-and-route power estimation of this FIR design implemented on Virtex-6 device XC6VHX380T and on a 45nm SOI technology. We choose XC6VHX380T in the Xilinx Virtex-6 family, because it is optimized for ultra high-performance DSP with highest speed grade of -3 available. The XC6VHX380T device has faster speed and more available DSP slices per column than the common XC6VLX240T on a Virtex-6 ML605 evaluation kit to allow a more valid and relevant comparison. The XC6VHX380T device has 864 DSP slices of DSP48E1 with Fmax being 600MHz with all pipelined registers used, and 28Mb of block RAM with Fmax also being 600MHz. Each of the DSP48E1 slices has a \(25 \times 18\) multiplier [35] and becomes the major building block for FIR filter as shown in Fig. 2-9. The device has six columns and 144 DSP48E1 slices per column, so we can fit each FIR of 101 taps inside a single column, and make efficient use of ACIN, ACOUT, PCIN, and PCOUT signals which are dedicated routing paths internal to the DSP48E1 column and are not accessible via fabric routing resources.

We implement the FIR filter in both the XC6VHX380T FPGA and the 45nm SOI
(a) The Direct Form structure of a \((m + 1)\)-tap FIR filter.

(b) The \((m + 1)\)-tap FIR filter in (a) with additional \((m + 1)\) pipeline stages.

(c) The \((m + 1)\)-tap FIR filter in (b) mapped on to DSP48E1 slices, with one additional pipeline stage on multipliers.

Figure 2-9: Mapping an FIR filter onto DSP48E1 slices on an FPGA. The two clock cycles of input delay generated inside the DSP48E1 slice and the dedicated cascade PCIN/PCOUT and ACIN/ACOUT connections internal to the column are utilized to achieve maximum performance irrespective of the number of coefficients [35].
technology, and power estimation numbers are reported in Table 2.4. We force the junction temperature to be 25 degrees Celsius for easier comparison; though, with similar heating dissipation strategy, the FPGA would get hotter and the increase in its leakage power would be even higher. For this FIR filter, where the design is dominated by multiplications and additions, we see 16× energy-savings at 2× higher throughput of the ASIC over the FPGA without even utilizing voltage scaling.

Table 2.4: Comparison of performance between FPGA and ASIC implementations of a 404-tap FIR filter bank example.

<table>
<thead>
<tr>
<th></th>
<th>FPGAs</th>
<th>ASICs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>Virtex-6 (40nm)</td>
<td>45nm SOI</td>
</tr>
<tr>
<td>Junction temperature (°C)</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Clock speed (MHz)</td>
<td>600</td>
<td>1250</td>
</tr>
<tr>
<td>Power (W)</td>
<td>6.367</td>
<td>0.782</td>
</tr>
<tr>
<td>Energy efficiency (nJ/sample)</td>
<td>10.6</td>
<td>0.63</td>
</tr>
<tr>
<td>Energy efficiency (pJ/sample/tap)</td>
<td>26.3</td>
<td>1.55</td>
</tr>
</tbody>
</table>

As we have clearly shown the limitations of FPGAs in throughput and energy-efficiency, in this chapter we have chosen to implement our digital baseband in ASICs to satisfy our power budget. In addition, the ASIC solution for the digital baseband also allows a truly integrated system which saves off-chip IO bandwidth and reduces the on-the-board further integration cost. As we will see later, in many transmitter signal processing applications even the ASIC implementation degrades the PAE by 1-2% for output powers of interest, making an FPGA implementation (which is at least 16x worse in energy) a true no-go.

### 2.3 Techniques to Achieve Higher Energy Efficiency

Let us now take a look at the techniques to achieve higher energy efficiency within the ASIC implementation. In the following sections we will review the OPWL approximation approach to compute nonlinear functions (Section 2.3.1), and reduce leakage power using backend optimization (Section 2.3.2). We will also discuss micro-architecture optimizations, look at pipelining versus parallelism, and apply this to out
AMO SCS design (Section 2.3.3). With a comparison of registers versus SRAMs, we will reduce the area footprint so that we can parallelize to meet the throughput requirements (Section 2.3.4).

### 2.3.1 Optimized Piece-Wise Linear Approximation

The typical low-throughput LINC SCS and recent AMO SCS implementations usually involve the use of CORDIC or a LUT map for these nonlinear functions. The maturity of the CORDIC algorithm and simplicity of the LUT approach make them suitable for LINC SCS applications with throughputs below 100Msamples/s and with low to medium resolution (8 bits or less). However, these approaches become less attractive or even prohibitive for wide-band applications with complex modulation schemes where the throughput is in the Gsamples/s range with high phase resolution (10 bits and more).

Especially dedicated to applications with high throughout and high precision requirements, the OPWL approximation approach we proposed in [22] provides a balance among precision, power, and area. All nonlinear functions involved in the SCS computations are smooth over almost the whole input range. This consideration led us to the OPWL approximation of the nonlinear functions.

Looking at the computation of a nonlinear function of an \(m\)-bit output with \(m\)-bit input \(x \in [0,1)\), \(m\)-bit input \(x\) can be decomposed to \(x_1\) and \(x_2\) as \(x = \left[ \begin{array}{c} \overbrace{x_1}^{m_1-\text{MSB bit}} \ , \ \overbrace{x_2}^{m_2-\text{LSB bit}} \end{array} \right] \), where \(m = m_1 + m_2\). Naturally, \(x_1\) divides the input range into \(2^{m_1}\) intervals and is the indexing number of those intervals. We proposed a fixed-point OPWL scheme such that:

\[
y_i = \frac{b_i \cdot 1}{m_1-\text{MSB bit}} + \frac{k_i(x_2 - S_i \cdot 1)}{m_2-\text{LSB bit}}, \quad i = 0, 1, ... 2^{m_1} - 1. \tag{2.10}
\]

Here, \(y_i = [y([i, 0]), y([i, 1]), \ldots, y([i, N_2 - 1])]^T\), \(x_2 = \frac{1}{N}[0, 1, \ldots, N_2 - 1]^T\), \(1 = [1, 1, \ldots, 1]^T \in \mathbb{R}^{N_2}\), \(N_1 = 2^{m_1}\), \(N_2 = 2^{m_2}\), \(N = 2^m\), \(m = m_1 + m_2\), \(k_i, S_i, b_i \in \mathbb{R}\) and \(y_i\) and \(x_2\) are vectors over fixed-point numbers. The underlying idea of this
formulation is to compute the $m$-bit output part by part. In the linear function of each interval, we use the term $b_i$ to represent the most significant $m_1$ bits of the function value, and the term $k_i \cdot (x_2 - S_i \cdot 1)$ to achieve the lower-significant $m_2$ bits of precision. Then $y_i$ is simply the concatenation of the two parts. The procedures to find the fixed-point representations of the three parameters $k_i, S_i, b_i$ in (2.10) are described in [22].

![Figure 2-10: Micro-architecture of the OPWL approximation in hardware.](image)

As shown in Fig. 2-10, there are essentially three arithmetic operations involved in the OPWL hardware: one LUT, one adder, and one multiplier. The LUT takes the $m_1$ MSBs of the input as the address and outputs the parameters $b_i \cdot k_i \cdot s_i$ in the corresponding interval. Since $m_1 \approx \lceil m/2 \rceil < m$, the LUT size is significantly smaller than the direct LUT case, especially for high-precision applications. From Fig. 2-10, we notice that for all arithmetic computations, the operands have only $m_1$, $m_2$, or $l_s + m_2$ bits, but not $m$ bits as input; short operands enable high throughput with low power consumption.

Our first AMO SCS chip *nlcom2*, implemented with 12-bit phase quantization using this OPWL approximation, achieves a minimum-energy point of 58pJ/sample measured at 800Msamples/s [22]. There are two different ways of implementing the LUT in the OPWL hardware: programmable and hardwired. The `getTheta` and `getAlpha` blocks in Fig. 2-3b implement fixed mathematical functions and require no programmability, thus are implemented with hardwired LUTs. As discussed in Section 2.1.2, the `getPhi` block not only computes the \( \frac{1}{1+\tan \varphi} \) functions but also com-
pensates for static nonlinearity of the phase modulator, thus is implemented with programmable LUTs. The high percentage of leakage power in this design (due to register-based programmable coefficient storage and low-Vt use) limits the voltage scaling and put its minimum-energy point at a high supply voltage of 0.7V, limiting the energy-efficiency gains from parallelism and supply scaling. By applying back-end leakage optimization, coefficient storage optimization, as well as various micro-architecture techniques (parallelism, removal of over-pipeline stages, etc.) in the later section, we illustrate the design trade-offs and further optimize the energy-efficiency and area of SCS design.

### 2.3.2 Backend Optimization

To reduce energy per operation on the AMO SCS chip, we use voltage scaling. Power-delay product, interpreted as the amount of energy spent in each transition, scales with supply voltage quadratically, only if we assume that only the dynamic power component of the power dissipation is important. However, our previous AMO SCS design in [22] has significant leakage power. Its significant leakage power component has hindered the supply scaling of AMO SCS from reaching a lower minimum-energy point at a lower supply voltage. To allow for the aggressive tradeoff of silicon area for better energy efficiency (Section 2.3.3), first we have to reduce the leakage power component in the SCS power dissipation.

The 45nm SOI CMOS process provides us with multi-threshold voltage libraries, with the high threshold voltage (HVT) standard cells consuming only 1/2 of the leakage power of the regular threshold voltage (RVT) standard cells and the ultra-high threshold voltage (UVT) standard cells consuming only 1/10 of the leakage power of the regular threshold voltage standard cells (Fig. 2-11). For the critical paths, the RVT standard cells have to be deployed to meet timing at cost of higher leakage power; however, part of the noncritical paths can be replaced using UVT standard cells without degrading the setup timing slack. Registers in the scan-register-based programmable LUTs are static during real-time high-speed operations and do not have to be fast. These storage-cells can be implemented with UVT standard cells.
instead of RVT to reduce leakage power.

Figure 2-11: The ultra-high threshold voltage standard cells consume only 1/10 of the leakage power of the regular threshold voltage standard cells, at the expense of reduced maximum performance.

As shown in the flow chart in Fig. 2-12, we perform aggressive leakage power optimization. In the earliest stage of the flow, we specify all storage-cells used for OPWL computations to be synthesized with only UVT standard cells. In the postroute stage, we reduce buffer and register sizing or replaced RVT standard cells with UVT on non-critical path logics. Using this aggressive backend optimization, we design an AMO SCS with 2GHz target clock rate. Based on the post place and route simulation, the energy and throughput of the design are shown in blue crosses in Fig. 2-15 as a function of supply scaling. The design has been heavily pipelined to meet the throughput of 2Gsamples/s at nominal supply.
Figure 2-12: The ASIC design flow with leakage reduction stages highlighted in red.

- Synthesis
  ▸ Floorplan and Power Grid
  ▸ Placement
  ▸ Clock Tree Synthesis
  ▸ Power Routing
  ▸ Signal Routing
  ▸ Leakage Optimization
    ▸ Storage-cells specified to synthesize with UVT cells only
    ▸ Sizing reductions or RVT cells replaced by UVT cells on noncritical path
  ▸ Filler Placement
    ▸ Sign-off Static Timing Analysis
2.3.3 Micro-Architecture

The three major components of power consumption in digital circuits are summarized in the following equation [36]:

\[
P_{\text{total}} = P_{\text{dynamic}} + P_{\text{leakage}}
\]

\[
= P_{\text{switching}} + P_{\text{short-circuit}} + P_{\text{leakage}}
\]

\[
= \alpha_{0 \rightarrow 1} C_L V_{dd} f_{clk} + I_{sc} V_{dd} + I_{\text{leakage}} V_{dd}
\]

(2.11)

where \(\alpha_{0 \rightarrow 1}\) is the node transition activity factor, \(C_L\) is the loading capacitance and \(f_{clk}\) is the clock frequency. As we are using a standard ASIC tool and flow on static logic with full swing, the voltage swing \(V\) in the equation is the same as the supply voltage \(V_{dd}\). The leakage current \(I_{\text{leakage}}\) depends on the NMOS and PMOS transistor threshold voltages \(V_{Tn}\) and \(|V_{Tp}|\) as well as fabrication technologies. The short circuit \(I_{sc}\) flows directly from supply to ground when both NMOS and PMOS transistors are active simultaneously. The short-circuit power is generally much smaller than switching power, and as we lower supply voltage we guarantee \(V_{dd} < V_{Tn} + |V_{Tp}|\), the short-circuit power is eliminated. We approximate Equation (2.11) with the following:

\[
P_{\text{total}} = P_{\text{dynamic}} + P_{\text{leakage}}
\]

\[
\approx P_{\text{switching}} + P_{\text{leakage}}
\]

\[
= \alpha_{0 \rightarrow 1} C_L V_{dd}^2 f_{clk} + I_{\text{leakage}} V_{dd}
\]

(2.12)

where we consider only full-swing static CMOS circuits and ignore short-circuit power.

The power-delay product, interpreted as the amount of energy spent in each transition, is then expressed as:

\[
E = C_{\text{eff}} V_{dd}^2 + \frac{I_{\text{leakage}} V_{dd}}{f_{clk}}
\]

(2.13)

with \(C_{\text{eff}} = \alpha_{0 \rightarrow 1} C_L\) being the effective capacitance switched. We assume the maximum performance or clock frequency \(f_{clk}\) achievable at \(V_{dd}\) is given by the following
approximation [36]:

\[
f_{\text{clk}} = k \left( \frac{V_{dd} - V_{Tn}}{V_{dd}} \right)^{\alpha},
\]

(2.14)

where \( k \) and \( \alpha \) are empirical parameters.

Long channel transistors are well described with \( \alpha = 2 \), while short channel transistors in the deep submicron technologies are better described with a smaller \( \alpha \) not much larger than 1. With this approximation on the maximum performance of circuits, now we have

\[
E = C_{\text{eff}}V_{dd}^2 + \frac{I_{\text{leakage}}V_{dd}^2}{k(V_{dd} - V_{Tn})^\alpha}
\]

(2.15)

where the dynamic part decreases quadratically with lowering supply voltage \( V_{dd} \); however, the leakage part increases with lowering \( V_{dd} \) when \( V_{dd} \) is relatively close to \( V_{Tn} \). This describes the minimum-energy point beyond which lowering supply voltage further will not reduce the power-delay product. Therefore, we can reduce supply voltage to achieve lower energy per operation, but at a lower throughput \( f_{\text{clk}} \).

At the minimum-energy point, we have to replicate multiple copies of the circuits to meet the throughput requirement.

For example, if we have a digital system with four functional stages, as shown in Fig. 2-13a to meet a throughput \( f_{\text{ref}} \) at a supply voltage of \( V_{\text{ref}} = \frac{T}{4} \) where \( T \) is the clock period, the dynamic power is \( C_{\text{eff}}V_{\text{ref}}^2f_{\text{ref}} \). If we pipeline the system as shown in Fig. 2-13b, we can possibly meet a throughput of almost \( 2f_{\text{ref}} = \frac{1}{T/2} \) with the same supply voltage of \( V_{\text{ref}} \), given that the delay of stages F and G equals the delay of stages H and K and the sum of minimum setup time and clock-to-output delay of the pipeline register is much smaller than the clock period. Instead of running at a throughput of \( 2f_{\text{ref}} \) at \( V_{\text{ref}} \), we could lower the voltage supply to, for illustration purpose, \( 0.7V_{\text{ref}} \), to lower the throughput to \( f_{\text{ref}} \). Then, the dynamic power is \( C_{\text{eff}}(0.7V_{\text{ref}})^2f_{\text{ref}} = 0.49C_{\text{eff}}V_{\text{ref}}^2f_{\text{ref}} \). If we pipeline the system further as shown in Fig. 2-13c, we can achieve a throughput higher than \( 2f_{\text{ref}} \) with the same supply voltage \( V_{\text{ref}} \), but much less than \( 4f_{\text{ref}} = \frac{1}{T/4} \) if either the delays of each stage are far from equal or the sum

47
of minimum setup time and clock-to-output delay of the pipeline register becomes a significant portion of the clock period. If we have a tight area constraint and a demanding throughput requirement to meet, “over-pipeline” can be the only option. However, this comes at the cost of higher dynamic power, as compared to a reasonably pipelined version. If we have a demanding throughput requirement but loose area constraint, it is better to “depipeline” to a more reasonably pipelined version where delays of each stage are almost equal and the register delay is a very small portion of clock period. With this version, we can parallelize the system as shown in Fig. 2-13d and relax the timing constraint on each copy to half the throughput. With half the throughput requirement, a lower supply voltage can be used where lower dynamic power is consumed.

The baseline design of multi-Vt AMO SCS with 500ps timing constraint is shown (in blue crosses) in Fig. 2-15. Instead of directly lowering supply voltage to lower the energy, we can also relax the timing constraint on the design and either relax the circuit sizing on critical paths or combine some pairs of pipelined stages into one stage. Relaxing the cycle time from 500ps to 1000ps, the design (in red crosses) is then synthesized with weaker buffers and fewer replicate registers. If we combine the pipelined stages or depipeline, we are aggressively reducing the number of registers in the design, and hence the dynamic energy. For target throughputs below 1Gsamples/s, the depipelined design (in green crosses) is most energy-efficient and the block diagrams of the depipelined design are shown in Fig. 2-14.

All SCS designs mentioned have roughly the same area and routing utilization. In cases where area is a very tight constraint and parallelism cannot be utilized to increase throughput, especially at throughputs above 1Gsamples/s, the aggressively pipelined design is more energy-efficient than the relaxed designs with increased supply voltage.

Since the area of the 1-way design is relatively small (0.41mm² at 54% utilization), we select a depipelined AMO SCS design that has the lowest energy, and implement it in a 45nm SOI technology. The design has a gate count of 227,374. With a combination of micro-architecture and backend optimization, this multi-Vt AMO
(a) A digital system with 4 functional blocks: F, G, H, K.

(b) A pipelined version of (a).

(c) A possibly “over-pipelined” version of (b).

(d) A parallelized version of (b).

Figure 2-13: A digital system with various amount of pipelining.
(a) The depipelined *getTheta* block, with three pipeline stages, relaxed from eight stages of our previous design in [22].

(b) The depipelined *getAlpha* block, with seven pipeline stages, relaxed from 15 stages in [22].

(c) The depipelined *getPhi* block, with two pipeline stages, relaxed from four stages in [22].

Figure 2-14: The hardware block diagram of the depipelined AMO SCS design.
Figure 2-15: Micro-architecture trade-offs for OPWL AMO SCS design.
SCS chip achieves a record energy-efficiency and throughput of 32pJ/sample at 0.6V supply voltage and 400Msamples/s.

2.3.4 Area Reduction for Parallelism

Given the above depipelined micro-architecture we have chosen, with a lower supply voltage, we can achieve a minimum-energy point, but at a lower throughput. From here, we have to parallelize the design aggressively to meet the target throughput [37], with area becoming the constraint. In our previous work [22] as well as in the SCS designs previously mentioned in Section 2.3.2, scan-register-based programmable LUTs take up most of the area. These shift registers do not have to be fast; they are read-only in real-time high-speed operations and only written off-line at a lower speed. They only have to hold table values, so they do not consume much dynamic power in real time and enable aggressive supply scaling. However, they do consume significant area and can potentially limit the energy and throughput achievable through parallelism in area constrained designs. If, however, we use SRAMs for programmable LUTs, we can save area significantly, but we are limited to supply scaling to around 0.7V which prevents the design from reaching the minimum-energy point (Table 2.5).

To investigate these trade-offs we implement an AMO SCS design with SRAM-based lookup tables with a gate count of 80,152. This design consumes an active area of 0.21mm$^2$ at a placed and routed density of 38%, 2× smaller than the area of the register-based AMO SCS but with 25% higher in minimum-energy. In Table 2.6, we compare the register-based AMO SCS and the SRAM-based AMO SCS design with state-of-the-art designs, at a throughput of 400Msamples/s.

Now in Table 2.7, at a throughput of 1.6Gsamples/s, we introduce an area area constraint such that all designs consume roughly the same area of $\sim 0.8mm^2$. The SRAM-based AMO SCS can be easily 4-way parallelized to achieve the same high throughput as [22] with a significant saving in area over other design, while consuming less energy/sample. For a relatively high throughput of 1.6Gsamples/s and phase resolution requirement of 12 bits, the SRAM-based AMO SCS consumes 63mW of power which is 40pJ/sample and is almost a 60% reduction from the 95pJ/sample
Table 2.5: Comparison of real-time read-only storage-cells: registers versus SRAM.

<table>
<thead>
<tr>
<th></th>
<th>scan-register-based</th>
<th>SRAM-based</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hardware</strong></td>
<td>registers, combinational multiplexers</td>
<td>Latches, tri-state buffers, row decoder,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>column decoder, sense-amplifiers</td>
</tr>
<tr>
<td><strong>Speed</strong></td>
<td>Very fast</td>
<td>Fast</td>
</tr>
<tr>
<td><strong>Area</strong> (45nm SOI CMOS Technology)</td>
<td>∼ 6um$^2$ per bit</td>
<td>∼ 0.7um$^2$ per bit</td>
</tr>
<tr>
<td><strong>Power</strong></td>
<td>More leakage</td>
<td>Less leakage</td>
</tr>
<tr>
<td><strong>Architecture flexibility</strong></td>
<td>Adding extra read ports, is straightforward</td>
<td>Fixed, one write port, and one or two read ports</td>
</tr>
<tr>
<td><strong>SUPPLY scaling</strong></td>
<td>Down to 0.6V and lower</td>
<td>Starts to fail at 0.7V</td>
</tr>
</tbody>
</table>

Table 2.6: Area and minimum-energy comparison with other works on digital AMO SCS, with no area constraint.

<table>
<thead>
<tr>
<th></th>
<th>This work (register-based)</th>
<th>This work (SRAM-based)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technology</strong></td>
<td>Scaled to 45nm CMOS</td>
<td>45nm CMOS</td>
</tr>
<tr>
<td><strong>Throughput</strong></td>
<td>Scaled to 400Msamples/s</td>
<td>400</td>
</tr>
<tr>
<td><strong>Phase Resolution (bit)</strong></td>
<td>Scaled to 12-bit</td>
<td>12</td>
</tr>
<tr>
<td><strong>Power (mW)</strong></td>
<td>43.2</td>
<td>23</td>
</tr>
<tr>
<td><strong>Energy (pJ/sample)</strong></td>
<td>106</td>
<td>58</td>
</tr>
<tr>
<td><strong>Area (mm$^2$)</strong></td>
<td>20.4</td>
<td>0.75</td>
</tr>
</tbody>
</table>
energy footprint of our previous SCS in [22] and almost a 30% reduction from the 55pJ/sample footprint of the register-based SCS. The supply scaling of these SCS designs is shown in Fig. 2-16. The traditional AMO SCS as reported in [29] does not scale well in area and is impossible to meet the high throughout with the area constraint.

Table 2.7: Energy comparison with other works on digital AMO SCS, with all designs, except [29], consuming roughly same area (∼ 0.8mm²).

<table>
<thead>
<tr>
<th></th>
<th>[29]</th>
<th>[22]</th>
<th>This work (register-based)</th>
<th>This work (SRAM-based)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>Scaled to 45nm CMOS</td>
<td>45nm CMOS</td>
<td>45nm CMOS</td>
<td>45nm CMOS</td>
</tr>
<tr>
<td>Throughput (Gsamples/s)</td>
<td>Scaled to 1.6Gsamples/s</td>
<td>1.6 (2-way parallel)</td>
<td>1.6 (4-way parallel)</td>
<td></td>
</tr>
<tr>
<td>Phase Resolution (bit)</td>
<td>Scaled to 12-bit</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Power (mW)</td>
<td>170</td>
<td>152</td>
<td>88</td>
<td>63</td>
</tr>
<tr>
<td>Energy (pJ/sample)</td>
<td>106</td>
<td>95</td>
<td>55</td>
<td>40</td>
</tr>
<tr>
<td>Area (mm²)</td>
<td>82</td>
<td>0.75</td>
<td>0.82</td>
<td>0.84</td>
</tr>
</tbody>
</table>

The energy efficiency gain of parallelism relies on the fact that the SRAM-based AMO SCS is very small in area. Being small in area, its routing and its clock tree are relatively simple, allowing easy deserialization and serialization with minimum cost. It is important to note that in general, deserialization is absorbed into the upsampling process with the shaping filter and consumes almost no power (Fig. 2-17c), while the power footprint of serialization running at full speed still requires some consideration. With modern silicon technology and area constraints, 2-way or 4-way parallel are still reasonably within reach, beyond which the integration cost on serialization would be overwhelming and correct retiming among more than four ways would be difficult.

For AMO SCS and other similar examples of sequential digital signal processing which involves no recursion or feedback, there are no control and routing interactions between the parallel copies. The only overhead is deserialization and serialization at interfaces. In such cases, the dynamic-only power for a fixed throughput \( f_{\text{ref}} \) as a
Figure 2-16: The throughput and energy with supply scaling for the AMO SCS chip in our first OPWL paper [22], multi-Vt AMO SCS (assumed 2-way parallel), and SRAM-based AMO SCS (assumed 4-way parallel).

Function of supply voltage $V$ and the degree of parallelism $N$ is modeled as follows [37]:

$$P(N, V) = NC_{ref}V^2f_{ref} + C_{interface}(N)V^2f_{ref}$$

(2.16)

where $C_{ref}$ is the capacitance of a single copy and $C_{interface}(N)$ is the overhead in deserialization and serialization which operates at full throughput and is generally increasing polynomially with the number of parallel copies $N$. The improvement of power efficiency due to parallelism is then expressed as:

$$P_{normalized}(N, V) = \frac{NC_{ref}V^2f_{ref}}{N} + C_{interface}(N)V^2f_{ref}$$

$$= \left( 1 + \frac{C_{interface}(N)}{C_{ref}} \right) \left( \frac{V}{V_{ref}} \right)^2.$$  

(2.17)

At very low supply voltages close to transistor threshold voltage, the number of parallel copies $N$ required to meet throughput usually increases too fast, and the interface overhead prevents any further energy efficiency gain by supply scaling. A more detailed analysis on the overhead of parallelism, with consideration of control
(a) Deserialization and serialization of a 2-way parallelized AMO.

(b) The FIR shaping filter $H(z)$ could be separated into 2-way $H_0(z^2) + zH_1(z^2)$, and the no-inter-symbol-interference property of the shaping filter requires $H_0(z^2) = 1$ with appropriate normalization.

(c) After having applied Noble Identity of Decimation on (b), the deserialization is absorbed.

Figure 2-17: Deserialization of 2-way parallel system is absorbed into the $2 \times$ upsampling and shaping filter and has minimum power impact, while serialization remains and consumes power.
and routing interaction among the parallel copies, has been studied in [37].

2.4 Architecture Overhead: AMO versus LINC

Let us now look at a comparison of overhead of AMO SCS versus the LINC SCS. As shown in Section 2.1.1, the AMO SCS design requires more operations, but AMO PAs enable higher average PA efficiency than LINC PAs [8, 9]. It is interesting to compare the overhead of the digital implementation of the AMO SCS versus the LINC SCS. Our previous AMO SCS design in [22], as well as the current AMO SCS design described in the previous sections, can be also digitally configured to perform the signal component separation task for LINC PA systems. However, in energy-constrained LINC PA systems, system designers might not be able to afford the energy overhead using an AMO SCS instead of a customized LINC SCS. It turns out that we do pay a price for the higher efficiency in the case of AMO PA because it needs a more complex SCS. As shown in Fig. 2-3 and Table 2.1, a customized LINC SCS does not require amplitude selection; as amplitudes of the two component vectors are equal and fixed in the LINC system, its outphasing angle computation in Equation (linc2) is much simpler than its AMO counterpart in Equation (amo2).

To quantify the architecture overhead, we implement a register-based multi-Vt LINC SCS design using the described depipelined micro-architecture and backend optimization with a gate count of 201,313. The design consumes an active area of 0.38mm$^2$ at a placed and routed density of 51%. This multi-Vt LINC SCS chip achieves a record energy-efficiency and throughput for SCS of 22pJ/sample at 0.6V supply voltage and 400Msamples/s. Table 2.8 shows that this LINC SCS achieves $10\times$ power and $15\times$ area savings compared to the design in [28] adjusted for technology and throughput.

As the programmable $getPhi$ blocks are the same and dominate the area of both SCS, the register-based multi-Vt AMO SCS consumes only 7% more area than LINC SCS. Overall, the AMO SCS consumes 40% more energy than LINC SCS. The energy saving of LINC SCS over AMO SCS is mostly due to the smaller dynamic power of its
Table 2.8: Area and energy comparison with other works on digital LINC SCS.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>90nm CMOS</td>
<td>90nm CMOS</td>
<td>Scaled to 45nm CMOS</td>
<td>45nm CMOS</td>
</tr>
<tr>
<td>Throughput (Msamples/s)</td>
<td>50</td>
<td>50</td>
<td>Scaled to 400Msamples/s</td>
<td>400</td>
</tr>
<tr>
<td>Phase Resolution (bit)</td>
<td>8</td>
<td>Scaled to 12-bit</td>
<td>Scaled to 12-bit</td>
<td>12</td>
</tr>
<tr>
<td>Power (mW)</td>
<td>0.95</td>
<td>22.8</td>
<td>91.2</td>
<td>8.9</td>
</tr>
<tr>
<td>Energy (pJ/sample)</td>
<td>19</td>
<td>456</td>
<td>228</td>
<td>22</td>
</tr>
<tr>
<td>Area (mm²)</td>
<td>0.06</td>
<td>1.44</td>
<td>5.76</td>
<td>0.38</td>
</tr>
</tbody>
</table>

getAlpha block, as shown in Fig. 2-18. Measured design energy with supply voltage scaling of implemented AMO SCS and LINC SCS are shown in Fig. 2-19. AMO SCS is more power hungry than LINC SCS; however, for transmitters with high PAPR modulation schemes and reasonably large output power, the overall system will be more energy-efficient using the AMO PAs instead of LINC PAs. For example, with an energy efficiency of 32pJ/sample versus 22pJ/sample, a 2-way parallelized AMO SCS operating at 800Msamples/s will consume 26mW while a LINC SCS will consume 18mW. For most PAs with reasonably large output power, the PA efficiency improvement by adopting AMO over LINC would be large enough to justify the 8mW overhead in digital baseband power.

2.5 Impact on Overall System Efficiency by the Energy Footprint of Digital Baseband

Wrapping things up, we feel it is important to quantify the digital baseband efficiency in the context of overall transmitter efficiency. With an energy efficiency of 32pJ/sample, a 2-way parallelized AMO SCS operating at 800Msamples/s will consume 26mW. In Fig. 2-20, we plot the raw RF PAE versus Psat in blue crosses for PAs above 10GHz on silicon-based technologies, from IEEE references in year 2006.
Figure 2-18: Power breakdown of the AMO SCS design and the LINC SCS design.

Figure 2-19: The measured throughput and energy with supply scaling for multi-Vt depipelined AMO SCS, LINC SCS chip.
to year 2012. The black contour lines list the PAs where the overall system efficiency will be degraded by 0.1% or 1% respectively by the 26mW power of our AMO power. For example, for any PAs with output power larger than 100mW and less than 20% power efficiency, the overall power efficiency of the system will degrade by less than 1% by including the SCS power. To visualize the efficiency penalty due to the energy footprint of digital baseband, the power efficiency of the system after having included 26mW power of our AMO SCS is plotted in green; for comparison, the power efficiency of the system after having included 86mW of the traditional AMO SCS [29] is plotted in red.

![Figure 2-20: Power efficiency degradation by a 800Msamples/s-throughput AMO SCS of the improved 32pJ/sample energy efficiency.](image)

To enable a new class of efficient wide-band communication systems based on outphasing PAs, in this chapter we have explored ways to implement high-throughput outphasing baseband functions with the smallest energy and area footprints. By utilizing the changes in micro-architecture and aggressive back-end power optimization techniques, we achieve a record energy-efficiency and throughput for AMO SCS
of 32pJ/sample at 0.6V supply voltage and 400Msamples/s, with area of 0.41mm$^2$. A LINC SCS chip is also designed with record energy-efficiency and throughput of 22pJ/sample at 0.6V supply voltage and 400Msamples/s, with area of 0.38mm$^2$, with comparison showing the AMO SCS consumes 40% more energy than LINC SCS. However, for systems with high PAPR modulation and reasonably large output power, it is still more energy-efficient to use AMO PAs instead of LINC PAs of comparable peak power. For high-throughput area-constrained applications, our SRAM-based AMO SCS design achieves $2 \times$ area reduction over register-based design at same throughput, but with slightly higher minimum energy.

SCS is one of the many power hungry tasks to perform in the digital baseband of the transmitter. We have shown hereto that our approach is efficient in energy and area. This approach of digital signal processing is applicable to many different existent applications as well as to new emerging applications.

Another power hungry task is digital compensation to achieve better system linearity. Once compensation is proposed and optimized at the architecture level, we can then apply all aforementioned techniques to reduce the area and energy footprint of our proposed digital predistorter design. In the following chapter, we will show how our digital predistorter improves linearity performance and overall energy-efficiency of transmitters.
Chapter 3

Efficient Digital Compensation of Transmitter Nonlinearity

3.1 Overview of Digital Compensation

Efficiency of a power amplifier (PA) improves as its input and output power push higher, but the linearity of the system gets worse as the PA enters the nonlinear operation realm. At high output power, there are two main linearization approaches to meet the spectral requirements. The common approach is input power back-off (IBO). This method decreases the input power of the amplifier with a large sacrifice in efficiency and uses multiple amplifiers in parallel. A high energy-efficiency power combiner is achievable, if its inputs are all in-phase up to small mismatches.

Because we are striving for power efficiency, we will be focusing on the other approach: digital predistortion (DPD). This method uses predistorted input signals and will be the focus of this chapter. The input power level is the same as without any input power back-off, so there is no sacrifice in efficiency if DPD consumes no additional power. Our emphasis is on how to design the predistorter to consume as little power as possible while performing the required compensation task. In high-throughput communication systems with use of more complex modulation for enhanced spectral efficiency, the requirement on the system linearity becomes more stringent. With system modeling techniques, the nonlinearity in the analog system can be compen-
sated for with DPD as aforementioned or a predistorter system in baseband, usually implemented with a digital system. In this work, we implemented an energy- and area-efficient digital predistorter for the outphasing PA. When the spectral requirements are very stringent, a combination of power back-off and predistortion approaches may be necessary.

3.1.1 Linearity Metrics

To characterize the linearity of the PA system, we look at two major metrics: error vector magnitude (EVM) and adjacent-channel-power-ratio (ACPR). The EVM measures the ratio of root-mean-square (rms) error $Error_{\text{rms}}$ of the received constellation versus the maximal magnitude of the ideal constellation $S_{\text{max}}$ as:

$$\text{EVM} = \frac{Error_{\text{rms}}}{S_{\text{max}}} \times 100\%. \quad (3.1)$$

Fig. 3-1 illustrates the physical meaning of the EVM definition. The ACPR characterizes the spectral regrowth through a nonlinear communication system. The nonlinearity in the system causes spurious spectrum emissions to adjacent channels, and the ACPR measures this interference as the ratio of the average power in the adjacent channels versus in the main channel as:

$$\text{ACPR}_{\text{dB}} = 10 \log_{10} \frac{\text{Average Power}_{\text{adjacent channels}}}{\text{Average Power}_{\text{main channel}}} \cdot \quad (3.2)$$

Fig. 3-2 shows the physical meaning of the ACPR definition, where the main and adjacent channels are defined by the particular communication standard. We use these two metrics to evaluate the linearity performance of the system before and after digital predistortion.

To illustrate the tradeoff between linearity and power efficiency clearly, we look at Fig. 3-3 where we show a typical relation curve between the ACPR of a single PA and its power efficiency. Before doing any linearization, as power efficiency increases (gets more positive towards 100%) with increasing input power, the ACPR degrades
Figure 3-1: EVM is the rms magnitude of error vectors, with the peak symbol amplitude normalized to be 1.

Figure 3-2: An illustration of the ACPR definition.
(gets less negative). The IBO approach is to move the operating point of the single amplifier along the ACPR-efficiency curve towards better ACPR (more negative value) and worse efficiency (smaller towards 0%). Once the spectral requirement is met at the chosen amplifier operating point, multiple instances of amplifiers and an N-way power combiner are required to achieve the desired output power. The DPD approach does not change the PA operating point, but instead shifts the ACPR-efficiency curve downwards with properly predistorted amplifier inputs. A better ACPR can be achieved with almost no degradation in power efficiency if the power of the digital predistorter is much smaller than the total power consumption of the PA. This second approach is particularly attractive when a very large output power, for long range or for channels with high loss, is required for the PA.

![Figure 3-3: Tradeoff between efficiency and ACPR of a single PA.](image)

### 3.1.2 Popular Digital Compensation Techniques

A digital predistorter acts as a right inverse of the nonlinear baseband-equivalent system, rather than a postdistorter which acts as a left inverse of the system. When
the inverse and the system are concatenated together, the nonlinearities cancel out, such that the output is a desired linearly amplified version of the input (Fig. 3-4).

Figure 3-4: When the inverse and the system are concatenated together, the output appears as a linear amplification of the input.

For systems with narrow bandwidth, memoryless nonlinearity can often be compensated for with static predistortion [38–40]. LUTs on symbols or on samples, are often enough to realize the static predistorter in hardware. With knowledge of the PA characteristics, the performance target, and the input signal statistics, [41] attempted to optimize the wordlength of the lookup table. References [42–44] optimized nonuniform spacing of LUT entries to reduce LUT size, but their results were limited to either simulation on memoryless PA models or measurements on real PAs with bandwidth less than 4MHz.

At the other end of the spectrum, for wideband communication systems, memory effects dominate in the nonlinear baseband-equivalent system model, so advanced system modeling techniques are needed [45–47]. To model the PA itself and its inverse, past works in PA predistortion have applied common nonlinear dynamical system structures, such as Volterra series [48], Wiener, Hammerstein, Wiener-Hammerstein, and Hammerstein-Wiener structures (Fig. 3-5). There is also a Memory Polynomial (MP) model with reduced complexity compared to Volterra series, which requires significantly less hardware resources for real-time implementation, but is less versatile [49,50].

However, despite wide interest in digital predistortion from both the circuits design community and the system modeling community, there are few works with real
hardware implementation. In the old realm of PAs with very low symbol rate, DPD did not require very high throughput, and because the bandwidth of the PA was much wider compared to signal bandwidth, the complexity of nonidealities was low. Hence, DPD had a very small power footprint compared to watt-level PAs, and its implementations were always reported with results in MATLAB/Simulink. Nowadays, as we move into the realm of high throughput and high precision, the power footprint of DPD will grow significantly. It is essential to consider the power of DPD when optimizing the overall system power efficiency.

In our work, we compensate for an outphasing PA at carrier frequency of 45GHz. At the time of this writing, there is no other literature on DPD for outphasing 45GHz power PAs yet for direct comparison. The closest two areas for comparisons are outphasing PAs at any carrier frequency, and Q-band (33GHz to 50GHz) PAs with any architecture. In Table 3.1, we see the most recent works on outphasing transmitters with DPD. Those works provide good results in improving ACPR of outphasing transmitters [51–54]. However, those works are limited to low data rate applications,
less than 100Msymbols/s symbol rate and less than 200Mbps data rate; and there is no hardware implementation of DPD reported. In Table 3.2, we see the most recent works on Q-band transmitters with DPD. Some of those works provide good results in improving ACPR or improving EVM. However, those works are also either limited to low data rate applications, less than 20Msymbols/s symbol rate and less than 100Mbps data rate [55,56]; or they achieve high data rate around 1Gbps, but do not report significant improvement of either ACPR or EVM [57,58]; of all these works, there is no hardware implementation of DPD reported. In comparison, in this work, we report the hardware implementation of DPD for an outphasing 45GHz PA with more than 500Msymbols/s symbol rate and 3Gbps data rate and report significant improvements in ACPR and EVM metrics.

Table 3.1: Recent literature on outphasing transmitters with DPD.

<table>
<thead>
<tr>
<th>PA carrier class (GHz)</th>
<th>symbol rate (Msymbols/s)</th>
<th>modulation scheme</th>
<th>data rate (Mbps)</th>
<th>ACPR without/with DPD (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[51] class-F 2.14</td>
<td>10</td>
<td>64QAM</td>
<td>60</td>
<td>-22.3 / -42.54</td>
</tr>
<tr>
<td>[52] class-E 2.14</td>
<td>20</td>
<td>64QAM</td>
<td>120</td>
<td>-26 / -49.5</td>
</tr>
<tr>
<td>[53] class-D 2.4</td>
<td>20</td>
<td>64QAM</td>
<td>120</td>
<td>-33 / -50</td>
</tr>
<tr>
<td>[54] class-E 9.8</td>
<td>78</td>
<td>$\frac{\pi}{4}$DQPSK</td>
<td>156</td>
<td>-30 / -45</td>
</tr>
<tr>
<td>This class-E 45</td>
<td>550</td>
<td>64QAM</td>
<td>3300</td>
<td>-30.6 / -44.0</td>
</tr>
</tbody>
</table>

3.1.3 Nonidealities in Outphasing Transmitters

Due to our limited access to a functional testing system at Q-band (45GHz) and W-band (93GHz), we have been confined in our efforts of outphasing transmitters characterization to simulations. Fig. 3-6 shows the simulation setup for the outphasing system under compensation, similar to the setup in [24]. We use this framework to both investigate the overall system nonlinearity, as well as test our nonlinear predistorter. As shown in Fig. 3-6, random symbols drawn from the 64QAM constellation first pass through the shaping filter, which operates at a higher sampling rate to achieve a shaped spectrum. Then the SCS decomposes the shaped samples
Table 3.2: Recent literature on 45GHz-carrier transmitters with DPD. ACPR numbers are not reported (NR) in some literatures.

<table>
<thead>
<tr>
<th>PA type/class</th>
<th>carrier (GHz)</th>
<th>symbol rate (Msymbols/s)</th>
<th>data rate (Mbps)</th>
<th>EVM without / with DPD (%)</th>
<th>ACPR without / with DPD (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[55]</td>
<td>single Q-Band</td>
<td>unspecified</td>
<td>16</td>
<td>64</td>
<td>-15.5/-35.6</td>
</tr>
<tr>
<td>[56]</td>
<td>single 40-45</td>
<td>unspecified</td>
<td>8</td>
<td>48</td>
<td>8.2/4.9/5.7/4.2</td>
</tr>
<tr>
<td>[57]</td>
<td>single 42-48</td>
<td>class-B</td>
<td>625</td>
<td>1250</td>
<td>NR/5.5</td>
</tr>
<tr>
<td>[58]</td>
<td>single 45</td>
<td>class-AB</td>
<td>98.2</td>
<td>982</td>
<td>6.3/1.26</td>
</tr>
<tr>
<td>This</td>
<td>outphasing</td>
<td>class-E</td>
<td>550</td>
<td>3300</td>
<td>4.5/1.0</td>
</tr>
</tbody>
</table>

and produces the two phase signals for LINC system. Phase commands then pass to the simulator as the inputs to the PA system. The PA system consists of two phase modulators; two switching PAs; and a power combiner, which produces the final transmitted signal. To obtain the received samples with high precision, we use the ideal demodulation method, also explained in [24] to demodulate the transmitted signal.

In our setup, the phase modulator, the PAs, and the power combiner are simulated with Cadence Spectre Accelerated Parallel Simulator. All other blocks are processed in MATLAB. The two systems are simulated at the carrier frequencies of 45GHz and 93GHz, both with 550Msymbols/sec bandwidth and 2× symbol oversampling rate.

For optimization of any system, it is important to first understand which metrics to optimize at which abstraction level, in order to achieve the best tradeoff for a given set of requirements. For an outphasing transmitter, it is worthwhile to understand where the nonlinearities come from and decide whether they should be compensated at the analog circuit level or at the system level.

In order to quantify the individual contribution of each building block to the ACPR degradation, we set up the simulation of the 45GHz outphasing PA as in
Figure 3-6: Illustration of the transmitter system under compensation, with blocks highlighted in the red box simulated in Spectre and other blocks in MATLAB.

Fig. 3-7. For the preamplifiers (preamp) and PAs, Spectre simulations of circuit schematics were used. For the power combiner, we have three choices: a model of zero-degree combiner originally designed and implemented to minimize insertion loss [59], a model of the Wilkinson combiner, or a model of an ideal combiner with perfect isolation between inputs. For the phase modulation, we have two choices: the SPICE-level accurate model of the actual implementation or a Verilog-A model of an ideal phase modulator. Ideal power supply network has been assumed in this study. More relevant to AMO than to LINC, bump inductance of the switch power supply network contributes to ACPR degradation significantly, and its effect has been well studied in [24]. To find the dominating contributor of ACPR degradation, we run through the permutations of the available choices for each analog part in an outphasing transmitter system, as shown in Table 3.3, and simulate the system performance. The simulated ACPR numbers with different permutations of SPICE-level and ideal models are shown in Fig. 3-8. We can see that nonlinearity of the phase modulator dominates over poor isolation of the zero-degree combiner and the saturation of the inverter-style preamplifiers.
Figure 3-7: Simulation setup to quantify the ACPR degradation.

Table 3.3: Choices of modeling, with different level of complexities, for each analog part in an outphasing transmitter.

<table>
<thead>
<tr>
<th>model of phase modulator</th>
<th>model of preamp and PA</th>
<th>model of power supply switch network</th>
<th>model of power combiner</th>
</tr>
</thead>
<tbody>
<tr>
<td>→ Verilog-A</td>
<td>→ spectre</td>
<td>→ ideal power supply</td>
<td>→ ideal combiner</td>
</tr>
<tr>
<td>→ spectre</td>
<td></td>
<td>→ 5pH bump</td>
<td>→ Wilkinson</td>
</tr>
<tr>
<td></td>
<td></td>
<td>→ 20pH bump</td>
<td>→ zero-degree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>→ 60pH bump</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3-8: The simulated ACPR numbers along the signal chain with different permutations of SPICE-level and ideal models. Note that the ACPR numbers with Wilkinson combiner and ideal combiner are similar.
With the setup described above, we look at the major contributing factors of degraded EVM and ACPR in the LINC system. We consider mixer characteristics, saturation of the inverter-style buffer chain, and limited isolation between the two inputs of the combiner. We explore these causes of nonlinearity and make efforts to address them at the most appropriate design level, whether it is at the analog circuit level or at the system level. Path delay mismatch between the two outphasing paths, delay mismatch between the phases and the amplitudes, and amplitude switching in the AMO system are not discussed here, as they are thoroughly considered and handled in the literature [24, 29, 60].

The first component in the analog front end, also the first source of nonlinearity, is the mixer. The mixer circuitry of the phase modulator upconverts a digital baseband signal into the carrier frequency range. The mathematical function of the mixer is to multiply the baseband signal and carrier signal. Isolation between the carrier input port and the baseband input port is limited, and there is also a certain amount of feed-through directly from the carrier input port to the mixer output port. The reverse feed-through from the mixer output port back to the carrier input port also causes undesired self-mixing. All of these effects contribute to nonlinearity in phase modulation. Unfortunately, addressing these at the analog circuit level is extremely difficult. Digital compensation at the system level is necessary to compensate for the nonlinear phase modulation.

In the signal chain, between the mixer and PA are the preamplifiers. The preamplifiers, typically implemented as inverters driven into saturation to give rail-to-rail outputs, are used in the signal path to achieve power gain. This contributes significant nonlinearity to the system. To mitigate the effect at the circuit level, a resistor is added across the input and the output of each inverter to make it less saturated at the expense of additional power consumption. Beyond this modification, other linearization techniques at the analog circuit level would trade off too much power for gained linearity.

At the end of the signal chain, after the PA is the power combiner. Power efficiency has long been the top priority of the RF PA designer. To improve power efficiency
in a straightforward manner, designers have focused on reducing insertion loss of the power combiner, so wherever applicable, a zero-degree combiner is used [59]. The zero-degree combiner is compared with the Wilkinson combiner in Table 3.4. The Wilkinson combiner provides no impedance transformation and its loss is higher, but it is suitable for outphasing transmitters with high linearity requirement because of its good isolation among input ports. Intuitively, the zero-degree combiner uses the shortest transmission line to connect the inputs to the output, minimizing insertion loss and saving silicon area. The zero-degree combiner is the best choice for the RF PA designer, if power efficiency and area are the only requirements. However, the zero-degree combiner provides almost no port isolation between inputs. The poor isolation between the inputs of the combiner allows for strong interaction between two separate paths in a LINC system. This strong interaction between two paths contributes significant nonlinearity to the system, and is extremely difficult to address at the system level. To optimize for performance of the outphasing system requiring a combined single two-output compensator, a power combiner with good isolation, such as a Wilkinson combiner, has to be implemented instead. Our approach trades off some power efficiency for linearity and better isolation between the two outphasing paths.

<table>
<thead>
<tr>
<th>Table 3.4: Comparison of Wilkinson combiner versus zero-degree combiner.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilkinson combiner</td>
</tr>
<tr>
<td>Additional functionality</td>
</tr>
<tr>
<td>Area</td>
</tr>
<tr>
<td>Loss</td>
</tr>
<tr>
<td>Isolation</td>
</tr>
</tbody>
</table>

Any remaining nonlinearities, unaddressed at the circuit level, have to be addressed at the system level by the digital predistorter.


3.2 Digital Compensator for Outphasing Transmitters

The iterative off-line compensation method has been studied in [24] and applied to the 45GHz LINC transmitter in simulation. Using the iterative method, we identify the system with sequence-based compensation. In [24], the \( PM1 \) and \( PM2 \) blocks (Fig. 3-6) are ideal Verilog-A models instead of spectre models to speed up simulation. It has been shown that the Frobenius norm of the Jacobian of the LINC outphasing function approaches infinity when the amplitude of input sample comes close to zero. Therefore, avoiding the region close to zero should help speed up the iteration of off-line sequence-based compensation to converge [24].

Table 3.5 shows that for an input sequence without zero-avoidance property the iterations do not converge and compensation does not improve ACPR significantly. The comparison results are again depicted in Fig. 3-9a and 3-9b showing the EVM performance before and after the off-line compensation, both compared with the ideal 64QAM constellation. The two EVM figures correspond to an improved EVM performance from 4.5% to 1.0% [24]. Fig. 3-10 shows the ACPR performance improvement before and after off-line compensation, from -30.6dB to -44.0dB [24].

Table 3.5: ACPR and EVM performance comparisons between using input sequence with and without zero-avoidance property for 45GHz LINC system [24].

<table>
<thead>
<tr>
<th></th>
<th>Zero-avoidance</th>
<th>No zero-avoidance</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACPR (dB)</td>
<td>-30.6 ( \rightarrow ) -44.0</td>
<td>-30.1 ( \rightarrow ) -39.6</td>
</tr>
<tr>
<td>EVM (%)</td>
<td>4.5 ( \rightarrow ) 1.0</td>
<td>4.2 ( \rightarrow ) 1.7</td>
</tr>
</tbody>
</table>

As it has been established that a zero-avoidance shaping filter is necessary to facilitate iterations, the setup of the system under compensation is shown in Fig. 3-11. In Section 3.2.1, we explore how to implement a zero-avoidance shaping filter in hardware in an energy-efficient way. Then we describe the implementation of the digital predistorter for outphasing transmitters in Section 3.2.2 and limitations of this linearization methodology in 3.2.3.
Figure 3-9: EVM of the Q-band (45GHz) LINC transmitter under study, uncompensated versus off-line compensated with zero-avoidance input sequence. Ideal symbols are in red and the output symbols are in blue.

Figure 3-10: Uncompensated output versus compensated output ACPR of the LINC system, with real-time zero-avoidance input sequence.
3.2.1 Design and Implementation of a Zero-Avoidance Shaping Filter for Outphasing Transmitters

In our work, we apply zero-avoidance to the input sequence and use predistortion for the 45GHz PA and the 93GHz PA in the LINC mode. Fig. 3-12 shows the zero-avoidance zone where the shaped samples have absolute value smaller than a positive threshold denoted as \( R_z \). The zero-avoidance algorithm seeks to move the samples in the zone to the outside.

The zero-avoidance shaping filter replaces the common shaping filter in the transmitter. From a sequence of symbols, it upsamples and generates a sequence of samples with absolute value above a certain positive threshold, and still shapes the spectrum with no inter-symbol-interference (ISI). The heuristics of the design, illustrated in Fig. 3-13, was first outlined in [24]. The real-time on-line zero-avoidance shaping filter starts with a spectrum-shaped sequence of samples, then identifies those samples with absolute value below the positive threshold \( R_z \). The preliminary corrections are
simply the identified samples properly scaled:

$$V_c = R_z \frac{V}{A} - V = \left( R_z \frac{1}{A} - 1 \right) V$$

(3.3)

where $A = |V|$ is the amplitude of the sample $V$. Then the preliminary corrections go through a correcting filter $h_c[n]$. When added back, the new sample sequence has no samples with absolute value below the positive threshold $R_z$ and maintains a shaped spectrum. The correcting filter $h_c[n]$ is designed with the same spectral requirement and the no-ISI constraint as the original shaping filter $h_s[n]$, so the resulting sequence has no ISI and meets the spectral mask requirement. The functional diagram of this algorithm is shown in Fig. 3-14.

To maintain good ACPR after correction and minimize the average distance of the samples remaining inside the circle of radius $R_z$, we optimize over $m_s = 4N_s - 1$, the number of taps in the shaping filter $h_s[n]$ and $m_c = 4N_c + 1$, the number of taps in the correcting filter with $h_c[n]$, with $N_s$ and $N_c$ being positive integers. Ideally, the larger the $m_c$, the longer the correcting filter $h_c[n]$ is and the better we can maintain good ACPR after correction. The large the $m_s$, the longer the shaping filter $h_s[n]$ is and the better we are able to optimize the shaping filter and reduce the maximum
Figure 3-13: Illustration of the zero-avoidance shaping filter algorithm [24].

Figure 3-14: The functional diagram of a zero-avoidance shaping filter.
amplitude of the interpolating sample during oversampling. Aware of the no-ISI and symmetric properties of these filters and having normalized the center tap value, the filters have only \( N_s \) and \( N_c \) independent tap values where:

\[
H_s(z) = \sum_{n=-(2N_s-1)}^{2N_s-1} h[n] z^{-n} \\
= 1 + \sum_{n=1}^{N_s} h[2n-1] (z^{-2n+1} + z^{-2n-1})
\]  
\[ (3.4) \]

\[
H_c(z) = \sum_{n=-2N_c}^{2N_c} h[n] z^{-n} \\
= 1 + \sum_{n=1}^{N_c} h[2n] (z^{-2n} + z^{2n})
\]  
\[ (3.5) \]

Numerically, we see diminishing returns on zero-avoidance performance as we increase \( N_s = (m_s + 1)/4 \) and \( N_c = (m_c - 1)/4 \) beyond 15, as shown in Fig. 3-15.

In the hardware implementation of this zero-avoidance algorithm, to have reasonable balance between performance and complexity, we choose \( m_s \) to be 59 and \( m_c \) to be 61 (Fig. 3-16). The optimized coefficients of \( h_s[n] \) and \( h_c[n] \) at \( m_s = 59 \) and \( m_c = 61 \) are shown in Fig. 3-17, and the effectiveness of this realized zero-avoidance shaping filter in moving samples outside the threshold radius is shown in Fig. 3-18. Similarly, because the shaping filter \( h_s[n] \) is no-ISI and symmetric, at an oversampling factor of 2, only \( (m_s + 1)/4 = 15 \) multiplications for the shaping filter rather than \( m_s = 59 \) multiplications need to be performed (Fig. 3-19). Because the correcting filter \( h_c[n] \) is also no-ISI and symmetric, at an oversampling factor of 2, only \( (m_c - 1)/4 = 15 \) multiplications rather than \( m_c = 61 \) multiplications need to be performed (Fig. 3-20). Although the non-ISI and symmetric property of these filters are leveraged to reduce power consumption, the filters consume a significant amount of power even with very small area because the clocking power of registers has not been reduced. Using ultra-high threshold voltage standard cells for the scan chain of coefficient values, we reduce the leakage power. At the high fanout node, we deliberately insert replicate registers to improve the performance on timing at a cost of slight increased power consumption.

The described zero-avoidance shaping filter is implemented in the nlcom6 chip, fabricated with 45nm SOI process. The filter has a gate count of 167,716 and standard
Figure 3-15: The ACPR using the zero-avoidance shaping filter, the maximum amplitude of shaped samples, and the average amplitude as percentages of $R_z$ of samples with amplitude below $R_z$. 
Figure 3-16: The hardware implementation of the zero-avoidance shaping filter.

Figure 3-17: Optimized coefficients of $h_s[n]$ and $h_c[n]$ at $m_s = 59$ and $m_c = 61$. 
cell area of 160,606 \text{\,um}^2. Though its gate count is only 2× that of the SRAM-based AMO block, its post-place and route estimated energy consumption of 224pJ/sample is 5× that of the SRAM-based AMO block. The energy footprint of the zero-avoidance shaping filter is not negligible.

### 3.2.2 Implementation of Digital Compensator for Outphasing Transmitters

To verify the model for a real world system, we implement the digital baseband containing the zero-avoidance shaping filter described above, the AMO signal component separator, and the predistorter, and then integrate them with a Q-band (45GHz) and a W-band (93GHz) analog front-end including digital-analog interface, phase modulator, AMO and LINC PAs, and power supply switches as an overall integrated system solution in 45nm SOI technology for a dual-band mm-wave transmitter. The hardware implementation of the predistorter provides the functionality of the model we tested with the simulation data. Additionally, to prepare for circumstances that are different from what Spectre simulations predict, we attempt to make the hardware as flexible as possible, not limiting it to the simulated situations.
(a) The 7-tap shaping filter with the center tap normalized $h_s[0] = 1$.

(b) Apply the no-ISI property ($h_s[n] = 0$ for all nonzero even $n$) on (a).

(c) Apply the symmetric property ($h_s[-n] = h_s[n]$) on (b).

Figure 3-19: An $m_s$-tap shaping filter requires only $(m_s + 1)/2$ adders and $(m_s + 1)/4$ multipliers, because of its no-ISI and symmetric properties and because the center tap is normalized $h_s[0] = 1$, as illustrated above with $m_s = 7$. 
(a) The 9-tap correcting filter with the center tap normalized $h_c[0] = 1$.

(b) Apply the no-ISI property ($h_c[n] = 0$ for all odd $n$) on (a).

(c) Apply the symmetric property ($h_c[-n] = h_c[n]$) on (b).

Figure 3-20: An $m_c$-tap correcting filter requires only $(m_c-1)/2$ adders and $(m_c-1)/4$ multipliers, because of its no-ISI and symmetric properties and that the center tap is normalized $h_c[0] = 1$, as illustrated above with $m_c = 9$. 

86
Figure 3-21: The block diagram of the predistorter hardware implementation, configured in the LINC mode.
The predistorter design we use on the *nlcom6* chip is shown in Fig. 3-21. The design assumes that the two outphasing paths are independent and non-interacting because we are using a Wilkinson combiner instead of a zero-degree combiner.

Our predistorter consists of two major parts, corresponding to the nonlinear transformation and LTI system structure as proposed in [24]. The predistorter has two configuration modes: compensate for a LINC system and compensate for an AMO system. When programmed to compensate for a LINC system, the predistorter takes the outphasing angles $\phi_1, \phi_2$ as the inputs and produces the correction signals $\Delta \phi_1$, and $\Delta \phi_2$ which are added to $\phi_1, \phi_2$ respectively and passed to the rest of the digital baseband system. The nonlinear transformation takes signals $\phi_1, \phi_2$, and their delayed versions as inputs, using two complex-valued nonlinear functions and produces two complex outputs, or effectively four real outputs for each PA. The nonlinear functions are implemented with piece-wise quadratic approximations in four dimensions (current phase, 1-sample-delayed phase, 2-sample-delayed phase, and 3-sample-delayed phase) as a straightforward extension from the one dimensional piece-wise linear algorithm used in the SCS implementation. When programmed to compensate for an AMO system, the nonlinear transformation part takes signals $a_1, a_2, \phi_1, \phi_2$, and their delayed versions as inputs, and the nonlinear functions are implemented with piece-wise quadratic approximation in three dimensions (current phase, 1-sample-delayed phase, and 2-sample-delayed phase). The address length of LUT for piece-wise quadratic approximation is 13 bits; different configuration modes of the LUT are shown in Table 3.6.

The second part of the predistorter structure is an LTI system with discontinuities at $\pm \pi$. With this special fixed LTI system with long memory to handle effects of zero-order hold and ideal bandpass filter characteristics in the modulation and demodulation processes, the nonlinear transformation requires only short memory with a small number of coefficients. As shown in Fig. 3-21, the LTI system takes the four real outputs from each PA and produces the correction signals $\Delta I$ and $\Delta Q$. We use 32 short 5-tap FIRs and four long 101-tap brick-wall low pass filters (LPFs). The last block translates the correction signals from Cartesian coordinates to polar

88
Table 3.6: LUT configuration of nonlinear transformation part of predistorter for LINC and AMO operation.

<table>
<thead>
<tr>
<th>configuration mode</th>
<th>address length of LUT in use (bit)</th>
<th>composition of LUT address</th>
</tr>
</thead>
<tbody>
<tr>
<td>LINC</td>
<td>12 ( = (3 \times 4) )</td>
<td>3 MSB of current phase,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 MSB of 1-sample-delayed phase,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 MSB of 2-sample-delayed phase,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 MSB of 3-sample-delayed phase</td>
</tr>
<tr>
<td>AMO</td>
<td>13 ( = (2 + 1 + 1) + (3 \times 3) )</td>
<td>2 bits of current amplitude,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 MSB of 1-sample-delayed amplitude,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 MSB of 2-sample-delayed amplitude;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 MSB of current phase,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 MSB of 1-sample-delayed phase,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 MSB of 2-sample-delayed phase</td>
</tr>
<tr>
<td>AMO Alternative</td>
<td>13 ( = (1 \times 7) + (3 \times 2) )</td>
<td>1 MSB of current amplitude,</td>
</tr>
<tr>
<td>(for amplitude-dominant long memory)</td>
<td></td>
<td>1 MSB of 1-sample-delayed amplitude,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 MSB of 2-sample-delayed amplitude,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 MSB of 3-sample-delayed amplitude,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 MSB of 4-sample-delayed amplitude,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 MSB of 5-sample-delayed amplitude,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 MSB of 6-sample-delayed amplitude;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 MSB of current phase,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 MSB of 1-sample-delayed phase</td>
</tr>
</tbody>
</table>
coordinates.

With the above predistorter structure, we implement the whole digital baseband system (Fig. 3-22) including the SCS functionality as well as the nonlinear compensation capability. The system accepts the input symbol from an on-chip 1k-entry SRAM table, which we can populate off-line, and the zero-avoidance shaping filter generates the shaped samples to the SCS. The SCS processes the computations and provides the phase signals along the way to the predistorter. The nonlinear predistorter outputs the corrected phases. Finally the getPhi block computes the output using the phase inputs.

The die photograph of this integrated transmitter chip nlcom6, realized in a 45nm SOI process, is shown in Fig. 3-23. The blocks are, from top to bottom, the 8-way W-band (93GHz) LINC PA block, the 93GHz phase modulator, the digital predistorter, the 45GHz phase modulator, and the 8-way Q-band (45GHz) AMO PA block. In order to leave enough room for design space exploration of the predistorter, all of the parameters in the predistorter are programmable and implemented with SRAMs. The chip fabricated in a 45nm SOI process measures 3mm×6mm. The digital baseband has a gate count of 3,977,014. The breakdown of gate area in the digital baseband blocks of nlcom6 is shown in Fig. 3-24. The majority of the area consumption is from the predistorter nonlinear transformation and FIR filtering. The static lookup table lookup1k block, for testing and support functions, is implemented as an SRAM with 1024 words and 28 bits per word; the getPhi block of the AMO SCS and the nonlinear transformation part cmpNL of the predistorter are realized by OPWL approximation, also with SRAMs. The nlcom6 digital baseband has a total of 800 kilobytes of memory with significant area saving over register-based memory, as shown in Table 3.7. With estimation from post-place-and-route analysis, Fig. 3-25 shows the power breakdown of the digital baseband, where the LTI part of the predistorter dominates the power consumption. With the zero-avoidance shaping filter in Section 3.2.1 and this implementation of the predistorter, in simulation, we are able to improve ACPR performance from -30.6dB to -44.0dB and reduce EVM from 4.5% down to 1.0%. The entire digital baseband, including the zero-avoidance shaping
Figure 3-22: The block diagram of the digital baseband with zero-avoidance shaping filter, SCS and nonlinear predistorter.
filter, the AMO SCS, and the digital predistorter, consumes 1.65W at a throughput of 1.1Gsamples/s (3.3Gbps with 64QAM modulation and oversampling ratio of 2). The energy efficiency is 1.5nJ/sample.

Figure 3-23: Die photograph of transmitter chip *nlcom6* with digital baseband including dynamic nonlinear predistorter with zero-avoidance shaping filter and integrated Q-band (45GHz) AMO PA and W-band (93GHz) LINC PA.

It is important to quantify the digital predistorter energy efficiency in the context of overall transmitter efficiency. With an energy efficiency of 1.5nJ/sample, an AMO predistorter at 800Msamples/s will consume 1.2W. In Fig. 3-26, we plot the raw RF
Table 3.7: Size of SRAM instances in \textit{nlcom6} digital baseband.

<table>
<thead>
<tr>
<th>Block</th>
<th>Size of SRAM Instance (bit)</th>
<th>Number of Instances</th>
<th>Memory Size (kB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>lookup1k</td>
<td>$2^{10} \times (13 \times 2 + 2)$</td>
<td>1</td>
<td>3.6</td>
</tr>
<tr>
<td>cmpNL</td>
<td>$2^{13} \times (9 + 7 \times 3 + 6 \times 6)$</td>
<td>8</td>
<td>540.7</td>
</tr>
<tr>
<td></td>
<td>$2^{13} \times (7 \times 1 + 6 \times 4)$</td>
<td>8</td>
<td>254.0</td>
</tr>
<tr>
<td>getPhi</td>
<td>$2^8 \times (9 + 10 + 9)$</td>
<td>2</td>
<td>1.8</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>800.0</td>
</tr>
</tbody>
</table>

Figure 3-24: Gate area breakdown of digital baseband of transmitter \textit{nlcom6}.
efficiency versus output power in blue crosses for PAs above 10GHz on either bulk CMOS, SOI CMOS, or SiGe technologies, from IEEE references in year 2006 to year 2012. The black contour indicate list the PAs where the overall system efficiency will be degraded by 1% or 2% respectively, by the 1.2W power of our predistorter. For example, for any PAs with output power larger than 1000mW and less than 14% power efficiency, the overall power efficiency of the system will degrade by less than 2% by including the 1.2W predistorter power. For a targeted outphasing PA with 4W output power at 16% PAE, the power efficiency of the system will degraded by 0.7%. To visualize the efficiency penalty due to the energy footprint of digital predistorter, the overall power efficiency of the system after having included 1.2W predistorter power is plotted in green circles. Note that the uncompensated PA operates at same input power and has worse ACPR than the PA with DPD. For same ACPR, the uncompensated PA would have to operate with IBO which leads to lower overall PAE than with our DPD.
IEEE References in 2006
- 2012, PAE efficiency,
>10GHz, >100mW

Efficiency degraded
by 1% with
1.2W compensator power

Efficiency degraded
by 2% with
1.2W compensator power

Targeted 4W output
power at 16% PAE

Power efficiency of
system with 1.2W compensator power

Figure 3-26: Power efficiency degradation by an 800Msamples/s-throughput AMO predistorter of the 1.5nJ/sample energy efficiency. Note that the uncompensated PA operates at same input power and has worse ACPR than the PA with DPD. For same ACPR, the uncompensated PA would have to operate with IBO which leads to lower overall PAE than with our DPD.
3.2.3 Limitations

The fundamental limitation of compensation is system instability. Given a certain degree of instability existing in the system, there is a maximum on the model quality we are able to achieve.

This compensation methodology applied to other systems has the limiting requirement that the nonlinearity of the system should be mild. In order for the sequence-based iterations to converge, the error function must have a Lipschitz constant less than 1. If this convergence criterion cannot be ensured in the system of interest, then the iterations fail and we will not be able to use iterations to find the off-line predistorter.

3.3 Digital Compensation of RF Power Amplifiers

In this section, we use the ADL5606 power amplifier evaluation board [61] as a case study to demonstrate the effectiveness of our approach to improve linearity of a radio-frequency (RF) transmitter.

3.3.1 Experimental Setup

Before we characterize and compensate for the nonlinearities of the PA system, we need to quantify the nonlinearity of our measurement setup, and make sure that the nonlinearity of our instruments is sufficiently smaller than the nonlinearity of the system under test.

To characterize the nonlinearity of the oscilloscope, a clean sinusoid with minimum harmonic distortion is required. We generate a sinusoid at 1GHz fundamental with an E4438C Vector Signal Generator. To reduce the second harmonic distortion of this input signal, we use a VLFX-825 coaxial low pass filter cascaded with a VLFX-950 coaxial low pass filter to filter out the second harmonic (at 2GHz) and higher harmonics. We use a 11667B power splitter to split the signal into the DSA90804A oscilloscope and the N90304A Signal Analyzer, so that we can observe the signal in
time and frequency domains (Fig. 3-27a). Examining the harmonic content of this test signal on the signal analyzer confirms that the second and higher harmonics have been filtered down to be below -80dBc (Fig. 3-27b). The signal-to-noise ratio (SNR) of the DSA90804A oscilloscope, at 4000mV full-scale, is specified to be only 50dB. However, we can reduce the noise floor by averaging the repetitively captured signals; improving SNR by 3dB for every two averages. With 1024 averages, the SNR can be sufficiently improved by 30dB from 50dB to 80dB; with 4096 averages or more, the SNR can be sufficiently improved to 86dB or more. With the cascaded coaxial low pass filters, we generate a test signal with very low harmonic distortion. And with the number of averages on oscilloscope being 4096, the noise floor of oscilloscope acquisition is pushed to be much lower than the power of the possible third harmonics distortion.

To check the fidelity of the oscilloscope, we look at the captured signal in both frequency and time domains. First, we look at the Fourier transform of the filtered signal on the oscilloscope (Table 3.8). The power of a sinusoidal wave of 4000mV peak-peak amplitude on 50 Ohm is 16.0dBm. The power of the input test signal is about 8dBm. The Fourier transform of the signal acquired by the oscilloscope shows a third harmonic content with -52dBm power level. The spurious free dynamic range (SFDR) of the oscilloscope with 4000mV full-scale is -60dBc or -68dBFs, regardless of the sampling rate (20GS/s or 40GS/s) or acquisition bandwidth (4GHz or 8GHz) used.

We also examine the statistics of the acquired time-domain waveform on the oscilloscope. As shown in Table 3.9, the larger the acquisitive bandwidth the larger the noise on the amplitude and time measurements. The time-interleaving acquisition circuitry of the instrument appears to be well aligned and calibrated; otherwise, the 40GS/s sampling setting could possibly produce noise and spurs at a level higher than the 20GS/s setting. This is consistent with what we saw in the frequency domain where the SFDR is slightly better with the 40GS/s sampling setting.

For all measurements described later, we configure the DSA90804A oscilloscope at a sampling rate of 40GS/s and an acquisition bandwidth of 4GHz and 1024 av-
(a) A filtered sinusoid is used to characterize the oscilloscope.

(b) The input test sinusoid displayed on signal analyzer.

Figure 3-27: To quantify the nonlinearity of the oscilloscope in our measurement setup, an input test sinusoid is filtered heavily to reduce the second and higher harmonics.
Table 3.8: Third harmonic distortion introduced by the oscilloscope with different sampling rate and different acquisition bandwidths.

<table>
<thead>
<tr>
<th>Sampling rate; Acquisition bandwidth</th>
<th>3rd harmonic SFDR</th>
<th>3rd harmonic SFDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>20GS/s; 8GHz</td>
<td>-52.0dBm</td>
<td>-60.0dBc</td>
</tr>
<tr>
<td>20GS/s; 4GHz</td>
<td>-51.9dBm</td>
<td>-59.9dBc</td>
</tr>
<tr>
<td>40GS/s; 8GHz</td>
<td>-52.8dBm</td>
<td>-60.8dBc</td>
</tr>
<tr>
<td>40GS/s; 4GHz</td>
<td>-52.7dBm</td>
<td>-60.7dBc</td>
</tr>
</tbody>
</table>
erages, where the scope nonlinearity dominates over SNR but is still well below the nonlinearity of system under test.

Table 3.9: Noise on voltage and time measurement by the oscilloscope with different sampling rates and different acquisition bandwidths.

<table>
<thead>
<tr>
<th>Sampling Rate (GS/s)</th>
<th>Acquisition Bandwidth (GHz)</th>
<th>standard deviation in peak-peak amplitude (mV)</th>
<th>standard deviation in period (ps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>8</td>
<td>13.2</td>
<td>3.6</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>10.3</td>
<td>2.5</td>
</tr>
<tr>
<td>40</td>
<td>8</td>
<td>10.2</td>
<td>3.0</td>
</tr>
<tr>
<td>40</td>
<td>4</td>
<td>7.4</td>
<td>2.1</td>
</tr>
</tbody>
</table>

The block diagram of the RF transmitter under test is shown in Fig. 3-28. We use the ML605 evaluation kit to generate 16-bit in-phase and 16-bit quadrature samples at 245.76Msamples per second in the digital domain and send to the FMCOMMS1 module [62]. These samples are then converted into currents in the analog domain using the current digital-to-analog converters (DACs). The 16-bit accurate current is mixed with 1966.08MHz carrier at the phase modulator. The upconverted signals are passed through the preamplifier and then through the ADL5606 PA.

### 3.3.2 Phase Modulator Calibration

With the setup described above, we can now characterize the RF transmitter under test (Fig. 3-28). A sequence of random symbols in 64-QAM are generated and passed through a shaping filter in MATLAB. The shaped samples are programmed into the ML605 evaluation kit. The 16-bit in-phase and 16-bit quadrature of samples at 245.76Msamples/s are passed through the FPGA Mezzanine Card (FMC) interface to the FMCOMMS1 Module with ADL5606 PA board. The PA output is attenuated 10dB by the 8495A Manual Step Attenuator and is subsequently split into the N9030A PXA signal analyzer and the DSA90804A oscilloscope, respectively. The digitalized PA output is saved by the oscilloscope and then loaded into MATLAB for *ideal demodulation* [24] and downsampled into symbols. The carrier frequency, the
Figure 3-28: Block diagram of the transmitter including ML605 evaluation kit, FMCOMMS1 module and ADL5606 evaluation board.

Sampling rate, and the symbol rate are all integer multiples of the sequence frequency, as shown in Table 3.10.

Table 3.10: The synchronous setup of FMCOMMS1 module with the ADL5606 board.

<table>
<thead>
<tr>
<th>sequence (oscilloscope trigger)</th>
<th>frequency</th>
<th>remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>symbol</td>
<td>7.5kHz</td>
<td>each sequence contains (2^{14}) symbols</td>
</tr>
<tr>
<td>sample</td>
<td>(122.88\text{MHz} = 7.5\text{kHz} \times 2^{14})</td>
<td>contains (2^{14}) symbols</td>
</tr>
<tr>
<td>carrier</td>
<td>(245.76\text{MHz} = 7.5\text{kHz} \times 2^{15})</td>
<td>oversampling factor of 2</td>
</tr>
<tr>
<td>oscilloscope acquisition bandwidth</td>
<td>4GHz</td>
<td></td>
</tr>
<tr>
<td>oscilloscope sampling rate</td>
<td>40GS/s</td>
<td></td>
</tr>
</tbody>
</table>

With the setup described above and shown in Fig. 3-29, static compensation of the phase modulator is required after the shaping filter. The nonidealities of the phase modulator, if not calibrated out, would deteriorate transmitter performance, especially EVM. Any static nonidealities of the modulator should be calibrated before
we perform any dynamic compensation. Specifically, the nonidealities include offsets, in-phase quadrature gain mismatch, and in-phase quadrature imbalance. Their effects on EVM, for a 64QAM constellation, are illustrated in Fig. 3-30.

As outlined in Section 2.1.2, an ideal phase modulator outputs

\[ y(t) = I_{in} \sin(\omega t) + Q_{in} \cos(\omega t) \]  

at carrier frequency of \( \omega \) radians, given in-phase input \( I_{in} \) and quadrature input \( Q_{in} \). Here we are dealing with a single PA instead of an outphasing PA. As large amplitude variation at the phase modulator output is not tolerable, two current DACs for in-phase and quadrature (\( |I_{in}| + |Q_{in}| \) is not constant), instead of one DAC as in Section 2.1.2, are used.

If the gains of in-phase and quadrature components are \( g_I \) and \( g_Q \), the offsets on in-phase and quadrature are \( I_0 \) and \( Q_0 \), and the in-phase quadrature imbalance is \( \theta_0 \),
(a) Ideal constellation.

(b) Offset.

(c) IQ gain mismatch.

(d) IQ imbalance.

Figure 3-30: Static nonidealities of phase modulator.
then the nonideal modulator output is the following:

\[
y(t) = (g_I I_{in} + I_0) \sin(\omega t + \theta_0) + (g_Q Q_{in} + Q_0) \cos(\omega t)
\]

\[
= (g_I \cos \theta_0 I_{in}) \sin(\omega t) + (I_0 \cos \theta_0) \sin(\omega t)
\]

\[
+ (g_Q Q_{in} + g_I \sin \theta_0 I_{in}) \cos(\omega t) + (I_0 \sin \theta_0 + Q_0) \cos(\omega t).
\]  

(3.7)

Or, expressed in matrix form, the static nonidealities of in-phase output \( I_{out} \) and quadrature output \( Q_{out} \) are

\[
\begin{bmatrix}
I_{out} \\
Q_{out}
\end{bmatrix} = \begin{bmatrix}
\cos \theta_0 & 0 \\
\sin \theta_0 & 1
\end{bmatrix} \begin{bmatrix}
g_I \\
g_Q
\end{bmatrix} \begin{bmatrix}
I_{in} \\
Q_{in}
\end{bmatrix} + \begin{bmatrix}
I_0 \\
Q_0
\end{bmatrix}.  
\]  

(3.8)

With proper normalization \( g_I = 1/\cos \theta_0 \) on the gain of in-phase component, it becomes the following relation,

\[
\begin{bmatrix}
I_{out} \\
Q_{out}
\end{bmatrix} = \begin{bmatrix}
\cos \theta_0 & 0 \\
\sin \theta_0 & 1
\end{bmatrix} \begin{bmatrix}
1/\cos \theta_0 & 0 \\
-g_Q & g_Q
\end{bmatrix} \begin{bmatrix}
I_{in} \\
Q_{in}
\end{bmatrix} + \begin{bmatrix}
I_0 \\
Q_0
\end{bmatrix}.  
\]  

(3.9)

with four independent real-valued parameters: the gain mismatch \( g_Q \), the imbalance \( \theta_0 \), and the offsets \( I_0 \) and \( Q_0 \). This model of static nonidealities is illustrated in Fig. 3-31a.

To compensate for these static nonidealities, the following calibrated inputs \( I_{in,cal} \) and \( Q_{in,cal} \) should be computed for the modulator, instead of the original input \( I_{in} \) and \( Q_{in} \):

\[
\begin{bmatrix}
I_{in,cal} \\
Q_{in,cal}
\end{bmatrix} = \begin{bmatrix}
\cos \theta_0 & 0 \\
\frac{1}{g_Q} & \frac{1}{g_Q}
\end{bmatrix} \begin{bmatrix}
1/\cos \theta_0 & 0 \\
-g_Q & g_Q
\end{bmatrix} \begin{bmatrix}
I_{in} \\
Q_{in}
\end{bmatrix} - \begin{bmatrix}
I_0 \\
Q_0
\end{bmatrix}.  
\]  

(3.10)

After having done the multiplication between two matrices, it simplifies to,

\[
\begin{bmatrix}
I_{in,cal} \\
Q_{in,cal}
\end{bmatrix} = \begin{bmatrix}
1 & 0 \\
-g_Q \tan \theta_0 & 1
\end{bmatrix} \begin{bmatrix}
I_{in} \\
Q_{in}
\end{bmatrix} - \begin{bmatrix}
\cos \theta_0 I_0 \\
\frac{1}{g_Q} Q_0
\end{bmatrix}.  
\]  

(3.11)
\[ \cos(\omega t) \sin(\omega t + \theta_0) \]

\[ I_{in} \quad Q_{in} \]

\[ \frac{1}{\cos(\theta_0)} \quad I_0 \quad \sin(\omega t + \theta_0) \]

\[ Q_0 \quad \cos(\omega t) \]

\[ g_Q \quad Q_0 \]

\[ I_{out} \quad Q_{out} \cos(\omega t) \]

(a) A static model of a nonideal phase modulator

(b) static compensation of a nonideal phase modulator

Figure 3-31: A static model of nonidealities of phase modulator and hardware implementation of its static compensation.
In the hardware, this static compensation is straightforward to implement with three adders and two multipliers (Fig. 3-31b). Other representations of this type of static compensation have been outlined in [41, 63, 64]; they are mathematically equivalent to Equation (3.11).

To calibrate the modulator experimentally, a sequence of samples from a 64QAM constellation, without being spectrally shaped, is used as input. Each input sample is repeated for 128 sample periods (≈ 500ns); long enough to focus on the static nonidealities instead of any dynamic behavior. The maximum amplitude of input is kept below 1/4 of the full-scale of the current DAC. The small signal amplitude prevents excitation of other nonlinearities in the system. As shown in Fig. 3-32, this modulator has in-phase offset of 0.1% of full-scale and quadrature offset of 0.5% of full-scale, the quadrature component has a gain 0.11% higher than the in-phase component, and the angle between in-phase and quadrature is 2.6 degrees off 90 degrees. Comparing the input samples and the demodulated output samples, the four real-valued parameters, \( g_Q, \theta_0, I_0, \) and \( Q_0 \), are obtained from the experimental data with straightforward least squares optimization. For our efforts on compensation later in Section 3.3.3, this static compensation for the modulator is applied on the samples before the samples are programmed into the ML605 evaluation kit.

### 3.3.3 Off-line Iterative Sequence-based Compensation

With static nonidealities of the phase modulator compensated, our compensation efforts can then focus on the dynamic nonidealities without getting distracted by any static distortion. We characterize the PA at different input power levels. At input power of 5.2dBm, the PA produces an output power of 28.3dBm, and we call this operation point 0dB IBO. The relationship between input amplitude and output amplitude (AM-AM) and the relationship between input amplitude and the phase shift of output from input (AM-PM) at 0dB IBO are shown in Fig. 3-33. Clearly, as shown in Fig. 3-34, power back-off is an effective but costly approach to improve linearity: if we back off the input power of 5.2dBm by 12dB down to -6.8dBm, the ACPR of output improves from -23dB to -48dB and the EVM improves from 5.8%
to 1.9%, while power efficiency degrades from 24% to 3.3%.

In the IBO approach, trading for better linearity requires a significant sacrifice in power efficiency. For example, if the specifications are 23dBm output power and -48dB ACPR, with the power back-off approach, one needs to use four PAs with 17.8dBm (a 12dB IBO) output power each and a high-efficiency power combiner to deliver the combined 23.8dBm to meet the specification. The power efficiency of each PA operating at 17.8dBm output power is only 3.3%, compared to 11% if operating at 23.6dBm output power directly. An extra 5.3W is sacrificed to meet the linearity requirement. If the DPD power can be made smaller than the 5.3W IBO penalty, the total PAE of the transmitter can be improved.

In contrast, in the DPD approach, we apply the iterative off-line compensation method to improve linearity of the ADL5606 power amplifier board. With four iterations, we can achieve up to 25dB improvements in ACPR for this PA (Fig. 3-35) and reduce EVM down to 0.4% (Table 3.11). In particular, at 23.6dBm output power (6dB IBO), we can predistort the input signal to improve the ACPR from -35.1dB to -59.2dB after four iterations of compensation (Fig. 3-36). After compensation, the
Figure 3-33: AM-AM and AM-PM relationships, with the amplitude of the largest input symbol normalized to be 1.
Figure 3-34: ACPR of the PA improves while power efficiency decreases at larger IBO.

symbols with largest amplitudes still contribute most to EVM (Fig. 3-37).

Figure 3-35: ACPR of the PA improves over iterations of compensation.

At large output power above 26dBm (less than 3dB IBO), the ACPR is bad to start with, and the improvement is small because of the saturation of the current
Figure 3-36: Spectrum of uncompensated and compensated output of PA: at 6dB IBO, ACPR improves from -35.1dB to -59.2 after four iterations.

(a) After DPD, EVM of 1.47% at 0dB IBO.   (b) After DPD, EVM of 0.56% at 3dB IBO.

Figure 3-37: EVM plots of PA output after four iterations of compensation (with amplitude of the largest symbol at 0dB IBO normalized to be 1); ideal symbols are in red and the output symbols are in blue.
DACs during iterations. For large input power (small IBO), compensation gradually increases the peak power of input samples outside the maximum output range of the current DACs, as shown in the distribution plot (Fig. 3-38a). The iteration cannot be carried out when the input amplitude of some samples in the next iteration exceed the full-scale of current DAC output and saturated the mixer input. We circumvent the issue by clipping the input amplitude and carry on with iterations. Effectively we replace the samples with large amplitude with its closest allowable sample. It is expected that by clipping amplitude, the spectral improvement through iterations will be less. Because we pick the replacement sample closest to its original sample and because the probability of samples having large amplitude is low, the impact of amplitude clipping is limited and further iterations still give some improvement in ACPR.

![Graph](a) CCDF at 3dB IBO.  
(b) CCDF at 6dB IBO.

Figure 3-38: Complementary cumulative distributions of normalized power of samples (normalized with average power to be 0dB).

At a reasonably large range of output power but below 26dBm (more than 3dB
IBO), the improvement in ACPR is significantly more than 10dB (Fig. 3-35). In comparison with cases of less than 3dB IBO, at more than 3dB IBO, compensation also gradually increases the peak power of input samples but never exceeds the maximum output range of the current DACs, as shown in the distribution plot (Fig. 3-38b). At very small output power (12dB IBO), the ACPR is very good to start with, and the improvement is smaller because of the barrier of best achievable ACPR imposed by the precision, noise, and jitter.

Table 3.11: ACPR and EVM before and after off-line iterative compensation at different input power back-off.

<table>
<thead>
<tr>
<th>IBO (dB)</th>
<th>input power (dBm)</th>
<th>output power (dBm)</th>
<th>ACPR before (dB)</th>
<th>ACPR after (dB)</th>
<th>ACPR improvement (dB)</th>
<th>EVM before (%)</th>
<th>EVM after (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5.2</td>
<td>28.4</td>
<td>-23.2</td>
<td>-28.7</td>
<td>5.5</td>
<td>5.83</td>
<td>1.47</td>
</tr>
<tr>
<td>1</td>
<td>4.2</td>
<td>27.7</td>
<td>-25.0</td>
<td>-31.4</td>
<td>6.4</td>
<td>4.77</td>
<td>1.29</td>
</tr>
<tr>
<td>2</td>
<td>3.2</td>
<td>27.0</td>
<td>-26.9</td>
<td>-35.0</td>
<td>8.1</td>
<td>3.83</td>
<td>0.54</td>
</tr>
<tr>
<td>3</td>
<td>2.2</td>
<td>26.2</td>
<td>-29.4</td>
<td>-39.4</td>
<td>10.0</td>
<td>3.10</td>
<td>0.56</td>
</tr>
<tr>
<td>4</td>
<td>1.2</td>
<td>25.3</td>
<td>-31.0</td>
<td>-45.5</td>
<td>14.5</td>
<td>2.81</td>
<td>0.43</td>
</tr>
<tr>
<td>5</td>
<td>0.2</td>
<td>24.6</td>
<td>-33.2</td>
<td>-52.4</td>
<td>19.1</td>
<td>2.42</td>
<td>0.41</td>
</tr>
<tr>
<td>6</td>
<td>-0.8</td>
<td>23.6</td>
<td>-35.1</td>
<td>-59.2</td>
<td>24.2</td>
<td>2.37</td>
<td>0.36</td>
</tr>
<tr>
<td>9</td>
<td>-3.8</td>
<td>20.8</td>
<td>-40.5</td>
<td>-65.5</td>
<td>25.1</td>
<td>2.07</td>
<td>0.31</td>
</tr>
<tr>
<td>12</td>
<td>-6.8</td>
<td>17.8</td>
<td>-48.0</td>
<td>-66.0</td>
<td>18.0</td>
<td>1.87</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Looking at Fig. 3-35, we see that in two iterations most of the improvement can be achieved. At the third iteration, further improvement is smaller and the Lipschitz constant of the iteration has approached or exceeded 1 quickly (Fig. 3-39).

3.3.4 Real-time Compensator Model

To model the nonlinearities seen in this RF PA system, we use a non-obvious but analytically motivated model structure (Fig. 3-40) as proposed in [25].

This model is a sum of three submodels. Each submodel is a nonlinear system with one-sample memory cascaded by an LTI system. The first part of the predistorter submodel represents the nonlinear function basis terms as a product of a polynomial
Figure 3-39: The Lipschitz constants of the iterations.

Figure 3-40: Compensator structure proposed in [25].
with one memory in amplitude and a sum of in-phase and quadrature components:

\[ V_k[n] = \sum_{i=1}^{N} \left( c_{0,i,k} I[n] + d_{0,i,k} Q[n] + c_{1,i,k} I[n-1] + c_{1,i,k} Q[n-1] \right) a[n]^{m_{0,i}} a[n-1]^{m_{1,i}}, \]

(3.12)

where \( k = 0, 1, 2 \) for the three models; \( I[n], Q[n] \), and \( a[n] = \sqrt{I[n]^2 + Q[n]^2} \) are in-phase component, quadrature component, and amplitude of input sample; \( I[n-1] \) and \( Q[n-1] \) are the one-sample delayed version of \( a[n] \), \( I[n] \), and \( Q[n] \), respectively; \( c_{0,i,k}, c_{1,i,k}, d_{0,i,k}, \) and \( d_{1,i,k} \) are the design variables; \( 0 \leq m_{0,i} \leq m_0 \) and \( 0 \leq m_{1,i} \leq m_1 \) are the degrees of the polynomial in amplitude; and where \( N \) is the total number of the terms. With sum of degrees \((m_{0,i} + m_{1,i})\) up to 2, there are six amplitude polynomials: \( 1, a[n], a[n-1], a[n]^2, a[n]a[n-1], a[n-1]^2 \). In total we have 24 coefficients in the nonlinear function part of predistorter for each \( k \). The second part of the predistorter submodel are the LIT systems given by

\[ H_0(e^{j\Omega}) = 1, H_1(e^{j\Omega}) = j\Omega, H_2(e^{j\Omega}) = 1 - 3\Omega^2. \]

(3.13)

With 24 terms in the nonlinear function part for each LTI system and 3 LTI systems, we have 72 complex coefficients, or 144 real coefficients.

At 6dB IBO, the iterative off-line compensation improves the ACPR from -35.1dB to -59.2dB in four iterations. Using the real-time predistorter model with 144 real coefficients, we can fit the desired predistorted input accurate to 0.7%, and applying the input from this real-time model, the ACPR is improved to -47.8dB, which is more than 12dB improvement. At 3dB IBO, using this real-time model with a different set of 144 real coefficients, the ACPR is improved from -29.4dB to -36.5dB.

### 3.3.5 Implementation of Digital Compensator for RF PA

We implement the above predistorter model on a 40 nm-technology XC6VLX240T-1FFG1156 FPGA on the ML605 Virtex-6 FPGA evaluation kit.

For the nonlinear part of the model, with sum of degrees in amplitudes up to 2, there are six amplitude polynomials: \( 1, a[n], a[n-1], a[n]^2, a[n]a[n-1], a[n-1]^2 \);
because the $a[n - 1]$ and $a[n - 1]^2$ term are associated with insignificant coefficients, we do not include them so we are left with a set of four amplitude polynomials. This reduces the number of real coefficients from 144 to 96 without any reduction in goodness of fit of the model. For this LTI system in Equation (3.13), we use $H_2(e^{j\Omega}) = 1 - 3\Omega^2$ for the quadratic term because it makes $h_2[0] = 0$; other formulations, such as $H_2(e^{j\Omega}) = \Omega^2$ used in [25], can also be used but require an additional DSP slice on the FPGA. With $h_1[n]$ being odd symmetric with $h_1[0] = 0$ and $h_2[n]$ being even symmetric with $h_2[0] = 0$, when mapped to FPGA hardware a $(2m + 1)$-tap FIR approximating either $h_1[n]$ or $h_2[n]$ requires only $m$ DSP slices instead of $2m+1$ slices, and the $h_0[n] = 1$ does not require any DSP slices at all. Mapping a symmetric FIR efficiently onto ASIC was illustrated in Fig. 3-19 and Fig. 3-20; mapping a symmetric FIR onto FPGA is similar, but not exactly the same because of the structure of dedicated DSP slices (Fig. 3-42).

The implemented predistorter utilizes 144 slices of DSP48E1 blocks. At 246MHz operation (to support a symbol rate of 123Msymbols/s and a data rate of 738Mbps), it consumes 0.42W of power. At 6dB IBO, without digital compensation the PA delivers 23.6dBm output power at 11.0% efficiency. The additional 0.42W power consumed by the digital predistorter would decrease the efficiency down to 9.2% which is still much better than 3.3% required by 12dB IBO for the uncompensated PA to
(a) A 5-tap FIR filter $h_2[n]$ mapped onto FPGA with 5 DSP48E1 slices.

(b) Apply the even symmetric property $h_2[-n] = h_2[n]$ on (a), using the D input ports and the pre-adders in DSP48E1 slices; add 1 additional pipeline stage after pre-adders.

(c) Apply the property $h_2[0] = 0$ on (b); remove 3 unused pipeline stages.

Figure 3-42: Because of its symmetric properties and that the center tap vanishes, a $(2m+1)$-tap FIR filter $h_2[n]$ requires only $m$ DSP48E1 slices, as illustrated above with $m = 2$. 

116
achieve the same ACPR. In Fig. 3-43, with input power back-off, the ACPR improves with larger IBO and lower power efficiency. With off-line iterative compensation, we improve the ACPR and the operating points move down from uncompensated (shown as blue crosses) to off-line compensated (shown as magenta pluses) without considering the power of off-line predistorter. With this real-time predistorter implemented, we achieve a slightly less but similar amount of ACPR improvement with off-line compensation, and the operating points move up slightly from off-line compensated to real-time compensated (shown as red squares).

When considering the power footprint of the real-time predistorter, the overall power efficiency degrades, and the operating points lose efficiency from off-line compensated to real-time compensated. The operating points with real-time predistorter are still significantly more efficient than the operating points without compensation. For a given ACPR requirement, we can improve power efficiency without violating the ACPR requirement. For example, to meet an ACPR constraint of -48dB, the uncompensated PA has to back-off the input power by 12dB such that the power efficiency is 3.3%. The compensated PA has to back-off by only 6dB such that the overall transmitter power efficiency is 9.2% with the DPD power included, almost $3 \times$ the efficiency of uncompensated PA. For another example, to meet an ACPR constraint of -40dB, the uncompensated PA has to back-off the input power by 9dB such that the power efficiency is 6.2%. The compensated PA has to back-off by only 4dB such that the overall transmitter power efficiency is 12.1% with the DPD power included, almost $2 \times$ the efficiency of uncompensated PA.

In Fig. 3-44, the EVM numbers are plotted against power efficiency. The operating points with the real-time predistorter are significantly below the operating points without compensation: for a fairly high power efficiency of 20% and below, we can improve EVM without sacrificing power efficiency. The operating points with the real-time predistorter are significantly more efficient than the operating points without compensation: for a given EVM requirement, we can improve power efficiency without violating the EVM requirement.

It is important to quantify the energy efficiency of this digital predistorter in the
Figure 3-43: ACPR of the PA improved by real-time predistorter with gains in power efficiency for the same ACPR over the uncompensated PA.

Figure 3-44: EVM of the PA improved by real-time predistorter with a small penalty in power efficiency.
context of overall transmitter efficiency. In Fig. 3-45, we plot the raw RF efficiency versus output power in blue for PAs with carrier frequencies in the range of 1GHz to 6GHz in either bulk CMOS, SOI CMOS, or SiGe technologies, from IEEE references in year 2006 to year 2012. The black contour lines identify the PAs where the overall system efficiency will be degraded by 2% or 5% respectively, by the 0.42W power of this predistorter. For example, for any PAs with output power smaller than 1000mW and more than 24% power efficiency, the overall power efficiency of the system will degrade significantly by more than 2% by including the 0.42W predistorter power.

To visualize the efficiency penalty due to the energy footprint of digital predistorter, the power efficiency of the system after having included the 0.42W predistorter power is plotted in green circles. As we have seen in Section 2.2, if this digital predistorter is implemented on ASICs with dedicated custom design and routing instead of in FPGAs, its power footprint could be $10 \times$ smaller, and the efficiency penalty on the overall transmitter efficiency would be much smaller. It is important to note here that with DPD the PAs would operate potentially with less back-off so at a better efficiency point, which is not represented in the plot.

The Q-band (45GHz) outphasing amplifiers and the RF PA at 1.97GHz carrier are two examples of general interest to the community that we have hereto used to illustrate our framework. Applying the hereto developed methodology, a system-level circuit designer can easily identify and mathematically model other completely different analog systems. For an example, one can model a polar PA or analog-to-digital converter, and then derive a predistorter architecture from the identified mathematical structure. Once a predistorter architecture is proposed, designers can apply all aforementioned techniques to reduce the energy footprint of their digital predistorter design for their high-throughput and area-constrained application.

### 3.4 Constellation Design

With either the Q-band outphasing PA or the RF PA, we used symbols from the 64QAM constellation. The 64QAM constellation on a square grid is widely adopted in
Figure 3-45: Power efficiency degradation by a 246Msamples/s-throughput predistorter of the 1.7nJ/sample energy efficiency. Note that the uncompensated PA operates at same input power and has worse ACPR than the PA with DPD. To meet the linearity requirements, the uncompensated PA would have to operate with IBO which leads to lower overall PAE than with our DPD.

many communication standards, partly due to its simplicity and straightforwardness in digital encoding and decoding. In Fig. 3-37, we see that after compensation, the four symbols with the largest amplitude in 64QAM contribute most to EVM while all other symbols contribute almost no error to EVM.

In considerations to either bit error rate (BER) or output power, the 64QAM constellation is not optimized for requirements of modern communications. In particular, its $D_{\text{min}}$ is small and its PAPR is large. Small $D_{\text{min}}$ results in poor BER, and for a given PA with specified peak output power, large PAPR would result in small average output power and hence reduce the transmission range. Alternative constellations can do better by increasing $D_{\text{min}}$ or reducing PAPR or both.

To support alternative constellations in our methodology is simple, at least on the transmitter side, as only a LUT is required. This LUT would be inserted before the $2 \times$ upsampling and the shaping filter $H_s(z)$ in Fig. 3-14. Since it is before the upsampling from symbols to samples, the dynamic power consumption of the LUT is small because it is running at the symbol rate (half the sample rate for our designs).
The high precision requirements of modern communication systems demand the LUT output to be 12 bits or more, but the LUT input address is still 6 bits for a 64-symbol constellation. Since this LUT can be small with only 64 entries, there is almost no additional cost in area to the digital baseband.

Let us first define the metrics of constellation performance. The PAPR, usually referring to samples, is defined as the ratio of peak power of the samples $V$ to their average power, where

$$\text{PAPR(dB)} = 20 \log_{10} \frac{|V|_{\text{peak}}}{|V|_{\text{rms}}}.$$  

In this discussion, we refer the PAPR to symbols instead,

$$\text{PAPR(dB)} = 20 \log_{10} \frac{|S|_{\text{peak}}}{|S|_{\text{rms}}}.$$  

where the symbols $S$ are oversampled and spectrally shaped to obtain the samples $V$. The PAPR on samples is generally larger than the PAPR on symbols. The magnitude of the increase from the PAPR on symbols to the PAPR on samples generally depends on the shaping filter optimization over the input signal statistics, but can also depend on the constellation. For simplicity of this discussion, we assume that the magnitude of the increase depends only on the oversampling ratio and the characteristics of the shaping filter, and we will specifically discuss the PAPR on symbols not the PAPR on samples. To simplify the discussion further, we assume that all symbols in a constellation are equally probable.

PAPR and $D_{\text{min}}$ are the two important metrics we will be considering for constellations. The linearity of a PA mostly depends on its peak power or its peak amplitude. If we increase minimum symbol distance, $D_{\text{min}}$, by optimizing the constellation design and keeping the peak amplitude the same, we are able to reduce BER without sacrificing linearity. If we reduce PAPR by optimizing the constellation design and keeping the peak amplitude the same, we are able to increase the transmission range without sacrificing linearity. If the peak amplitude stays the same, the improvements are possible without the necessity of upgrading to a device with larger peak output.
power.

3.4.1 Alternative 64-Symbol Constellations to 64QAM

For convenience, we normalize the peak amplitude of any considered constellation to be 1. With this normalization in which the peak amplitude of the 64QAM constellation is 1, either the maximum in-phase amplitude or the maximum quadrature amplitude of the 64QAM constellation is $1/\sqrt{2} = 0.7071$, and its minimum symbol distance $D_{\text{min}}$ is $\sqrt{2}/7 = 0.2020$.

Only four out of 64 symbols in the 64QAM constellation are at the peak amplitude, while the rest are all more than 1dB below the peak amplitude. Besides the 12 symbols with the largest amplitude or the second largest amplitude, the other 52 symbols are all more than 2dB below the peak amplitude. If we move these 12 symbols away to other locations on this square grid, we can obtain a modified constellation; we will call this “64QAM-MOD1” (Fig. 3-46b). If we scale the 64QAM-MOD1 constellation to have the same peak amplitude of 1, the minimum symbol distance $D_{\text{min}}$ is increased by 4.4% and PAPR is reduced by 1.62dB from 64QAM. If we move the four symbols at peak amplitude away to other locations on this square grid, we can obtain another alternative constellation; we will call this “64QAM-MOD2” (Fig. 3-46c). The $D_{\text{min}}$ is increased by 9.4%, and PAPR is reduced by 1.17dB from 64QAM. If we move the four symbols at peak amplitude away to other locations but not restricted to the square grid, this “64QAM-MOD3” constellation (Fig. 3-46d) has $D_{\text{min}}$ increased by 13.4%.

With all symbols remaining on a square grid, the required change in the receiver architecture to accommodate either the 64QAM-MOD1 constellation or the 64QAM-MOD2 constellation is minimal. The 64QAM-MOD1 and the 64QAM-MOD2 allow up to 12 additional symbols and four additional symbols, respectively, for redundancy (highlighted on figures in yellow) when compatible with 64QAM, which can be used with coding to allow further reduction of bit error rate at the sacrifice of PAPR.

If a polar receiver is available instead of a quadrature receiver, amplitude and phase-shift keying (APSK) can be used for constellation design. For example, an alternative 64-symbol constellation, that we will call “64APSK1” (Fig. 3-47a), can
Figure 3-46: The 64QAM constellations on a square grid, and its modified variants. The peak amplitude, normalized to be 1, is shown by the outer green circle. The diameter of red circles is the minimum symbol distance $D_{\text{min}}$. If the designed constellations, given the peak amplitude, supports more than 64 symbols, the symbols with smallest amplitudes are chosen to be redundant symbols to minimize PAPR; redundant symbols are highlighted in yellow.
be created from overlapping the 3PSK, 10PSK, 17PSK, 24PSK, and 31PSK constellations closely packed on concentric circles. However, this straightforward APSK constellation has too many redundant symbols, and its $D_{\text{min}}$ has almost no improvement above 64QAM while its PAPR is significantly reduced by 2.44dB from 64QAM.

Reducing the number of redundant symbols, we can construct other 64-symbol APSK constellations with larger $D_{\text{min}}$. A “64APSK2" constellation can be constructed from overlapping the 9PSK, 16PSK, 23PSK, and 30PSK constellations as shown in Fig. 3-47b. A “64APSK3" constellation can be constructed from overlapping the 8PSK, 15PSK, 22PSK, and 29PSK constellations as shown in Fig. 3-47c. Similarly, a “64APSK4" constellation can be constructed from overlapping the 7PSK, 14PSK, 21PSK, and 28PSK constellations as shown in Fig. 3-47d. Finally, a “64APSK6" constellation can be constructed from overlapping the 6PSK, 13PSK, 20PSK, and 27PSK constellations as shown in Fig. 3-48b; with two redundant symbols, it has $D_{\text{min}}$ increased by 15.0% and PAPR reduced by 1.78dB from 64QAM. Given fixed peak symbol amplitude, we clearly see a tradeoff between $D_{\text{min}}$ and PAPR with these APSK constellations.

By removing the redundant symbols in the 64APSK4 constellation but keeping the $D_{\text{min}}$ unchanged, a “64APSK5" constellation (Fig. 3-48a) of wide interest [65–67] can be obtained.

Symbol redundancy can, however, be potentially used for error detection over a noisy channel. In the context of an outphasing transmitter, redundant symbols can be removed completely, or possibly swapped with a real-time algorithm, to achieve more effective zero-avoidance.

Instead of on a square grid or on concentric circles, symbols of a constellation can be placed on a hexagonal grid to achieve significantly larger $D_{\text{min}}$ given the same peak amplitude. A “64HEX1" constellation (Fig. 3-49a) is 13.6% larger in $D_{\text{min}}$ than the 64QAM constellation. Without any increase in peak amplitude, 64HEX1 allows up to eight additional symbols for redundancy when compared with 64QAM. If symmetry is unnecessary, a “64HEX2" constellation (Fig. 3-49b) can be created on the hexagonal grid shifted slightly to the left [68]. The 64HEX2 constellation is 17.9% larger in $D_{\text{min}}$.
Figure 3-47: Constellations on concentric circles. The peak amplitude, normalized to be 1, is shown by the green circle. The diameter of red circles is $D_{\text{min}}$. Redundant symbols are highlighted in yellow.
Figure 3-48: More constellations on concentric circles. The peak amplitude, normalized to be 1, is shown by the green circle. The diameter of red circles is $D_{\text{min}}$. Redundant symbols are highlighted in yellow circles.

than the 64QAM constellation.

Unfortunately, for a typical quadrature receiver, this 64HEX2 constellation would require 16 comparators with different thresholds on the in-phase component and eight comparators on the quadrature component, compared to seven comparators on either in-phase or quadrature for the 64QAM constellation (Fig. 3-50b). The threshold interval on the in-phase component of 64HEX2 constellation is almost halved from the threshold interval on the in-phase of 64QAM. The offset and sensitivity requirements on the comparator for 64HEX2 is higher than for 64QAM, possibly contributing to even higher power consumption on the receiver side.

Denser packing of symbols can be achieved, if we give up the symmetry requirement. Dense packing of congruent circles in a large circle is a classic problem in discrete mathematics, and the solution for the dense packing 64 circles in a circle has been reported in [69–71]. If we use this dense packing solution as a symbol constellation “64DENSE” (Fig. 3-51), we can increase $D_{\text{min}}$ by 24.4% over 64QAM and reduce PAPR by 1.48dB while keeping the same peak symbol amplitude. This comes with more significant changes in requirements on the receiver side though. The number
Figure 3-49: The constellations on a hexagonal grid. The peak amplitude, normalized to be 1, is shown by the green circle. The diameter of red circles is $D_{\text{min}}$. Redundant symbols are highlighted in yellow.

Figure 3-50: Different constellations require different receiver complexity: (a) quadrature receiver of 64QAM constellation requires seven thresholds on either in-phase or quadrature, with threshold interval being $\sqrt{2}/7 = 0.2020$; (b) quadrature receiver of 64HEX2 constellation requires 16 thresholds on in-phase and eight thresholds on quadrature, with threshold intervals being $5/42 = 0.1190$ and $5\sqrt{3}/42 = 0.2062$. 
and sensitivity of comparators required on the receiver side could make 64DENSE impractical.

Figure 3-51: The most closely packed 64-symbol constellation. The peak amplitude, normalized to be 1, is shown by the green circle. The diameter of red circles is $D_{\text{min}}$.

Different designs of 64-symbol constellation are summarized in Table 3.12. The constellation with largest $D_{\text{min}}$ or smallest PAPR, given the same peak amplitude, is not necessarily the best for all designs. It is a tradeoff between bit error rate, transmission range, and receiver complexity. Optimization on symbol constellation for BER and power efficiency requirement should consider the whole communication system including the channel and the receiver, in even more detail.

3.4.2 More Bits Per Symbol

The energy-efficiency disadvantage of QAM constellation on a square grid only gets worse, compared to densely packed constellation on a hexagonal grid, as we demand more and more bits in a symbol. The recent trend of higher utilization of spectral bandwidth demands more bits in a symbol. For example, the 256QAM constellation gets 8 bits per symbol, which is 2 more bits per symbol than the 64QAM constellation. The 33% increase in channel capacity using 256QAM constellation over 64QAM is
Figure 3-52: Tradeoff between the minimum distance $D_{\text{min}}$ (with peak symbol amplitude of each constellation normalized to be 1) and PAPR of 64-symbol constellation design.
Table 3.12: The minimum symbol distance $D_{\text{min}}$ and PAPR (before being spectrally shaped, and assuming that all symbols in a constellation are equally probable and redundant symbols are never used) of different constellation designs, with peak symbol amplitude of each constellation normalized to be 1.

### 16-symbol constellations.

<table>
<thead>
<tr>
<th>Constellation</th>
<th>$D_{\text{min}}$</th>
<th>Symbol Amplitude rms</th>
<th>PAPR (dB)</th>
<th>Redundancy in Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>16PSK</td>
<td>$2\sin(\pi/16)$</td>
<td>0.3902</td>
<td>1.0000</td>
<td>0.00</td>
</tr>
<tr>
<td>16QAM</td>
<td>$\sqrt{2}/3$</td>
<td>0.4714</td>
<td>0.7454</td>
<td>2.55</td>
</tr>
<tr>
<td>16APSK1 (5PSK+12PSK)</td>
<td>$2\sin(\pi/12)$</td>
<td>0.5176</td>
<td>0.8936</td>
<td>0.98</td>
</tr>
<tr>
<td>16DENSE [73]</td>
<td></td>
<td>0.5532</td>
<td>0.8700</td>
<td>1.21</td>
</tr>
</tbody>
</table>

### 64-symbol constellations.

<table>
<thead>
<tr>
<th>Constellation</th>
<th>$D_{\text{min}}$</th>
<th>Symbol Amplitude rms</th>
<th>PAPR (dB)</th>
<th>Redundancy in Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>64QAM</td>
<td>$\sqrt{2}/7$</td>
<td>0.2020</td>
<td>0.6547</td>
<td>3.68</td>
</tr>
<tr>
<td>64QAM-MOD1</td>
<td>$\sqrt{2}/5/3$</td>
<td>0.2108</td>
<td>0.7888</td>
<td>2.06</td>
</tr>
<tr>
<td>64QAM-MOD2</td>
<td>$\sqrt{2}/41$</td>
<td>0.2209</td>
<td>0.7489</td>
<td>2.51</td>
</tr>
<tr>
<td>64QAM-MOD3</td>
<td>$2/(7+\sqrt{3})$</td>
<td>0.2290</td>
<td>0.7301</td>
<td>2.73</td>
</tr>
<tr>
<td>64APSK1</td>
<td>$2\sin(\pi/31)$</td>
<td>0.2023</td>
<td>0.8673</td>
<td>1.24</td>
</tr>
<tr>
<td>64APSK2</td>
<td>$2\sin(\pi/30)$</td>
<td>0.2091</td>
<td>0.8543</td>
<td>1.37</td>
</tr>
<tr>
<td>64APSK3</td>
<td>$2\sin(\pi/29)$</td>
<td>0.2162</td>
<td>0.8405</td>
<td>1.51</td>
</tr>
<tr>
<td>64APSK4</td>
<td>$2\sin(\pi/28)$</td>
<td>0.2239</td>
<td>0.8241</td>
<td>1.68</td>
</tr>
<tr>
<td>64APSK5</td>
<td>$2\sin(\pi/28)$</td>
<td>0.2239</td>
<td>0.7964</td>
<td>1.98</td>
</tr>
<tr>
<td>64APSK6</td>
<td>$2\sin(\pi/27)$</td>
<td>0.2322</td>
<td>0.8032</td>
<td>1.90</td>
</tr>
<tr>
<td>64HEX1</td>
<td>$1/\sqrt{19}$</td>
<td>0.2294</td>
<td>0.7727</td>
<td>2.24</td>
</tr>
<tr>
<td>64HEX2</td>
<td>$5/21$</td>
<td>0.2381</td>
<td>0.7183</td>
<td>2.87</td>
</tr>
<tr>
<td>64DENSE</td>
<td></td>
<td>0.2512</td>
<td>0.7760</td>
<td>2.20</td>
</tr>
</tbody>
</table>

### 256-symbol constellations.

<table>
<thead>
<tr>
<th>Constellation</th>
<th>$D_{\text{min}}$</th>
<th>Symbol Amplitude rms</th>
<th>PAPR (dB)</th>
<th>Redundancy in Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>256QAM</td>
<td>$\sqrt{2}/15$</td>
<td>0.0943</td>
<td>0.6146</td>
<td>4.23</td>
</tr>
<tr>
<td>256APSK1</td>
<td>$2\sin(\pi/57)$</td>
<td>0.1102</td>
<td>0.7560</td>
<td>2.43</td>
</tr>
<tr>
<td>256DENSE</td>
<td></td>
<td>0.1213</td>
<td>0.7323</td>
<td>2.71</td>
</tr>
</tbody>
</table>

### 1024-symbol constellations.

<table>
<thead>
<tr>
<th>Constellation</th>
<th>$D_{\text{min}}$</th>
<th>Symbol Amplitude rms</th>
<th>PAPR (dB)</th>
<th>Redundancy in Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>1024QAM</td>
<td>$\sqrt{2}/31$</td>
<td>0.0456</td>
<td>0.5957</td>
<td>4.50</td>
</tr>
<tr>
<td>1024APSK1</td>
<td>$2\sin(\pi/117)$</td>
<td>0.0537</td>
<td>0.7323</td>
<td>2.71</td>
</tr>
<tr>
<td>1024DENSE</td>
<td></td>
<td>0.0600</td>
<td>0.7165</td>
<td>2.90</td>
</tr>
</tbody>
</table>
Figure 3-53: Alternative 256-symbol constellation designs.

rather attractive in many applications.

With more bits per symbol, the amount of improvement in $D_{\text{min}}$ through constellation optimization is even larger. If we normalize the peak amplitude of the standard 256QAM constellation to be 1, the maximum in-phase amplitude of the 256QAM constellation is $\frac{1}{\sqrt{2}} = 0.7071$ and $D_{\text{min}}$ is $\sqrt{2}/15 = 0.0943$. Overlapping 8PSK, 15PSK, 22PSK, ..., 50PSK, and 57PSK constellations with four redundant symbols as shown in Fig. 3-53a, the $D_{\text{min}}$ of a “256APSK1” constellation is 0.1102, which is 17% larger than the standard 256QAM constellation, and its PAPR is reduced by 1.8dB. Using the densest packing of 256 congruent circles in a large circle [71] as shown in Fig. 3-53b, the $D_{\text{min}}$ of a “256DENSE” constellation is 0.1213, which is 29% larger than 256QAM, and its PAPR is reduced by 1.5dB from 256QAM. The improvement is similar in 1024-symbol constellations. The $D_{\text{min}}$ of the 1024APSK1 constellation (overlapping 5PSK, 12PSK, 19PSK, ..., 110PSK, and 117PSK constellations, with 13 redundant symbols) is 18% larger than the standard 1024QAM constellation, and its PAPR is reduced by 1.8dB. Using the densest packing of 1024 congruent circles in a large circle [71], the $D_{\text{min}}$ of a “1024DENSE” constellation is 32% larger than 1024QAM, and its PAPR is reduced by 1.6dB from 1024QAM.
Beyond 1024 symbols, now consider an $m$-QAM constellation on a square grid, where $m = n^2$ and $n$ is an even integer. If we normalize the peak amplitude of the standard $m$-QAM constellation to be 1, the maximum in-phase amplitude of the $m$-QAM constellation is $\frac{1}{\sqrt{2}} = 0.7071$, and the $D_{\text{min}}$ is:

$$D_{\text{min}}[m\text{-QAM}] = \frac{1}{\sqrt{2}} \frac{2}{n-1} = \frac{2}{\sqrt{2}\sqrt{m-1}} \approx \sqrt{\frac{2}{m}}. \quad (3.16)$$

It is also straightforward to show that the rms amplitude of $m$-QAM is:

$$|S|_{\text{rms}}[m\text{-QAM}] = \sqrt{\frac{1}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} \left[ \left( \frac{1}{\sqrt{2}} \frac{2k-n-1}{n-1} \right)^2 + \left( \frac{1}{\sqrt{2}} \frac{2l-n-1}{n-1} \right)^2 \right]}
= \sqrt{\frac{1}{3} \frac{(n+1)}{(n-1)}} = \sqrt{\frac{1}{3} \frac{\sqrt{m+1}}{(\sqrt{m-1})}}. \quad (3.17)$$

As $m \to \infty$, the rms amplitude of $m$-QAM asymptotically approaches $\sqrt{1/3} \approx 0.5774$, so the PAPR is then 4.77dB.

For a densely packed $m$-DENSE constellation with its peak amplitude also normalized to be 1, at very large $m$, the packing of symbols gets on a regular hexagonal grid. The $D_{\text{min}}$ of $m$-DENSE constellation for very large $m$ is:

$$D_{\text{min}}[m\text{-DENSE}] = \sqrt{\frac{\pi}{m\sqrt{3}}} \approx 1.35\sqrt{\frac{2}{m}}, \quad (3.18)$$

which is 35% larger than the $D_{\text{min}}$ of the $m$-QAM constellation. The rms amplitude of the $\infty$-DENSE constellation is, estimated by approximating discrete with continuous variables such that:

$$|S|_{\text{rms}}[m\text{-DENSE}] = \sqrt{\frac{\int_0^1 r^2 * 2\pi r dr}{\int_0^1 2\pi r dr}} = \sqrt{\frac{1}{2}}. \quad (3.19)$$

and the PAPR is then 3.01dB, which is 1.76dB reduced from the $\infty$-QAM constellation.

The trend of more significant $D_{\text{min}}$ improvement of $m$-DENSE constellation over
Figure 3-54: $m$-DENSE constellation versus $m$-QAM constellation, with the peak amplitude of all considered constellations normalized to be 1.

$m$-QAM constellation with larger $m$ is illustrated in Fig. 3-54a; the trend of greater reduction in PAPR of $m$-DENSE constellation over $m$-QAM constellation, given the same peak amplitude that is larger increase in rms amplitude of $m$-DENSE over $m$-QAM, with larger $m$ is illustrated in Fig. 3-54b. As we move into the realm of packing more and more bits per symbol, it is worth the effort to consider the more densely packed constellations over the standard QAM constellations.

### 3.4.3 Limitations

For simplicity, so far we have only studied PAPR on symbols, and we have assumed that the magnitude of the increase from the PAPR on symbols to the PAPR on samples is constant, where:

$$\text{PAPR}_{\text{sample}} = \text{PAPR}_{\text{symbol}} + \Delta \text{PAPR}_{\text{shaping filter}}.$$  \hspace{1cm} (3.20)

We have thus far searched for constellations with smaller PAPR on symbols, in the hope of achieving smaller PAPR on samples. Implicitly, we have assumed $\Delta \text{PAPR}$
depends only on the oversampling ratio and the characteristics of the shaping filter and does not depend on the characteristics of the symbol constellation. That assumption needs to be challenged, as $\Delta$PAPR can also depend on the symbol constellation and should be studied to optimize the shaping filter specific to a constellation. We have also assumed that all symbols in a constellation are equally probable, and this assumption can also be replaced with the statistical distribution from real data. We leave the optimization of shaping filter in the context of nonequal input symbol probability to future work.
Chapter 4

Conclusion and Future Research Directions

4.1 Conclusion

To enable a new class of energy-efficient mm-wave communication systems based on outphasing PAs (PAs), we have demonstrated a methodology to implement high-throughput outphasing baseband functions with smaller energy and area footprints. Aware of the limitations of FPGAs in throughput and energy-efficiency, we have implemented our digital baseband in ASICs on a 45nm SOI technology, to allow for a truly integrated transmitter.

By utilizing the changes in micro-architecture as described in Section 2.3.3 and aggressive back-end power optimization techniques as described in Section 2.3.2, we achieve a record energy-efficiency and throughput for AMO SCS of 32pJ/sample at 0.6V supply voltage and 400Msamples/s, with area of 0.41mm². For high-throughput area-constrained applications, our SRAM-based AMO SCS design, described in Section 2.3.4, achieves 2× area reduction over a register-based design at the same throughput to allow more parallelism to meet the stringent throughput requirement.

We have aggressively optimized the design of a high-throughput AMO SCS with optimized energy and area footprints. These design techniques are readily applicable to other high-throughput digital system designs with high energy efficiency require-
We developed a linearization methodology for PAs. To illustrate the wide applicability of this proposed nonlinearity compensation methodology, we experimentally apply the proposed methodology to a PA with 1.97GHz carrier frequency and 737Mbps data throughput with 64QAM constellation. First, we proved the feasibility of RF PA compensation in digital baseband using the off-line sequence-based iterative compensation. With off-line compensation, the ACPR can be improved by up to 25dB, as shown in Section 3.3.3. We demonstrated the goodness of fit of the proposed dynamical model with experimental data on this RF PA under study. We implemented the dynamical model predistorter in hardware in an energy-efficient way to improve linearity. With the real-time predistorter implemented in an FPGA, the ACPR can be improved by up to 12dB as shown in Section 3.3.5, or equivalently the PA can operate at smaller back-off, improving the overall efficiency. Finally, in Section 3.4, we have also shown that alternative constellations can reduce BER and increase transmission range without upgrading to a PA with larger peak output power.

4.2 Future Research

There still remain interesting questions in digital baseband design.

The improvement in ACPR is small at very large output power (very small IBO) because of the saturation of current DACs during iterations as seen in Section 3.3.3. A Doherty PA tries to solve the dynamic range problem by enabling additional PAs only when input power gets larger. An interesting direction of future research would be compensation of Doherty PAs with no power back-off.

Our extraction of predistorter model parameters in Section 3.3.3 is off-line, and our compensation of the transmitter in Section 3.3.5 is open-loop. Process variation can be taken care of by calibration, but power supply voltage variations and temperature variations need to be compensated for with closed-loop adaptation. An efficient hardware implementation of real-time least-squares extraction of predistorter model parameters would enable closed-loop transmitter compensation with adaption. Given
the complexity of least-squares extraction of parameters, there is room to explore ways to implement parameters extraction and adaption efficiently. An important question and direction of future research would be how to reduce the oversampling bandwidth requirement in the feedback path.

We have been compensating for nonlinearity solely on the transmitter side, and thus have designed predistorters. Taking a step back, we could optimize the system linearity and BER performance of the whole communication system, with transmitter and receiver considered together. One hypothetical scenario could be that a predistorter, with less complexity to save power on the transmitter side, is designed to improve ACPR to satisfy spectral requirements, but does not sufficiently improve EVM enough, while a co-optimized postdistorter on the receiver side is designed to improve EVM further to meet the final BER requirement. It could be worthwhile to consider offloading some tasks from the transmitter side to the receiver side, rather than restricting compensation effort solely to the transmitter side. With co-optimization of the predistorter on the transmitter side and the postdistorter on the receiver side, it might be possible to achieve better energy efficiency for the whole communication chain.

Design of energy-efficient transmitter constellations, in the higher PAPR realm, needs to be further explored in the context of optimization together with the receiver. For the entire discussion on symbol constellation design in Section 3.4, for simplicity, we have assumed that the magnitude of the increase from the PAPR on symbols to the PAPR on samples is constant, with the increase depending only on the oversampling ratio and the characteristics of the shaping filter. With that assumption, we optimized the constellation to reduce the PAPR on symbols in order to reduce the PAPR on samples. That assumption needs to be challenged, and the magnitude of the increase from symbol PAPR to sample PAPR on constellation should be studied. The assumption that all symbols in a constellation are equally probable could also be replaced with the statistical distribution from real data. More development can be carried out on the optimization for PAPR of shaping filters for a given input symbol distribution. For the special case of an outphasing transmitter, co-optimization of
constellation and shaping filter can be carried out in the context of zero-avoidance, as an extension of Section 3.2.1. Once we have optimized the constellation and the filter for given input signal statistics, we could leverage our knowledge of the filter properties, and potentially implement the shaping filter in hardware in an energy-efficient way.
Bibliography


