

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Design Techniques for High Data Rates in Microwave and Millimeter-Wave Transmitters /

Permalink

<https://escholarship.org/uc/item/1ck1f76s>

Author

Dabag, Hayg-Taniel

Publication Date

2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

**Design Techniques for High Data Rates in Microwave and
Millimeter-Wave Transmitters**

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering (Electronic Circuits and Systems)

by

Hayg-Taniel Dabag

Committee in charge:

Peter M. Asbeck, Chair
James F. Buckwalter, Co-Chair
Prasad S. Gudem, Co-Chair
Robert R. Bitmead
William S. Hodgkiss
Bhaskar D. Rao

2014

Copyright
Hayg-Taniel Dabag, 2014
All rights reserved.

The dissertation of Hayg-Taniel Dabag is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Co-Chair

Chair

University of California, San Diego

2014

DEDICATION

To my parents.

TABLE OF CONTENTS

Signature Page	iii
Dedication	iv
Table of Contents	v
List of Figures	viii
List of Tables	xi
Acknowledgements	xii
Vita	xiv
Abstract of the Dissertation	xvi
Chapter 1	Introduction	1
	1.1 Spectral Efficiency of Digital Modulation	2
	1.2 Increasing the Bandwidth	4
	1.3 Challenges in mm-Wave CMOS PA Design	9
	1.4 Scope of the Dissertation	12
	1.5 Dissertation Organization	13
Chapter 2	Receiver Desensitization in Uplink Carrier Aggregation Due to Mixing of Two Transmit Signals in Cellular Handsets	15
	2.1 Background: Cellular Transceivers for Uplink Carrier Ag- gregation	15
	2.2 Prior Application of Noise/Distortion Cancellation Algo- rithms	19
	2.2.1 Principle of Adaptive Noise Canceller	19
	2.3 Proposed Cancellation Algorithm for UL CA Handsets	21
	2.3.1 Interference Estimation and Single-Input Single- Output Adaptive Distortion Canceller	22
	2.3.2 Multiple-Input Single-Output Adaptive Distortion Canceller	27
	2.3.3 Multiple Nonlinear Components	31
	2.3.4 Cancellation in the Presence of Desired RX Signal	31
	2.3.5 Cancellation in the Presence of Adjacent Channel Jammers	32
	2.4 Simulation Results	34
	2.4.1 Filter Adaptation With Perfect Time Alignment	35
	2.4.2 Filter Adaptation With Time Alignment Errors	36

2.5	Experimental Results	38
2.5.1	Measurement Setup	38
2.5.2	Cancellation Results	40
2.6	Rate of Convergence	49
2.7	Computational Effort	49
2.8	Conclusions	52
Chapter 3	Analysis and Design of Stacked-FET Millimeter-Wave Power Amplifiers	54
3.1	Background: mm-wave Silicon PAs	54
3.2	FET Stacking Concept	57
3.2.1	Prior Work on Stacked-FET PAs	57
3.2.2	Sizing of the Gate Capacitance C_k	60
3.2.3	Voltage Distribution	63
3.2.4	Benefits and Limitations of Stacking	64
3.2.5	Comparison of Stacking to Other Power Combining Techniques	66
3.3	Complex Intermediate Node Matching	68
3.3.1	Optimal Complex Intermediate Node Impedance	69
3.3.2	Optimal Intermediate Node Impedance Matching	73
3.3.3	Verification of Intermediate Node Matching Analysis	76
3.3.4	Comparison of Intermediate Node Matching Techniques	78
3.4	Technology and Amplifier Implementation	79
3.4.1	45-nm CMOS SOI Technology	79
3.4.2	PA Implementation	79
3.5	Experimental Results	85
3.5.1	Measurement Setups	85
3.5.2	Intermediate Node Matching	85
3.5.3	Comparing 2-, 3-, and 4-Stack PAs	87
3.6	Conclusions	91
3.7	Appendix 3.A: Optimal Drain Impedance	93
3.8	Appendix 3.B: Stacking Efficiency	95
Chapter 4	High Data Rate mm-Wave Wireless Transmission	97
4.1	Mark E Predistortion System	98
4.1.1	DPD Algorithms	101
4.1.2	M-QAM Test Signals	104
4.1.3	Predistortion of Mark E “Through” Test	106
4.1.4	System Accuracy Limits	109
4.2	Spatially Power Combined stacked-FET PAs	112

4.3	DPD Results of Spatially Power Combined stacked-FET PAs	116
4.4	Conclusions	125
Chapter 5	Conclusions and Future Work	127
5.1	Dissertation Summary	127
5.2	Future Work	129
5.2.1	Two UL CA and Three DL CA	129
5.2.2	Silicon mm-Wave Transmitters	131
Bibliography	133

LIST OF FIGURES

Figure 1.1: 64-QAM constellation for various SNRs	3
Figure 1.2: Signal quality requirements for correct data reconstruction . . .	4
Figure 1.3: 2011 US frequency allocations chart illustrates the heavy fragmentation	8
Figure 1.4: Possible frequency plan for carrier aggregation.	9
Figure 1.5: 2011 ITRS roadmap for f_T and f_{max} [5]	10
Figure 1.6: Nonlinear PA distorting input signal.	12
Figure 2.1: Block diagram of two transmitter system for UL CA.	16
Figure 2.2: Frequency view of third-order cross-modulation product (CM3) created by band 5 and 13 transmit signals across antenna switch.	17
Figure 2.3: The adaptive noise cancelling concept [12].	20
Figure 2.4: Block diagram of a transceiver using polar modulation for transmission [14].	21
Figure 2.5: Block diagram of adaptive distortion canceller.	22
Figure 2.6: Covariance of measured $cm3(n)$ and estimated $cm3'(n)$	24
Figure 2.7: Peak of covariance from Fig. 2.6 versus the group-delay difference between TX1" and TX2'.	25
Figure 2.8: Block diagram of SISO adaptive distortion canceller.	26
Figure 2.9: Block diagram of MISO adaptive distortion canceller.	28
Figure 2.10: Amount of cancellation given λ for various received signal powers.	33
Figure 2.11: Adaptive MISO filter with digital channel select filter for adjacent channel jammer suppression.	34
Figure 2.12: Spectral plots of the simulated distortion before and after cancellation using the SISO and the MISO filter.	36
Figure 2.13: Peak of covariance of $cm3_{test}$ and for various K	37
Figure 2.14: Simulated distortion before and after the MISO canceller using correct and incorrect group delay adjustment of K	38
Figure 2.15: Cancellation performance for different filter lengths when K is underestimated ($K = 3$) for the MISO canceller.	39
Figure 2.16: Sensitivity of MISO and SISO filter to errors in time alignment.	39
Figure 2.17: Measurement setup mimicking an UL CA handset.	41
Figure 2.18: Measured duplexer distortion before and after cancellation using MISO or SISO filter.	42
Figure 2.19: Measured switch and duplexer distortion before and after cancellation using either the MISO or the SISO filter.	43
Figure 2.20: Measured switch and duplexer distortion before and after cancellation using the SISO or the MISO filter with K alignment error by minus one sample.	43
Figure 2.21: Low-power received signal with distortion before and after cancellation (SINR before/after cancellation \approx -10 dB/ +9.4 dB).	44

Figure 2.22: High-power received signal with distortion before and after cancellation (SINR before/after cancellation ≈ 20 dB/32 dB). . . .	45
Figure 2.23: Captured signal with out-of-band jammer before and after channel select filtering and after adaptive distortion cancellation. . .	47
Figure 2.24: Captured CM3 in experiment using different antennas. Reflections change distortion shape. SISO filter length is 4. MISO filter length is 16.	48
Figure 2.25: Convergence of MISO filter output for various λ	50
Figure 3.1: 3-stack PA schematic. The rectangular boxes used in the input and output matching network are coplanar waveguides (CPWs).	58
Figure 3.2: Hittite high power amplifier [25]	59
Figure 3.3: Prior stacked-FET PAs	60
Figure 3.4: $C_k/C_{gs,k}$ for various $C_{gd,k}/C_{gs,k}$ for $g_{m,k} \cdot R_{opt} = 3$	63
Figure 3.5: Incremental increase in P_{sat} of k th stacked FET.	65
Figure 3.6: Comparison of f_{MAX} of two, three, and four stacked FETs using thin-oxide FETs and two stacked thick oxide / high-voltage (HV) FETs.	68
Figure 3.7: Incremental increase in P_{sat} of the k th Wilkinson combiner. . .	69
Figure 3.8: Simplified small-signal model of stacked transistors.	70
Figure 3.9: Cumulative stacking efficiency for various phase misalignments.	72
Figure 3.10: 2-stack PA schematic with different intermediate node tuning techniques.	73
Figure 3.11: PAE (a) and P_{out} (b) for $P_{in} = 9$ dBm using series L , shunt L , and shunt-feedback C_{ds} intermediate node tuning.	77
Figure 3.12: Schematic of 2-stack PA with shunt tuning elements between the two transistors.	80
Figure 3.13: Schematic of 3-stack PA with shunt tuning element between $M1$ and $M2$ and series tuning inductance between $M2$ and $M3$. . .	81
Figure 3.14: Schematic of 4-stack PA with shunt tuning element between $M1$ and $M2$ and series tuning inductance between $M2$ and $M3$. . .	82
Figure 3.15: $50\text{-}\Omega$ load and pad capacitance are transformed by a shunt stub (solid line) to a load impedance for optimal PAE inside the highlighted region.	83
Figure 3.16: Photomicrograph of 3-stack PA occupying $0.6\text{ mm} \times 0.5\text{ mm}$ including pads.	83
Figure 3.17: Simulated drain voltages (a) and drain currents (b) of 2-stack PA from Fig. 3.12 without $CPW2$	84
Figure 3.18: Large-signal measurement setup.	85
Figure 3.19: Measured gain and PAE as a function of output power at 46 GHz for the 2-stack PA with two shunt CPW s, with one shunt CPW , and no shunt CPW biased at: $V_{G,1} \sim 0.2\text{ V}$, $V_{G,2} = 1.8\text{ V}$, $V_{DD} = 2.8\text{ V}$, $I_{DC} = 8\text{ mA}$	86

Figure 3.20: Measured PAE and P_{sat} over frequency for 2-stack PA with two shunt CPW s, with one shunt CPW and no shunt CPW	87
Figure 3.21: Measured S-parameter for 2-stack, 3-stack, 4-stack PA; 2-stack: $V_{G,1}=0.3$ V, $V_{G,2}=1.6$ V, $V_{DD}=2.5$ V; 3-stack: $V_{G,1}=0.2$ V, $V_{G,2}=1.7$ V, $V_{G,3}=2.5$ V, $V_{DD}=3.5$ V; 4-stack: $V_{G,1}=0.3$ V, $V_{G,2}=1.7$ V, $V_{G,3}=2.7$ V, $V_{G,4}=4$ V, $V_{DD}=5$ V.	88
Figure 3.22: Measured gain and PAE versus P_{out} for 2- and 3-stack PA at 46 GHz and 4-stack PA at 41 GHz for class-AB bias	89
Figure 3.23: P_{out} versus number of stacked transistors.	90
Figure 3.24: Measured peak PAE and P_{sat} versus frequency for 2-, 3-, 4-stack PA.	91
Figure 4.1: Simplified block diagram of mm-wave predistortion system . . .	100
Figure 4.2: Modeling and inverse modeling of PA for DPD	101
Figure 4.3: Inverse modeling of PA using the MM signal as primary input .	104
Figure 4.4: Spectrum of M-QAM signal after RRC filtering with different α	105
Figure 4.5: Evaluation of linearity and memory of the Mark E system in “through” test; P_{out} de-embedded to Quinstar downconverter RF input.	108
Figure 4.6: Spectral response of system “through” test after the mm-wave driver and after the mm-wave downconverter.	110
Figure 4.7: Diagram of stacked-FET PA array with differential patch antennas [50]	113
Figure 4.8: Picture of antenna assembly around the PCB with 2x2 antenna array.	114
Figure 4.9: Measured EIRP at main antenna and estimated P_{out} vs. I_{dc} of eight 4-stack PAs for CW excitation	115
Figure 4.10: Comparison of MP and RGMP model to match MM DPD PA input signal	117
Figure 4.11: PA array output before and after DPD for 49-MS/s, 1024-QAM signal	121
Figure 4.12: PA array output before and after DPD for 82-MS/s, 1024-QAM signal	122
Figure 4.13: PA array output before and after DPD for 98-MS/s, 1024-QAM signal	123
Figure 4.14: PA array output before and after DPD for 98-MS/s, 256-QAM signal	124
Figure 4.15: EVM and ACPR vs. EIRP	126
Figure 5.1: 2 UL and 3 DL CA with CM2 desensing on of the receivers . .	130

LIST OF TABLES

Table 1.1: LTE FDD Frequency Bands September 2012 [3]	6
Table 1.2: LTE FDD Frequency Bands September 2012 Continued [3]	7
Table 2.1: Cancellation Performance of MISO Filter	46
Table 2.2: Order of Complexity of SISO and MISO Filter	52
Table 3.1: Evaluation of Load Impedances in Stacked-FET PA	62
Table 3.2: Reactive Intermediate Node Tuning	76
Table 3.3: Comparison To Previously Reported Silicon mm-Wave PAs	92
Table 4.1: Specifications of used modulated signals	106
Table 4.2: Summary of NRMSE of CW signal after digital filtering with various filter corner frequencies	111
Table 4.3: Summary of NRMSE/EVM with MM and RGMP DPD for var- ious symbol rates at a power of -3.5 dBm at the output of the mm-wave driver	112
Table 4.4: Summary of NRMSE of CW signal without digital filtering for various output powers of the PA array	116
Table 4.5: Summary of ACPR and EVM with and without DPD	118
Table 5.1: B3 and B8 TX power at different components of B7 receiver and resulting CM2 power	131

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Peter Asbeck, for his support and guidance throughout my graduate studies. I have been very fortunate to have him as my teacher and mentor during my years at UCSD. His technical expertise and guidance were tremendously helpful in achieving my research goals. His endless patience was truly an inspiration and is something I strive to achieve one day for myself. I cannot express my gratitude enough for all his support, and inspiration. Thank you Dr. Asbeck.

I am also exceedingly grateful to my co-advisors Dr. James Buckwalter and Dr. Prasad Gudem. Their assistance and encouragement were tremendously valued especially in regards to the ELASTx project and my distortion cancellation project.

I would also like to thank my dissertation committee: Dr. Bhaskar Rao, Dr. William Hodgkiss, and Dr. Robert Bitmead for providing precious and insightful feedback.

A special acknowledgment is needed for Dr. Byunghoo Jung and Dr. Dongwon Seo whos unwavering support during my undergraduate studies paved the way for my internships at Qualcomm and my doctorate at UCSD.

To my classmates, labmates, and countless others I have had the pleasure to work at UCSD, I would like to thank you all for being part of this journey. In particular Dr. Joohwa Kim for his invaluable training during my early years at the program and my current labmates Bassel Hanafi, Hamed Gheidi, Paul Draxler, Johana Yan, Sataporn Pornpornlikit, for their friendship and support.

On a personal note, I would like to thank my parents, Sona and Mihran, and my brother Tigran. I cannot express enough love for their unconditional support, love, and guidance. Their work ethic, determination, and loving spirit is something

I strive to imitate every day.

Lastly, I would like to include a special thanks to my friends in San Diego and Bochum. I will forever cherish the good memories we created over this long and arduous journey. Greg, Willie, Thanh, Chris, Marcel, Khaled, and everyone else who helped me relax, thanks!

The material in this dissertation is based on the following papers. Chapter 2 is mostly a reprint of the material as it appears in “All-Digital Cancellation Technique to Mitigate Receiver Desensitization in Uplink Carrier Aggregation in Cellular Handsets”, *Transactions on Microwave Theory and Techniques*, Dec. 2013.

Chapter 3 is mostly a reprint of the material as it appears in “Analysis and Design of Stacked-FET Millimeter-Wave Power Amplifier”, *Transactions on Microwave Theory and Techniques*, Apr. 2013.

Section 4.2 is in part a reprint of the material as it appears in ‘A CMOS 45 GHz Power Amplifier with Output Power >600 mW Using Spatial Power Combining’, accepted to *2014 IEEE MTT-S International Microwave Symposium (IMS)*.

The material in Section 4.3 will in part be used for a publication in preparation with the working title “Digital Predistortion for 1024-QAM of Millimeter-wave, Free-space-combined Stacked-FET PAs”.

The dissertation author was primary or collaborating author of these materials, and co-authors have approved the use of the material for this dissertation.

VITA

2008	Dipl. Ing. in Electrical Engineering, Ruhr University Bochum, Germany
2011	M.S. in Electrical Engineering, University of California, San Diego
2009-2014	Graduate Research Assistant, University of California, San Diego
2014	Ph. D. in Electrical Engineering, University of California, San Diego

PUBLICATIONS

H.-T. Dabag, H. Gheidi, S. Farsi, P. Gudem, and P. M. Asbeck, "All-Digital Cancellation Technique To Mitigate Receiver Desensitization in Uplink Carrier Aggregation in Cellular Handsets", *Transactions on Microwave Theory and Techniques*, vol. 61, no. 12, pp. 4754-4765, Dec 2013.

A. Agah, H.-T. Dabag, B. Hanafi, P. M. Asbeck, J. F. Buckwalter and L. E. Larson "Active Millimeter-Wave Phase-Shift Doherty Power Amplifier in 45-nm SOI CMOS", *IEEE Journal of Solid-State Circuits* vol. 48, no. 10, pp. 2338-2350, Oct 2013.

J. Jayamon, A. Agah, B. Hanafi, H. Dabag, J. Buckwalter, and P. Asbeck, "A W-band stacked FET power amplifier with 17 dBm Psat in 45-nm SOI CMOS", *IEEE 13th Topical Meeting on RF Systems (SiRF)*, 2013.

H.-T. Dabag, H. Gheidi, P. Gudem, and P. M. Asbeck, "All-Digital Cancellation Technique To Mitigate Self-Jamming In Uplink Carrier Aggregation in Cellular Handsets", *IEEE International Microwave Symposium*, 2013.

A. Agah, H. Dabag, P. Asbeck, L. Larson, and J. Buckwalter, "High-speed, High-efficiency millimeter-wave transmitters at 45 GHz in CMOS", *IEEE International Microwave Symposium*, 2013.

H.-T. Dabag, B. Hanafi, F. Golcuk, A. Agah, J. F. Buckwalter, and P. M. Asbeck, "Analysis and Design of Stacked-FET Millimeter-Wave Power Amplifier", *Transactions on Microwave Theory and Techniques*, vol. 61, no. 4, pp. 1543-1556, Apr 2013.

- J. Kim, H. Dabag, P. Asbeck, and J. F. Buckwalter, “Q- and W-band Power Amplifiers in 45-nm SOI CMOS Technology”, *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 12, pp. 1870-1877, June 2012.
- H.-T. Dabag, P. M. Asbeck, and J. F. Buckwalter, “Linear operation of high-power millimeter-wave stacked-FET PAs in CMOS SOI”, *IEEE International Midwest Symposium on Circuits and Systems*, 2012.
- A. Agah, B. Hanafi, H. Dabag, P. Asbeck, L. Larson, and, J. Buckwalter, “A 45GHz Doherty power amplifier with 23% PAE and 18 dBm output power, in 45nm SOI CMOS”, *IEEE International Microwave Symposium*, 2012.
- A. Agah, H. Dabag, B. Hanafi, P. Asbeck, J. Buckwalter, and L. Larson, “A 34% PAE, 18.6dBm 42-45 GHz stacked power amplifier in 45nm SOI CMOS”, *IEEE International Microwave Symposium*, 2012.
- S. Pornpromlikit, H.-T. Dabag, B. Hanafi, J. Kim, L. E. Larson, J. F. Buckwalter, and P. M. Asbeck, “A Q-Band Amplifier Implemented with Stacked 45 nm CMOS FETs”, *IEEE Compound Semiconductor IC Symposium*, 2011.
- H.-T. Dabag, J. Kim, L. E. Larson, J. F. Buckwalter and P. M. Asbeck, “A 45-GHz SiGe HBT Amplifier with Above 25% Efficiency and 30 mW Output Power”, *IEEE Bipolar / BiCMOS Circuits and Technology Meeting*, 2011.
- D. Seo, H. Dabag, Y. Guo, M. Mishra, and G. McAllister, “High-Voltage-Tolerant Analog Circuits Design in Deep-Submicrometer CMOS Technologies”, *IEEE Transactions on Circuits and Systems I*, vol. 54, no. 10, pp. 2159-2166, Oct 2007.
- H. Dabag, D. Seo, M. Mishra and J. Hausner, “Electrical Stress-free High Gain and High Swing Analog Buffer Using an Adaptive Biasing Scheme”, *IEEE International Symposium on Circuits and Systems*, 2007.

ABSTRACT OF THE DISSERTATION

**Design Techniques for High Data Rates in Microwave and
Millimeter-Wave Transmitters**

by

Hayg-Taniel Dabag

Doctor of Philosophy in Electrical Engineering (Electronic Circuits and Systems)

University of California, San Diego, 2014

Peter M. Asbeck, Chair
James F. Buckwalter, Co-Chair
Prasad S. Gudem, Co-Chair

In the quest to increase channel bandwidths in wireless communication systems, two important trends are to move towards wider continuous bands at mm-wave frequencies and to aggregate smaller bands at cellular frequencies. In this dissertation a few of the challenges and possible circuit and DSP solutions for efficient high data rate communication using these techniques are described.

First, an issue relating to cellular uplink carrier aggregation is discussed and a DSP based solution developed. Second, the design of a broad band CMOS PA for

mm-wave applications is presented. Third, the design of a mm-wave predistortion system and its use to predistort an array of mm-wave CMOS SOI PAs is described.

In the near term, cellular carriers plan on employing carrier aggregation to increase data rates. This can lead to significant receiver desensitization for a number of LTE band combinations, because of the cross-modulation products created by the nonlinearity of RF front-end components. To mitigate this effect, an all-digital cancellation algorithm is proposed in this thesis that cancelled the cross-modulation product and improved the signal-to-interference-plus-noise ratio (SINR) and error-vector-magnitude (EVM) of the desired received signal by up to 20 dB.

In the second part of the dissertation, the possibility of using mm-wave CMOS PAs for wideband communication is described. The design of CMOS stacked-FET PAs with an emphasis on appropriate complex impedances between the transistors is presented. The stacking of multiple FETs enables the use of higher supply voltages, which in turn allows higher output power and a broader bandwidth output matching network. A 4-stack amplifier design that achieves a saturated output power greater than 21 dBm while achieving a maximum power-added-efficiency (PAE) greater than 20% from 38 GHz to 47 GHz is reported.

Finally, the thesis describes predistortion of an array of stacked-FET PAs after spatial power combining. Predistortion improved the signal quality to a high level, which allowed the use of complex modulation schemes, which in turn allows high data rates in a spectrally efficient manner. After predistortion a 100-MHz wide, 1024-QAM signal was demodulated with an EVM of 1.3%, which corresponds to a data rate of 1 Gb/s.

Chapter 1

Introduction

It is just about 40 years ago that the first cellular call from a prototype mobile phone was made and about 30 years when the first commercial mobile phones became available. The initial phones were heavy, bulky, and their talk time was very short. After many iterations of improvements, cellular phones became smaller, had acceptable battery life, and became sufficient for (but limited to) voice calls and text messaging.

More recently, the simultaneous increase in available data rates for wired Internet connections enabled a variety of new Internet based multimedia and business services. The user demand to enjoy these contents in wireless fashion keeps driving the the demand for higher data rates for wireless communication systems. To fill this need, various techniques are under development. The attempted solutions generally center around two approaches: increasing the available bandwidth and increasing the number of bits transmitted in a given bandwidth, or a combination of the two.

In this chapter, a brief overview is provided to review some of the challenges on the path to higher wireless data rates. First, a brief review is given to explain

the limitation on obtainable data rates per occupied bandwidth. Second, techniques are discussed to increase the available bandwidth. Third, implementation difficulties focusing on the transmitter, in particular the power amplifier (PA), are described. The fourth and fifth section describe the scope and structure of the dissertation.

1.1 Spectral Efficiency of Digital Modulation

Even though our transmitters radiate signals at gigahertz frequencies the actual information transmitted generally only occupies multiple megahertz of unique user data. The method how the bits are encoded on the carrier in a bandwidth limited signal is called modulation. Earlier wireless communication standards used analog modulation schemes, which have been replaced by their digital counterparts. There exists a wide variety of modulation schemes with different tradeoffs. This section does not try to give a complete overview, but just enough background information to explain the benefits of more complex modulations and the practical challenges they pose. M-ary quadrature amplitude modulation (M-QAM) in some form is the basis of many modern communication standards. Fig. 1.1 shows a 64-QAM constellation with 64 unique symbols. In this constellation $\log_2(M)$ i.e. 6 bits per symbol can be encoded. Fig. 1.1(a) and Fig. 1.1(b) respectively show an example of received modulations (in I and Q plane) for signal to noise ratio (SNR) of 30 dB and 20 dB. From the figures it is apparent that correct assignment of each received symbol to its target is dependent on the SNR. For example the constellation shown in Fig. 1.1(a) allows error free reception and the signal shown in Fig. 1.1(b) will have a significant number of errors.

Fig. 1.2(a) plots the required SNR for M-QAM for a targeted bit error rate

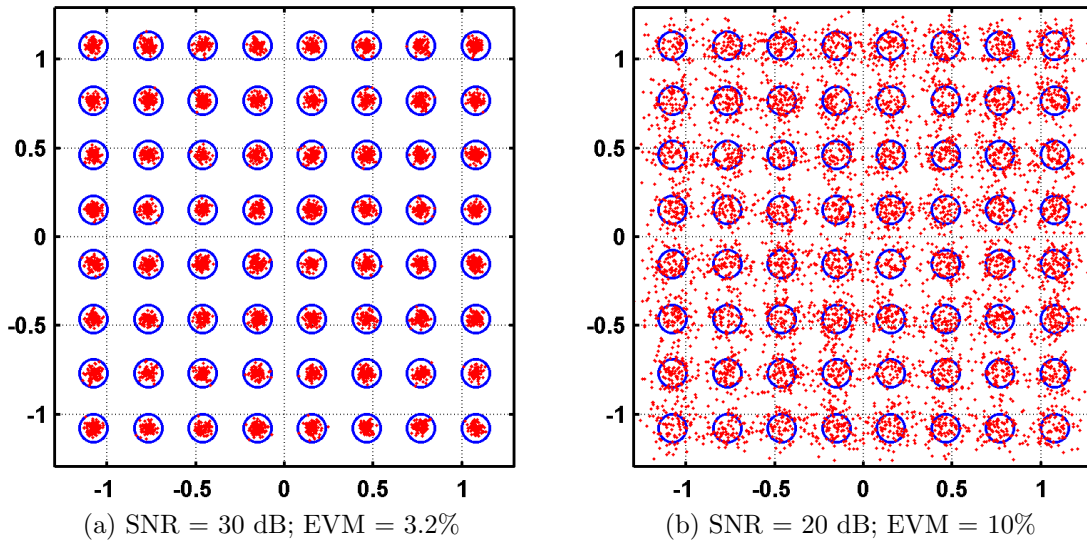


Figure 1.1: 64-QAM constellation for various SNRs

(BER), assuming the only disturbance is white Gaussian noise. A BER between 10^{-3} and 10^{-6} are frequently targeted in wireless systems [1]. An alternative representation to SNR is the error vector magnitude (EVM). In the constellation diagram it represents the normalized power difference between the actual received symbol and the ideal symbol. The peak EVM and root-mean-square (rms) average are often specified for a received signal. With (1.1) one can approximate the average EVM for a given SNR. Fig. 1.2(b) plots the BER versus EVM. One can see that for higher order modulation schemes the SNR and EVM requirements increase significantly even though the achieved increase in data rate is not as high.

$$EVM(\%) \approx 10^{-SNR(dB)/20} \cdot 100 \quad (1.1)$$

Another factor and the key motivator to go to higher order modulations is the required signal bandwidth. It is important to note that the bandwidth is independent of the modulation order. This leads to the concept of spectral efficiency, which is defined as the number of transmitted bits per bandwidth. Since higher

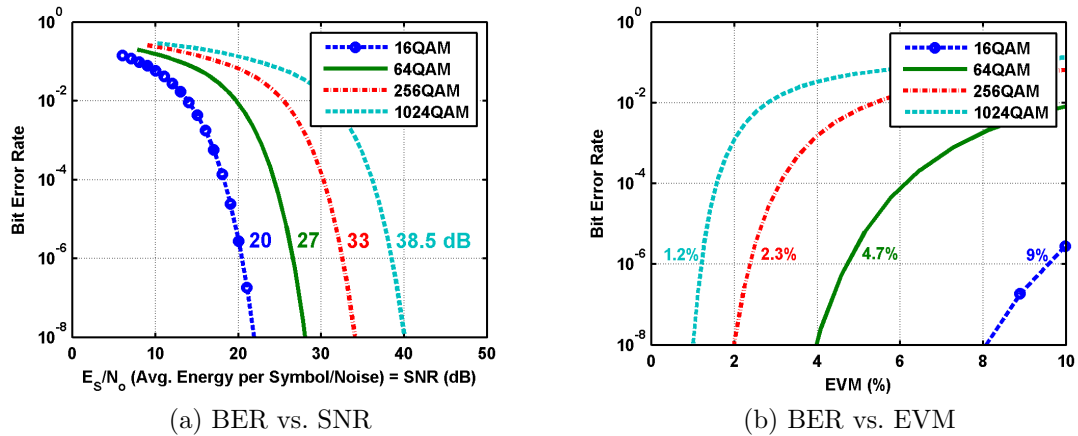


Figure 1.2: Signal quality requirements for correct data reconstruction; highlighted are the average SNR / EVM for a BER of 10^{-6} .

order modulation occupy the same amount of bandwidth the spectral efficiency increases with increased modulation order. However, the higher modulation order require higher SNR i.e. better EVM for correct demodulation. Achieving the good SNR/EVM over wireless links poses significant implementation challenges in particular in the power amplifier (PA). Some of these challenges are discussed in Section 1.3.

Other techniques such as multiple-input and multiple-output (MIMO) antenna arrays are being developed to further increase the spectral efficiency and orthogonal frequency-division multiplexing (OFDM) modulation is used to increase the effectively achievable data rates. The solutions in the dissertation do not go into issues relating to these techniques, but the achieved results can be used in conjunctions with MIMO and OFDM systems.

1.2 Increasing the Bandwidth

Since increasing the spectral efficiency is becoming increasingly more difficult and the increase in data rates are modest by further increase of the modulation

order, a lot of attention is focused on increasing the bandwidths used. In cellular system this is made difficult due to the scarcity of available bandwidth and its fragmentation as shown in Fig. 1.3. Therefore, the upcoming LTE-A standard allows aggregating up to five 20-MHz LTE channels [2]. Fig. 1.4 shows the four possible frequency plans of carrier aggregation. The simplest two would aggregate multiple continuous or non-continuous channels in a single band. The third combines continuous or non-continuous channels across multiple adjacent bands. The fourth combines multiple channels across non-continuous bands. Unfortunately, most bands are significantly smaller than 100 MHz as listed in Table 1.1 and Table 1.2. This regulatory restriction virtually forces the need for inter band carrier aggregation (CA) in cellular systems. Even if cost and power inefficiency in a simplistic implementation of CA with two parallel transceivers is accepted, CA of certain band pairs can cause problems. One specific issue in uplink CA (UL CA) and a potential solution is discussed in Chapter 2 of this dissertation.

Table 1.1: LTE FDD Frequency Bands September 2012 [3]

LTE FDD Band	Uplink Frequency (MHz)	Downlink Frequency (MHz)	Width (MHz)
1	1920 - 1980	2110 - 2170	60
2	1850 - 1910	1930 - 1990	60
3	1710 - 1785	1805 - 1880	75
4	1710 - 1755	2110 - 2155	45
5	824 - 849	869 - 894	25
6	830 - 840	865 - 875	10
7	2500 - 2570	2620 - 2690	70
8	880 - 915	925 - 960	35
9	1749.9 - 1784.9	1844.9 - 1879.9	35
10	1710 - 1770	2110 - 2170	60
11	1427.9 - 1447.9	1475.9 - 1495.9	20

Table 1.2: LTE FDD Frequency Bands September 2012 Continued [3]

LTE FDD Band	Uplink Frequency (MHz)	Downlink Frequency (MHz)	Width (MHz)
12	698 - 716	728 - 746	18
13	777 - 787	746 - 756	10
14	788 - 798	758 - 768	10
15 ¹	1900 - 1920	2600 - 2620	12
16 ¹	2010 - 2025	2585 - 2600	15
17	704 - 716	734 - 746	12
18	815 - 830	860 - 875	15
19	830 - 845	875 - 890	15
20	832 - 862	791 - 821	30
21	1447.9 - 1462.9	1495.9 - 1510.9	15
22	3410 - 3490	3510 - 3590	80
23	2000 - 2020	2180 - 2200	20
24	1626.5 - 1660.5	1525 - 1559	34
25	1850 - 1915	1930 - 1995	65
26	814 - 849	859 - 894	35
27	807 - 824	852 - 869	17
28	703 - 748	758 - 803	45

¹ Reserved

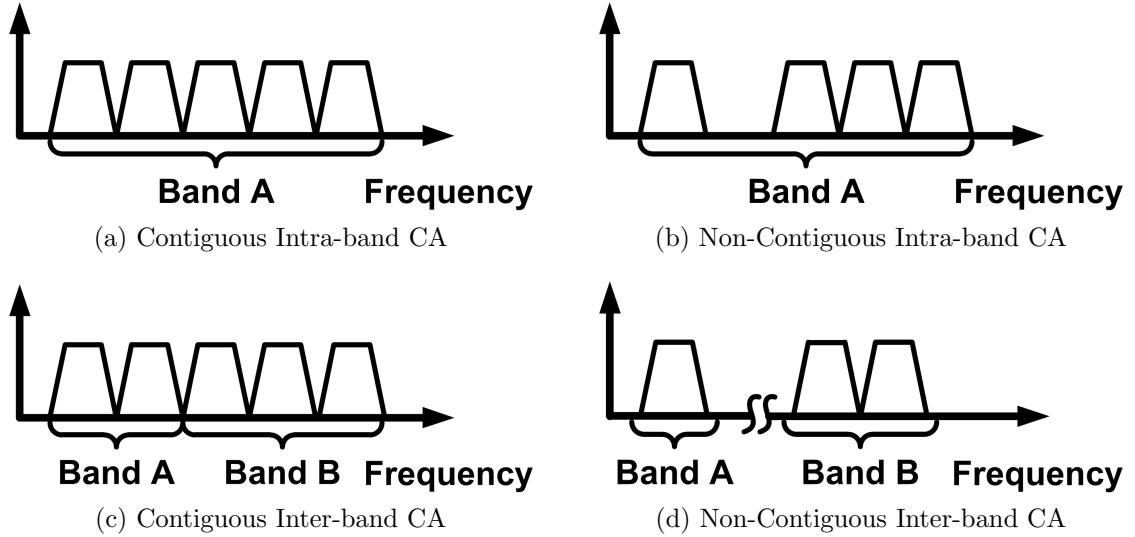


Figure 1.4: Possible frequency plan for carrier aggregation.

Even if all of the challenges posed by CA are solved, the maximum available bandwidth is still limited to 100 MHz. However, for satellite communication and wireless high resolution video transfer even higher data rates are required than available in the relatively low frequency bands around 1-5 GHz. Another effort focuses on cost effective solutions for mm-wave wireless system, in particular in the *Q* (33 - 50 GHz) and *V* (40 - 75 GHz) band. In those continuous channels of 1 GHz or larger are possible [4].

1.3 Challenges in mm-Wave CMOS PA Design

Due to the aggressive scaling of transistor sizes, reasonable gains are achievable in mm-wave bands. The unity current gain (f_T) and unity power gain (f_{max}) in today's CMOS process are in the order of 300 GHz and are predicted to continue increasing as illustrated in Fig. 1.5. This makes the use of CMOS transistors viable for mm-wave amplifiers.

Unfortunately, the decreasing feature sizes also reduce the voltage handling

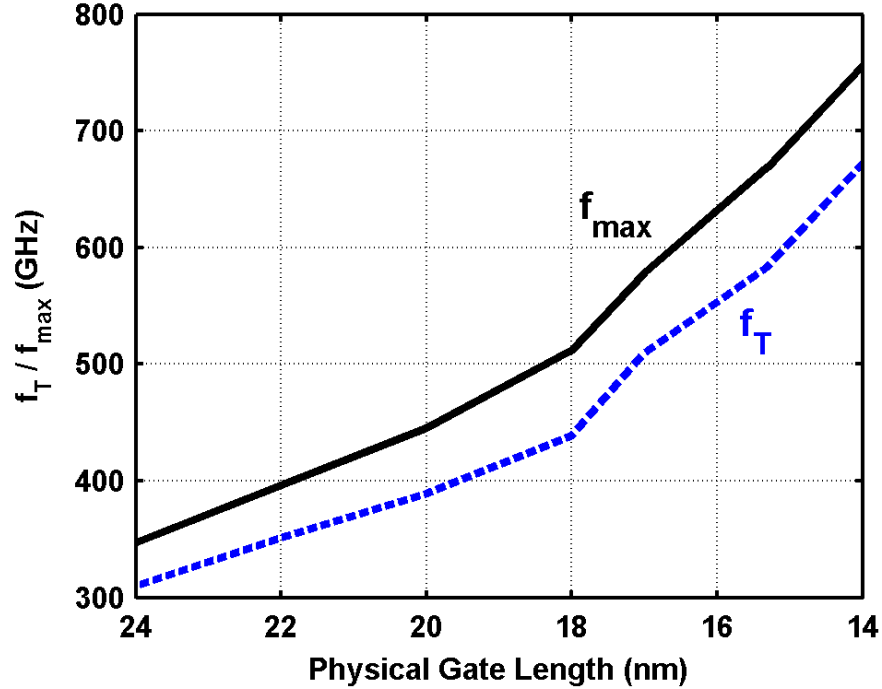


Figure 1.5: 2011 ITRS roadmap for f_T and f_{\max} [5]

capabilities of the transistor. There are several breakdown mechanisms in CMOS transistors such as the gate oxide breakdown, hot carrier degradation, and punch-through. To avoid these breakdown mechanisms, the drain-source, the gate-drain, and the gate-source voltage need to be kept below certain voltage levels. Unfortunately, those voltage limits decrease with the decreasing feature sizes of the transistor, which is particularly disadvantages for power amplifiers (PA).

In order to achieve high output powers one can either increase the output voltage swing and / or output current swing of the PA. The former is limited by the transistor breakdown voltages. The latter can be increased to some extend by increasing the transistor gate width. However, increasing the transistor width, and hence current, also increases the transistor capacitance, which decreases the transistor gain at high frequencies and makes the input matching more challenging.

In addition, increasing the transistor current also decreases the load line

impedance R_L as state in (1.2), where V_{max} is the maximum RF swing the device can tolerate and I_{max} is the maximum current the device can provide. Loading a PA with R_L ensures that it operates at its optimal efficiency and highest achievable power [6]. As mentioned before, an increase in current also decreases R_L , which makes the output matching network more challenging.

$$R_L = V_{max}/I_{max} \quad (1.2)$$

In summary, the reduced feature sizes for CMOS transistors are required to ensure sufficient gain at mm-wave frequencies. Unfortunately, this goes hand in hand with reduced voltage handling capability, which effectively reduces the output power and efficiency.

Since achieving high output powers at mm-wave is a challenge in itself, one should operate the PAs as close as possible to their saturated output power. Unfortunately, the linearity of amplifiers significantly degrades close to saturation. A PA in compression, as any nonlinear system, pass a distorted copy of its input signal to the output. Many of the distortion components land far way from the band of interest and can easily be filtered. However, some of the intermodulation products land very close to the band of interest. This is illustrated in Fig. 1.6. If 3-tones are fed to a PA, which only experiences a third order nonlinearity the output of the PA will generate the original 3-tones and additional tones in- and out-of-band. Fig. 1.6(b) shows the nonlinear PA output for a wideband input signal.

The in-band distortion corrupts the desired signal and reduces its signal to noise and interference ratio (SINR), which reduces the EVM in a similar relationship as the SNR in (1.1).

The out-of-band signal interferes with neighboring channels. The exact

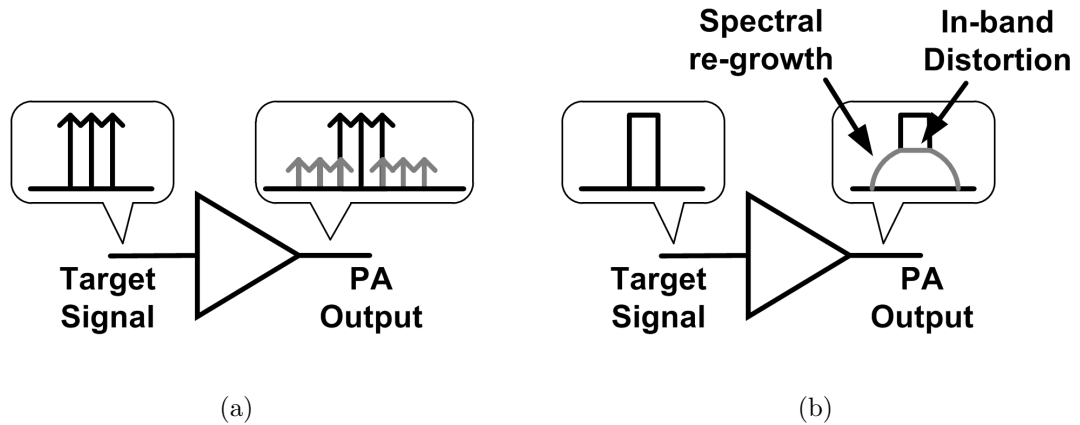


Figure 1.6: Nonlinear PA distorting input signal.

acceptable adjacent channel power leakage (ACP) depends on the used communication standard. But unless it is kept at reasonable low levels e.g. 33 dBm below the main signal, the achievable EVM of the adjacent channel is limited. Alternatively, one would have to place “guard bands” between two channels, which is disadvantageous due to the scarcity of available bandwidth.

1.4 Scope of the Dissertation

There are various challenges and potential solutions in the quest for higher data rates. This dissertation focuses on two aspects.

First, an immediate issue in an uplink CA (UL CA) transceiver is studied. The simultaneous data transmission in certain band pairs can cause a self-jamming of the receiver due to nonlinearity of the passive front-end. The nature of the problem allowed a DSP based solution, which has the advantage that cost effective implementations can be realized very quickly.

The second part of the dissertation focuses on the mm-wave communication system. There the availability of low cost and efficient PAs is a major challenge. This is addressed by adapting the “stacked-FET PA” architecture to mm-wave

operation, which resulted in record efficiencies and power for CMOS PAs at mm-waves. In addition digital predistortion is applied to an array of the “stacked-FET PAs” after spatial power combining, which allowed the use of high complexity signals for high spectral efficiency.

1.5 Dissertation Organization

In this chapter the critical issues for spectrally efficient high-speed microwave and millimeter wave wireless communication systems have been reviewed.

Chapter 2 considers cell phone transceivers suitable for uplink carrier aggregation (UL CA) to increase transmit data rates. UL CA can lead to significant receiver desensitization for a number of LTE band combinations, because of the cross-modulation products created by the nonlinearity of antenna switches and duplexers in the RF front-end. To mitigate this effect, an all-digital cancellation algorithm is proposed that relies solely on the digital representation of the signals, a peak covariance search for time alignment, and an adaptive distortion canceller.

Chapter 3 discusses stacked-FET CMOS mm-wave PAs with a focus on design of appropriate complex impedances between the transistors. The stacking of multiple FETs allows increasing the supply voltage, which in turn allows higher output power and a broader bandwidth output matching network. Different matching techniques for the intermediate nodes are analyzed and used in 2-, 3- and 4-stack single-stage Q -band CMOS power amplifiers (PAs).

In Chapter 4 a wideband digital predistortion system for mm-wave applications is described. With this system, an ensemble of stacked-FET PAs is predistorted after their output signals are spatially power combined. An auxiliary antenna is used to feed back part of the radiate signal to the digital predistortion

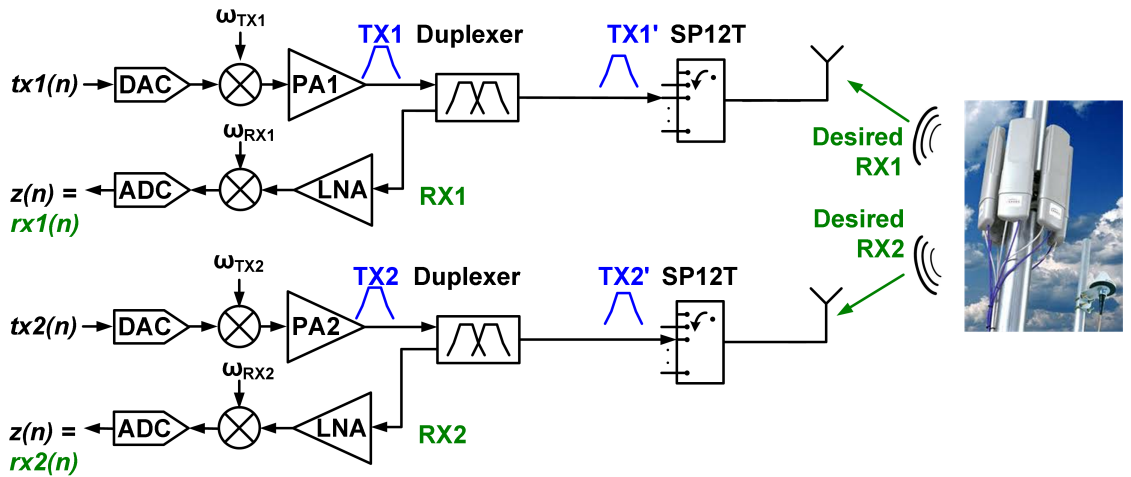
(DPD) system and the main lobe is monitored on a spectrum analyzer to confirm the effectiveness of the DPD.

Chapter 2

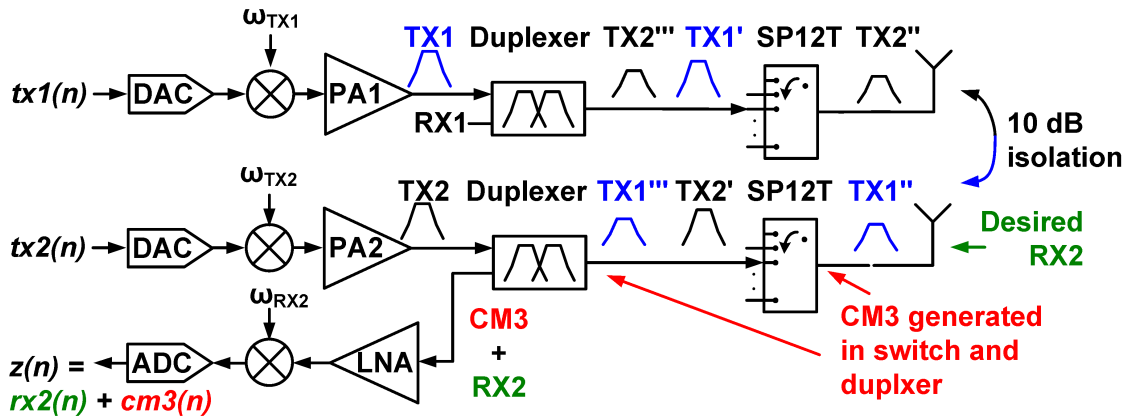
Receiver Desensitization in Uplink Carrier Aggregation Due to Mixing of Two Transmitted Signals in Cellular Handsets

2.1 Background: Cellular Transceivers for Uplink Carrier Aggregation

One approach to meet user demands for high data rates in handsets is the use of wider signal bandwidths. This cannot always be directly applied, since many carriers own noncontiguous spectra in various frequency bands. To enable the use of multiple bands, next-generation cellular standards support uplink carrier aggregation (UL CA) [2]. Fig. 2.1(a) shows a block diagram of a two transmitter system used for UL CA, which comprises two separate TX and RX chains as well as two antennas. Since modern cell phones need to cover a wide range of bands, each



(a) Ideal UL CA transceiver



(b) UL CA transceiver with cross modulation generated in the front-end

Figure 2.1: Block diagram of two transmitter system for UL CA.

antenna is preceded by a multiport switch, which allows transmitters for different bands to share the same antenna. As shown in Fig. 2.1(b), due to poor antenna isolation in a cell phone, a copy of the first transmitted waveform (TX1) will be received by the second transmitter (and vice versa).

If all system components were perfectly linear, the coupled version of TX1 would only act as an out-of-band jammer. However, nonlinear behavior of switches and duplexers creates cross-modulation distortion products that can land in the receive band for certain band pairings. One can calculate the center frequency of

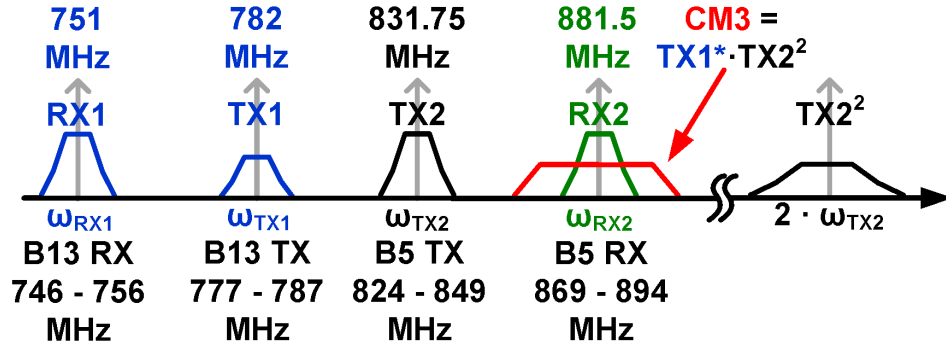


Figure 2.2: Frequency view of third-order cross-modulation product (CM3) created by band 5 and 13 transmit signals across antenna switch.

both third-order cross-modulation products (CM3s) using (2.1):

$$\omega_{CM3,1} = 2 \cdot \omega_{TX1} - \omega_{TX2}, \quad \omega_{CM3,2} = 2 \cdot \omega_{TX2} - \omega_{TX1} \quad (2.1)$$

where ω_x is the center frequency of that particular signal. Fig. 2.2 illustrates this for bands 5 (B5) and 13 (B13). For example, if ω_{TX1} is 782 MHz and ω_{TX2} is 831.75 MHz, one can calculate with (2.1) that $\omega_{CM3,2}$ is 881.5 MHz, which is inside of the B5 RX band. Other examples of problematic band pairs for UL CA are as follows: B2 and B4, and B8 and B20. The distortion power can be estimated by:

$$P_{CM3} = 2 \cdot P_{TX1} + (P_{TX1} - \text{path loss}) - 2 \cdot IIP3 \quad (2.2)$$

where P_x is the power of the particular signal in dBm, and IIP3 refers to the input-referred third-order intercept points of the component. State-of-the-art antenna switches and duplexers, respectively, have IIP3s of 70 dBm to 80 dBm [7,8]. When the TX power at each antenna is 24 dBm and considering 10-dB antenna isolation, the resulting CM3 power, according to (2.2), is -79 dBm, which is 28 dB above the thermal noise floor for a 5-MHz signal. Despite the high linearity of switches and duplexers, the distortion power is high enough to severely desensitize the

receiver. To overcome this, one could reduce the power of both transmitted signals by approximately 10 dB. However, this would severely compromise the link budget for UL CA. Improving the antenna isolation from 10 to 40 dB would also greatly mitigate the receiver desensitization. However, this is challenging due to the limited cavity size dictated by cell phone dimensions.

Alternatively, one could develop components with IIP3s greater than 85 dBm, which are currently not available. Even if such parts with the required linearity become available in the future, replacement of all required passive components with high linearity components would significantly increase the cost. Alternatively, one could take advantage of the deterministic nature of the sources of self-jamming and employ digital cancellation techniques to mitigate the receiver desensitization [9–11]. In this chapter the digital cancellation technique is extended to mitigate receiver desensitization caused by two unrelated transmit signals in UL CA. A multiple-input single-output (MISO) digital filter is added to the receiver DSP to reduce RF front-end hardware cost and complexity. This chapter is divided into eight sections. The previous section described a cellular UL CA transceiver and explains the cause of the desensitization of one of the receivers. Section 2.2 goes into the background of adaptive noise cancellation and highlights two exemplary uses of this technique. Section 2.3 describes the modified cancellation algorithm when used to avoid self-jamming in UL CA transmitters. In Section 2.4, the sensitivity of the cancellation algorithms to time alignment errors is studied in representative simulated experiments. Section 2.5 describes laboratory measurements using handset components to demonstrate up to 20 dB of cancellation, by utilizing the proposed algorithm. In Section 2.6, the convergence rate of the algorithm is discussed, and Section 2.7 briefly describes the order of complexity of the algorithms. Section 2.8 summarizes the conclusions of the use of DSP based

cross-modulation cancellation.

2.2 Prior Application of Noise/Distortion Cancellation Algorithms

Since UL CA is an upcoming technique, no published work has focused on the self-jamming issue described in Section 2.1. Therefore, there is no prior cancellation algorithm for this particular case. However, to give a brief overview and explain the origin of adaptive noise cancellation a short summary of the paper “Adaptive Noise Cancelling: Principles and Applications” is provided in this section [12].

2.2.1 Principle of Adaptive Noise Canceller

Fig. 2.3 illustrates the problem and the adaptive noise canceller. A primary input receives a desired signal (s) contaminated by noise (n_0). If one had exact knowledge of n_0 one could simply subtract it from the contaminated received signal. Unfortunately, exact knowledge of n_0 is generally not available. However, in many cases a “reference noise” (n_1), which is strongly correlated to n_0 can be obtained. If for example the difference between n_0 and n_1 can be compensated with a filter, one could process the reference input n_1 subtract it from the primary input. The output of the adaptive noise canceller z would equal the desired signal s . Generally, the structure and parameters of the required filter are unknown. However, by feeding back z to the filter and adapting it one can find a close approximation to the ideal solution such that z approximates s .

At first glance it is not clear how z can be used to adapt the filter for optimal performance. Widrow et. al provide a detailed derivation in [12]. The key

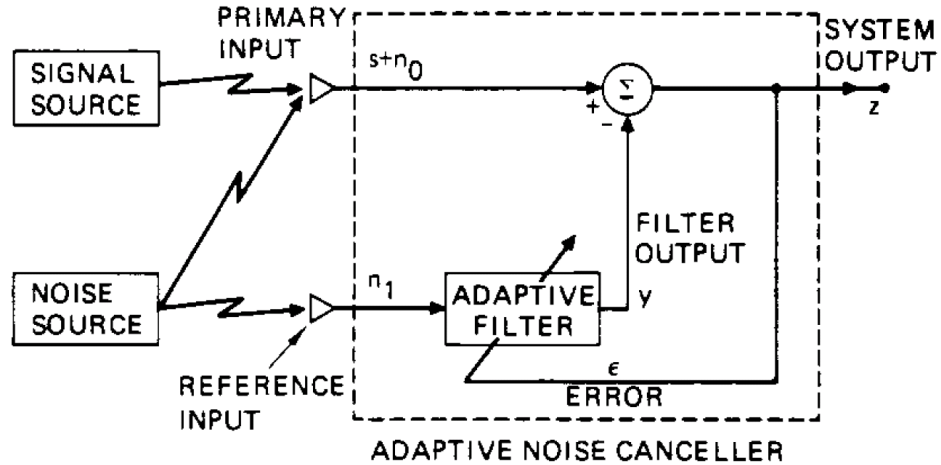


Figure 2.3: The adaptive noise cancelling concept [12].

observation is that n_1 and n_0 are strongly correlated to each other and neither of them is correlated to s . Therefore, by subtracting y from the primary input, the power in z can only decrease by cancelling n_0 from the primary input. Various algorithms have been proposed of the past decades on how to quickly and efficiently adapt the filter coefficients such as the least squares method, the least mean square (LMS) algorithm, and the recursive least squares (RLS) algorithm [13].

A recent use of the adaptive noise canceller for cell phone applications has been presented in [14]. Fig. 2.4 shows a block diagram of a homodyne transceiver. The duplexer isolates the transmitter from the receiver in system when the transmitter and receiver operate simultaneously. However, due to its finite suppression some of the TX signal leaks into the receiver. This is referred to as TX-leakage. In a homodyne architecture the desired received signal is directly downconverted to baseband, where it is sampled by an ADC. The downconverted TX-leakage can be suppressed with conventional filters in the analog or the digital domain. However, due to nonlinearity of the receiver a second order intermodulation component (IMD2) caused by the TX-leakage lands on top of the desired received signal after downconversion and cannot be suppressed with conventional filters. The authors

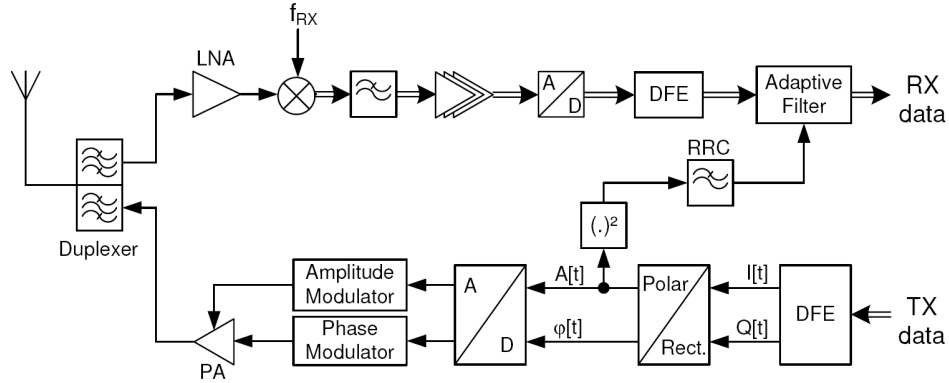


Figure 2.4: Block diagram of a transceiver using polar modulation for transmission [14].

of [14] propose to use the adaptive noise canceller to suppress the interference caused by the IMD2 of the TX-leakage. In this case the reference signal is easily obtainable by squaring the known TX signal. The adaptive filter compensates for the difference in frequency response of the actual IMD2 and the estimated IMD2. Lederer et al. show that this technique reduced the receiver desense due to even order nonlinearity of the receiver by 3-5 dB.

2.3 Proposed Cancellation Algorithm for UL CA Handsets

The interference cancellation for the self-jamming for UL CA handsets described in Section 2.1 is based on the adaptive distortion cancellation approach reviewed in the previous section. Fig. 2.5 shows the block diagram of the distortion canceller for the UL CA case, where a desired received signal $rx2(n)$ is contaminated by a deterministic and predictable signal, in this case the $cm3(n)$ created from the switch or other nonlinear components (where n is time sample index). If an estimate ($cm3^3(n)$) of the contaminating distortion can be measured

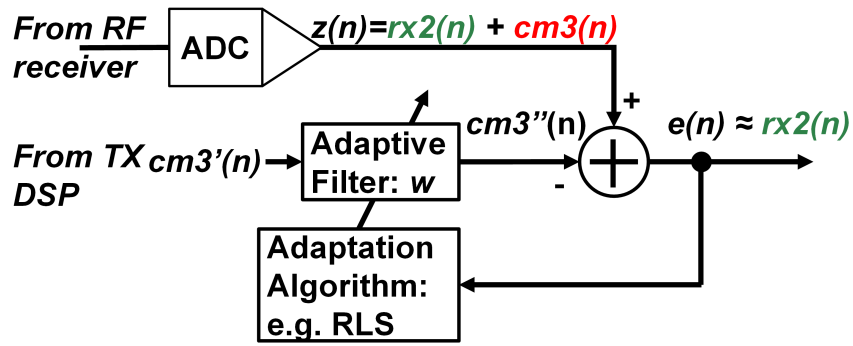


Figure 2.5: Block diagram of adaptive distortion canceller.

or computed, one can subtract it from the distorted received signal $z(n)$ such that only the desired received signal is left at the output of the filter $e(n) \approx rx2(n)$. The distortion estimate can be generated within the transmitter DSP, since both TX1 and TX2 are known. Unfortunately, the distortion estimate ($cm3'(n)$) does not perfectly match the real distortion, since the latter experiences some filtering in the transmitter, power amplifier, and passive front-end components. An adaptive FIR filter w of length M can be used to modify the reference to compensate for the linear response of the system. The optimal filter weights of w , which minimize the effect of $cm3(n)$ in $e(n)$, can be found either in block fashion using the pseudoinverse method discussed below, or are determined iteratively with algorithms such as recursive least squares (RLS). Since $rx2(n)$ is uncorrelated with $cm3'(n)$, the desired signal is not disturbed by the filtering.

2.3.1 Interference Estimation and Single-Input

Single-Output Adaptive Distortion Canceller

The major challenge for good cancellation is the computation of a good estimate of the distortion ($cm3'(n)$). Since RF signals TX1'' and TX2' at the ports of the switch are always in the passband of the system, it is a reasonable

approximation that the signals are not noticeably changed from the known digital baseband signals $tx1(n)$ and $tx2(n)$. Note that one only needs to consider one of the cross-modulation products, since the others lie outside of the receive band and are attenuated by the duplexer. One can estimate the baseband equivalent of the relevant distortion as

$$cm3'(n) \approx tx1^*(n) \cdot tx2^2(n). \quad (2.3)$$

Unfortunately, this is not a sufficiently accurate estimate, since the transmitted signals and the distortion experience an unknown delay in the transmitter and receiver hardware. The DSP distortion estimate needs to compensate for this delay. This is a common time alignment problem and can be represented as follows:

$$cm3'\{J\}(n) \approx tx1^*(n - J) \cdot tx2^2(n - J) \quad (2.4)$$

where J corresponds to the number of delay taps required to time align the distortion estimate with the measured distortion in the received signal $z(n)$. Various time alignment algorithms can be applied, such as the early - late algorithm [15] or other covariance-based algorithms. In this case, the covariance of the measured distortion and the distortion estimate is computed, as shown in Fig. 2.6. In a particular hardware implementation, the group delay is fairly constant, and the search space of J is small.

In addition to the general group delay between the distortion estimate and the measured distortion, it is critical to note that the signal TX1" and TX2' experience slightly different group-delay profiles by the time they reach the switch. This group delay difference is critical and has a significant impact on the quality of the distortion estimate, and therefore needs to be compensated.

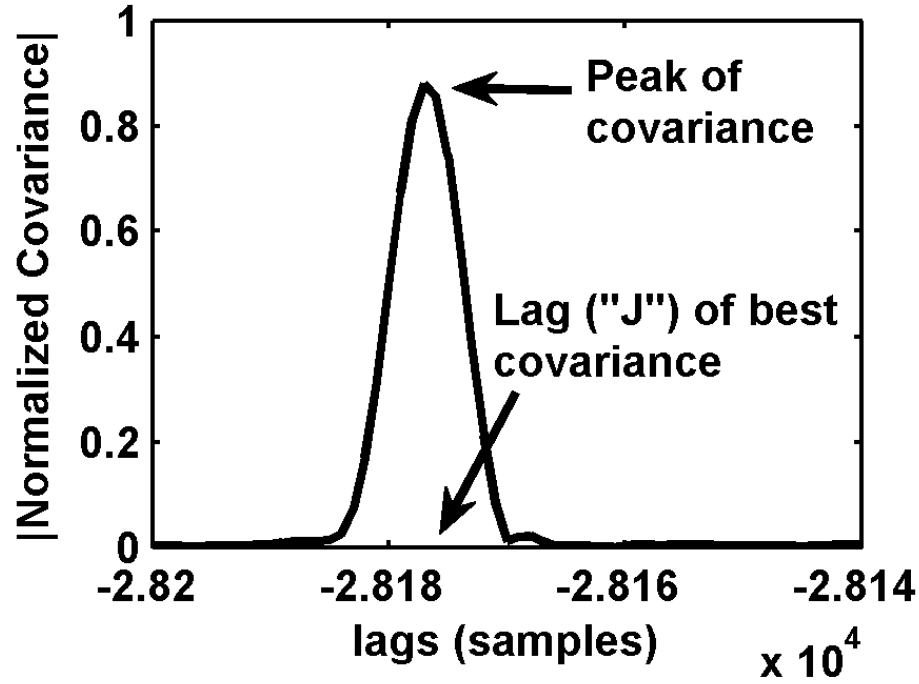


Figure 2.6: Covariance of measured $cm3(n)$ and estimated $cm3'(n)$.

Equation 2.5 includes a new variable K , which can be used to compensate for the group-delay difference of the two signals:

$$cm3'(n)\{J, K\} \approx tx1^*(n - J - K) \cdot tx2^2(n - J). \quad (2.5)$$

One approach to find a good value for K is based on the covariance. The peak of the covariance shown in Fig. 2.6 depends on K . Fig. 2.7 shows the peak of the covariance of the measured distortion $cm3(n)$ and distortion estimate $cm3'\{J, K\}$ for various values of K . For a large search space of K , this would be computationally expensive. Fortunately, K is based on known and relatively constant delays of the transmitter components, and only a fine alignment within a few samples is required.

Fig. 2.8 shows the complete diagram of the cancellation algorithm, where

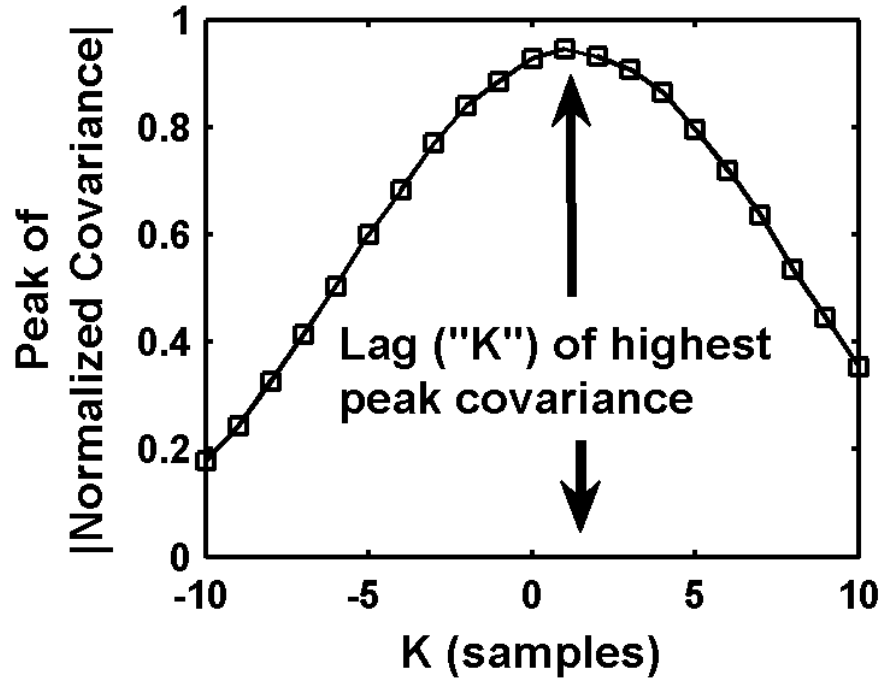


Figure 2.7: Peak of covariance from Fig. 2.6 versus the group-delay difference between TX1' and TX2'.

J and K are determined with the methods described above. The adaptive distortion canceller shown in Fig. 2.8 has been presented in [16]. It implements a single-input single-output (SISO) distortion canceller, where $cm3'(n)$ is calculated prior to the adaptive filter. It has been demonstrated to work very effectively in directly coupled transmitters. However, the SISO algorithm does not give the adaptation sufficient degrees of freedom to separately modify $tx1(n)$ and $tx2(n)$. This is required when the two transmitters are coupled through the antenna with strong spectral shaping or in cases where multipath effects are significant or TX2 is reflected by its own transmit antenna due to insufficient matching.

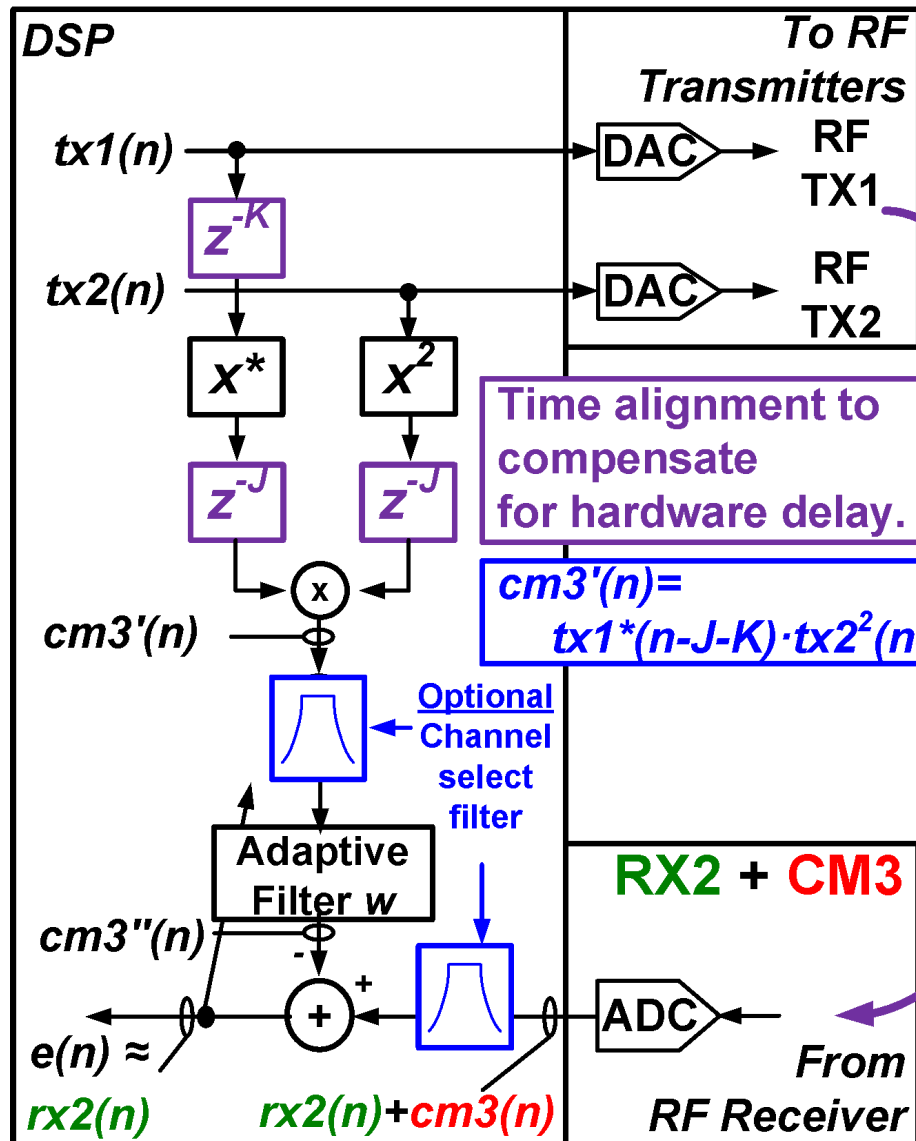


Figure 2.8: Block diagram of SISO adaptive distortion canceller.

2.3.2 Multiple-Input Single-Output Adaptive Distortion Cancellation

In this section, a multiple-input single-output (MISO) adaptive distortion canceller is proposed as an extension to the SISO canceller. Conceptually, $tx1^*(n)$ (the complex conjugate of $tx1(n)$) and $tx2^2(n)$ pass through their own adaptive FIR filters w and v before they are multiplied to form the estimated cancellation signal $cm3''(n)$. The MISO approach has sufficient degrees of freedom in the adaptation to compensate for the transfer function and multipath effects in antenna coupled transmitters. Fig. 2.9 shows a block diagram of the MISO filter structure. An additional advantage of the MISO algorithm with separate filters for $tx1^*(n)$ and $tx2^2(n)$ is that it can compensate for incorrect estimation of K if the filter length of v and w is sufficient. This can even be used to skip the search for K using the covariance method explained above. Instead of using a separate search for K , one can take advantage of the fact that in a given hardware environment K will be relatively fixed in a range of K_{min} to K_{max} . One can set K to K_{min} and extend the filter length of w by $K_{max} - K_{min}$ taps (often one or two taps are sufficient). This has the advantage that no separate covariance based search for K is necessary; however, this increases the order of complexity of the parameter estimation.

Whether or not a separate covariance-based search for K is performed, the challenging part is the optimal estimation of the filters w and v . Unfortunately, direct estimation of the filter $M + N$ coefficients of w and v is difficult due to the nonlinear dependency of the parameters. Following the logic of [17], the input-

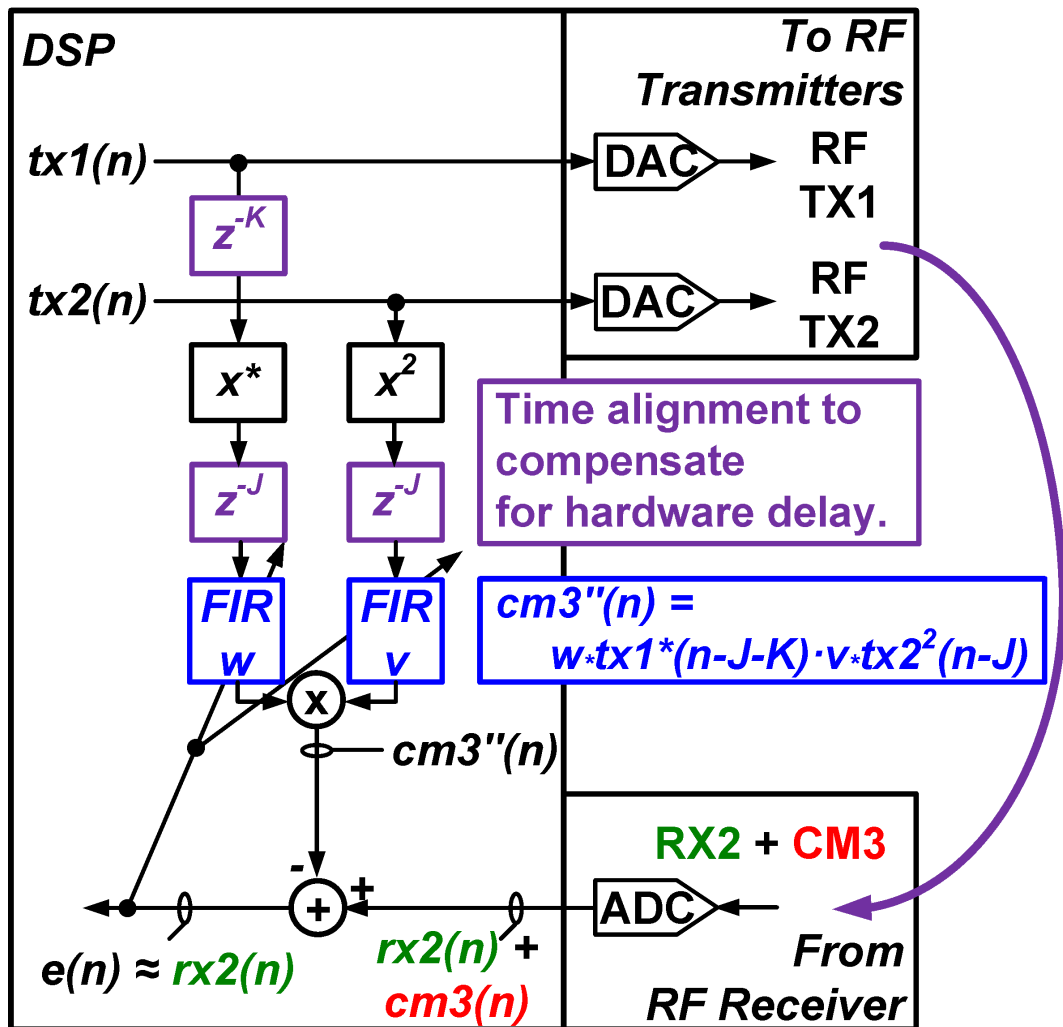


Figure 2.9: Block diagram of MISO adaptive distortion canceller.

output relationship of MISO filter can be written as

$$\begin{aligned} cm3(p) &= \left(\sum_{i=0}^{M-1} w_i \cdot tx1^*(p-i) \right) \left(\sum_{j=0}^{N-1} v_j \cdot tx2^2(p-j) \right) + n(p) \\ &= \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} w_i \cdot v_j \cdot tx1^*(p-i) \cdot tx2^2(p-j) + n(p) \end{aligned} \quad (2.6)$$

where $n(p)$ is the modeling error, e.g., caused by noise. Using the following notation:

$$\theta = [w_0v_0, \dots, w_0v_{N-1}, \dots, w_{M-1}v_0, \dots, w_{M-1}v_{N-1}]^T \quad (2.7)$$

$$\phi_p = [tx1^*(p)tx2^2(p), \dots, tx1^*(p-(M-1))tx2^2(p-(N-1))]. \quad (2.8)$$

Equation (2.6) can be written as

$$cm3(p) = \phi_p \theta + n(p). \quad (2.9)$$

The latter form has the advantage that the problem is in linear regression form.

Given a P -point data set and defining

$$CM3_P \triangleq [cm3(0), cm3(1), cm3(2), \dots, cm3(P-1)]^T \quad (2.10)$$

$$N_P \triangleq [n(0), n(1), \dots, n(P-1)] \quad (2.11)$$

$$\Phi_P \triangleq [\phi_0, \phi_1, \dots, \phi_{P-1}]^T, \quad (2.12)$$

the parameter estimation problem can be written as

$$CM3_P = \Phi_P \theta + N_P. \quad (2.13)$$

The best estimate of the parameters in least-mean-squares sense can be found with the pseudoinverse

$$\hat{\theta} = (\Phi_P^T \Phi_P)^{-1} \Phi_P CM3_P. \quad (2.14)$$

The transformed problem with a "combined $w \cdot v$ filter" instead of two separate filters requires the estimation of $M \cdot N$ parameters instead of $M + N$ parameters. This, in turn, implies a higher computational cost. However, in this particular application, the order of the filters M and N is low enough that the difference in computational effort is not prohibitively high given the capability and efficiency of the DSP in today's cell phones. An alternative to the block computation of the parameters using the pseudoinverse is the RLS algorithm. In each iteration, the RLS algorithm computes an updated estimate of the parameters $\hat{\theta}$ based on the previous estimate and the current measured data. The RLS algorithm estimates the parameter in the $\hat{\theta}$ i th iteration as

$$\hat{\theta}_i = \hat{\theta}_{i-1} + P_i \Phi_i [cm3(i) - \Phi_i \hat{\theta}_{i-1}] \quad (2.15)$$

where P_i is

$$P_i = \lambda^{-1} \left[P_{i-1} - \frac{\lambda^{-1} P_{i-1} \Phi_i^* \Phi_i P_{i-1}}{1 + \lambda^{-1} \Phi_i P_{i-1} \Phi_i^*} \right]. \quad (2.16)$$

Since the RLS algorithm continuously adapts the parameters $\hat{\theta}_i$, the algorithm can quickly adapt to the changing external environment, such as the hand position of the user. This is crucial, since the hand position has a significant impact on the coupling between the two transmitters, which, in turn, affects the distortion.

2.3.3 Multiple Nonlinear Components

As mentioned above, there can be multiple sources of nonlinearity such as the duplexer and the switch. Referring to Fig. 2.1, it is important to realize that distortion from the switch is created from TX1'' and TX2'. The distortion caused by the duplexer is created from TX1'' and TX2, where the apostrophes denote small differences between the signals caused by the switch and the duplexer transfer functions and group delays. In particular, this means that an estimate for each distortion signal should have slightly different values for J and K . Depending on the sampling rate of the system and the components, the differences in J and K might be in the subsample range. An alternate interpretation of this is that TX2 and TX1 experience an echo and multiple copies of the signals create multiple *cm3s* in a single component. FIR filters are frequently used to model multipath and echo effects. Therefore, the structure of the MISO filter inherently is capable of compensating for distortion created by multiple components.

2.3.4 Cancellation in the Presence of Desired RX Signal

So far, the discussion has focused only on the distortion cancellation. However, the main objective is to achieve good reconstruction of the desired signal received from the base station. The input to the adaptive filter contains the distortion and the desired received signal. The latter behaves like noise to the parameter estimation algorithm. In extreme cases, when the desired received signal is significantly stronger than the distortion, it degrades the cancellation performance of the distortion canceller. The output power of the adaptive filter is

$$e^2 = EMSE + \sigma_{RX2}^2 + \sigma_n^2 \quad (2.17)$$

where σ_{RX2}^2 and σ_n^2 are the powers in the received signal and the thermal noise floor. EMSE is the power of the residual distortion after cancellation. After the RLS algorithm converges, this can be expressed as

$$EMSE_{RLS} \approx \frac{(\sigma_{RX2}^2 + \sigma_n^2)(1 - \lambda)L}{2} \quad (2.18)$$

where L is the number of parameters to be estimated and λ is the forgetting factor [18]. From (2.18), one can determine the amount of distortion cancellation as

$$Cancellation_{CM3} \approx \frac{(\sigma_{RX2}^2 + \sigma_n^2)}{\sigma_{CM3}^2} \cdot \frac{(1 - \lambda)L}{2}. \quad (2.19)$$

Fig. 2.10 shows the amount the cancellation for a given received signal power above or below the distortion for L equal to 4 and various values of λ . The graph illustrates that, for a strong received signal and λ equal to 0.99, the EMSE is as strong or stronger than the CM3 itself. This means that the adaptive filter would actually degrade the SINR of the desired received signal. To achieve good cancellation for strong received signals, one needs to increase λ to 0.9999. Unfortunately, this also increases the convergence time of the algorithm. However, in those cases, the SINR of the desired received signal is high enough that longer convergence time is acceptable. This will be discussed in more detail in Section 2.6.

2.3.5 Cancellation in the Presence of Adjacent Channel Jammers

In addition to the desired received signal, one also has to consider adjacent channel jammers. These can be significantly stronger than the desired received signal and would disturb the cancellation algorithm similar to the RX signal shown

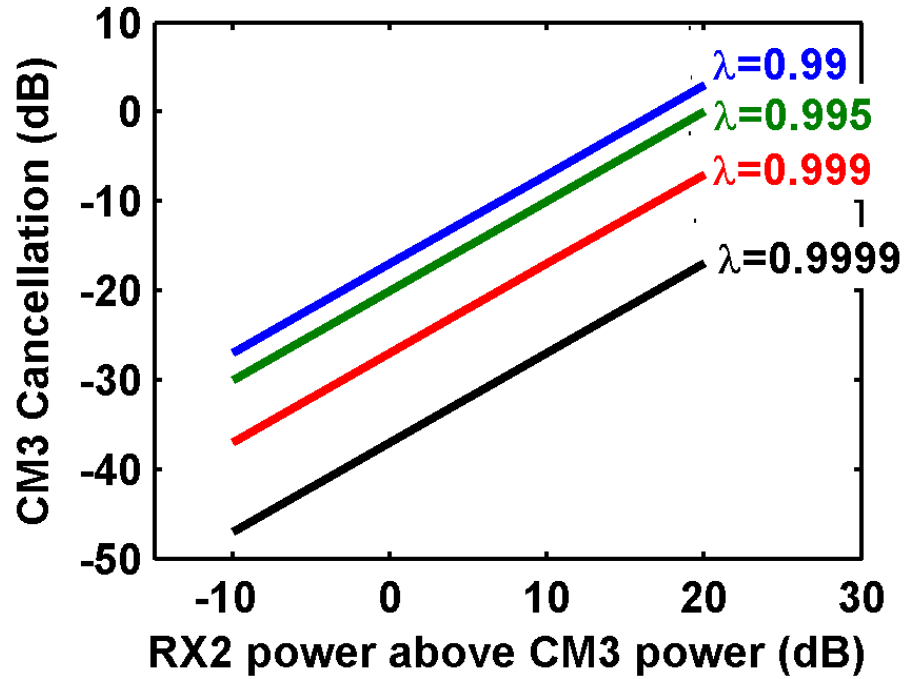


Figure 2.10: Amount of cancellation given λ for various received signal powers.

in Fig. 2.10. However, in the case of adjacent channel jammers, we can benefit from the suppression of the jammer by the digital channel filter at the input of the receiver. This also requires a modification to the cancellation algorithms to include the impact of the channel select filter. In the case of the SISO filter, this is straightforward and illustrated in Fig. 2.8. It is sufficient to insert the channel select filter at the receiver input and after $cm3'(n)$ is generated.

In the MISO distortion canceller, it is slightly more complicated. The adaptive distortion cancellation portion, including channel select filters for the MISO distortion canceller, is shown in Fig. 2.11. The measured received signal $z(n)$ is processed by a channel select filter to suppress the jammer. It also suppresses parts of the $cm3(n)$ signal that lie outside of the band of interest. In the MISO case, each of the vectors of the form $tx1(p)^*tx2^2(q)$ still contains the information outside of the band of interest that are not present at the filter input $d(n)$. These

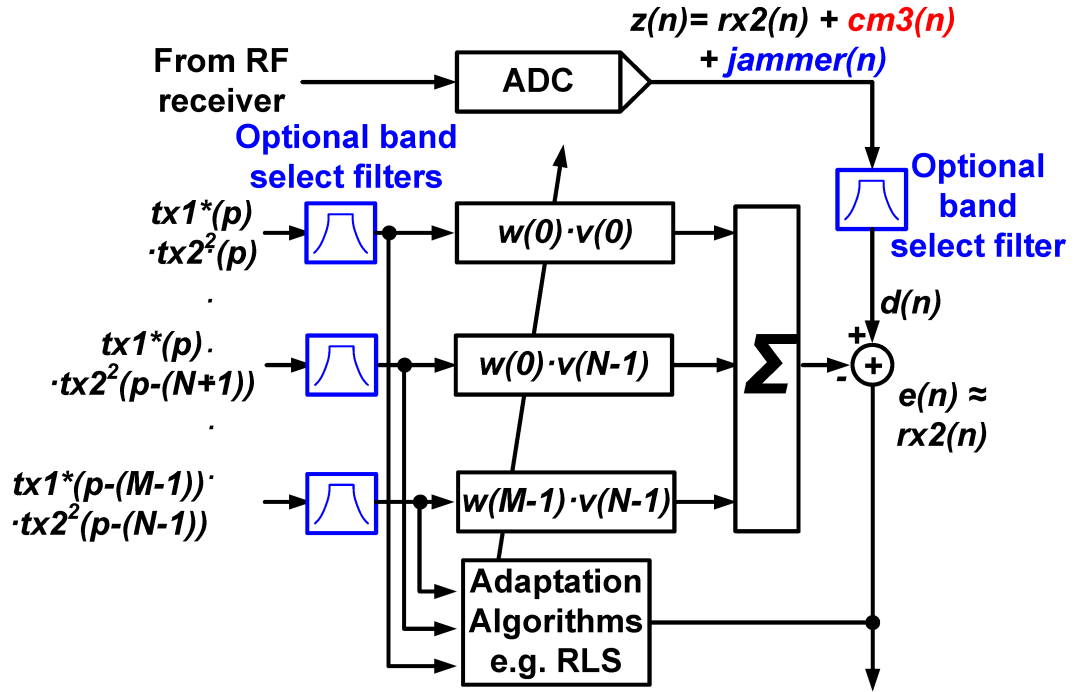


Figure 2.11: Adaptive MISO filter with digital channel select filter for adjacent channel jammer suppression.

out-of-band components need to be filtered out so as to not interfere with the parameter estimation process. This is similar to the approach presented in [19].

2.4 Simulation Results

In this section, MATLAB simulations are used to prove the robustness of the proposed algorithms under different conditions. The section focuses particularly on aspects relating to the time alignment. The cancellation performances in the presence of received signals and out-of-band jammers are discussed in detail in Section 2.5 using measured data sets and are not included here for brevity.

To simulate the cancellation performance, a representative simulated experiment is generated from two 5-MHz LTE signals representing $tx1(n)$ and $tx2(n)$. The two signals pass through test filters emulating the analog hardware and cou-

pling transfer function with the coefficients:

$$w_{test} = [10.5e^{-j155^\circ} \quad 0.25e^{-j15^\circ}] \quad (2.20)$$

$$v_{test} = [10.3e^{-j135^\circ} \quad 1.0e^{j35^\circ}]. \quad (2.21)$$

The filtered signals are multiplied to generate a test distortion $cm3test$. Numerical white noise is added 80 dB below the distortion. The distortion to noise ratio is significantly higher than in the experiment in order to have sufficient dynamic range in the simulation to illustrate some of the subtle differences in performance of the SISO and MISO filter.

2.4.1 Filter Adaptation With Perfect Time Alignment

To evaluate the performance of the parameter extraction alone, no time misalignment is introduced. Therefore, the search for the group-delay values J and K is not performed. To find the optimal parameters for the two different filter structures, the pseudoinverse approach was used with 1000 training points. Fig. 2.12 shows the following three equivalent baseband signals: the emulated distortion, the signal after cancellation using the SISO algorithm from Fig. 2.8, and the signal after the MISO filter from Fig. 2.9. The SISO algorithm achieves 20-dB cancellation, while the MISO filter achieves cancellation of 80 dB, down to the noise floor. It is understandable that the SISO algorithm only achieves a modeling accuracy of 20 dB. That might be sufficient in many cases where reasonably linear components are used, and the coupling between the two transmitters is well behaved. In those cases, the SISO algorithm could be the better choice, since it is computationally less expensive.

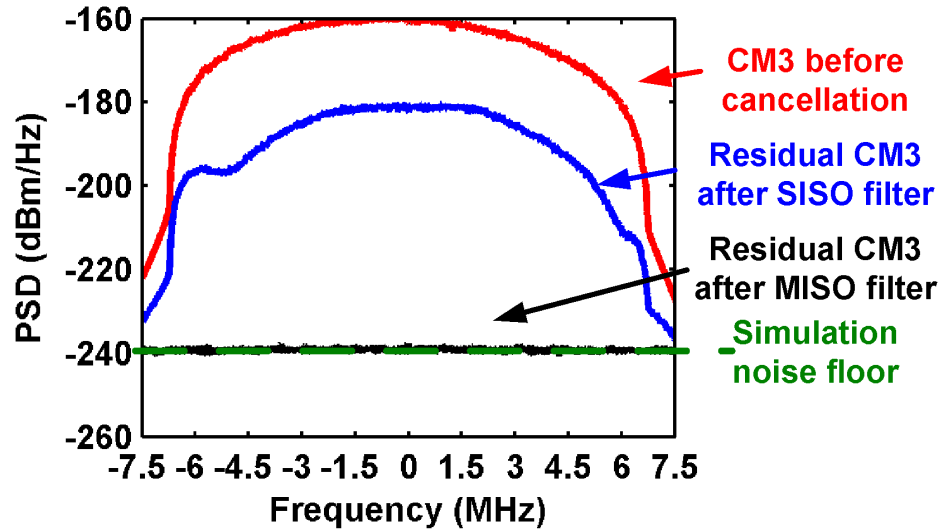


Figure 2.12: Spectral plots of the simulated distortion before and after cancellation using the SISO and the MISO filter.

2.4.2 Filter Adaptation With Time Alignment Errors

As described in Section 2.3, appropriate time alignment is critical. In order to test the time alignment algorithm presented in Section 2.3, two test distortion signals $cm3test$ are generated. The first test assumes that $tx1(n)$ and $tx2(n)$ are unfiltered, but time-shifted, with $J = 10000$ and $K = 4$. In this case, K is estimated correctly, since the covariance of $cm3test$ to $cm3'$ peaks for $K = 4$, as shown in Fig. 2.13 (the line with the circles). The adaptive distortion canceller performs similarly as in the previous case and cancels the distortion down to the noise floor. The second test introduces again the same time misalignment as before, but in addition $tx1^*(n)$ and $tx2^2(n)$ are filtered, respectively, with $wtest$ and $vtest$ from (2.20) and (2.21). In this case, J is correctly estimated. Fig. 2.13 shows the normalized peak covariance for various values of K . Based on that graph, one would incorrectly conclude that K equal to 5 is the best choice instead of K equal to 4. Fig. 2.14 shows the cancellation results for K equal to 4 and 5. As can be seen, better results are achieved if K is set to 4. However, in practical

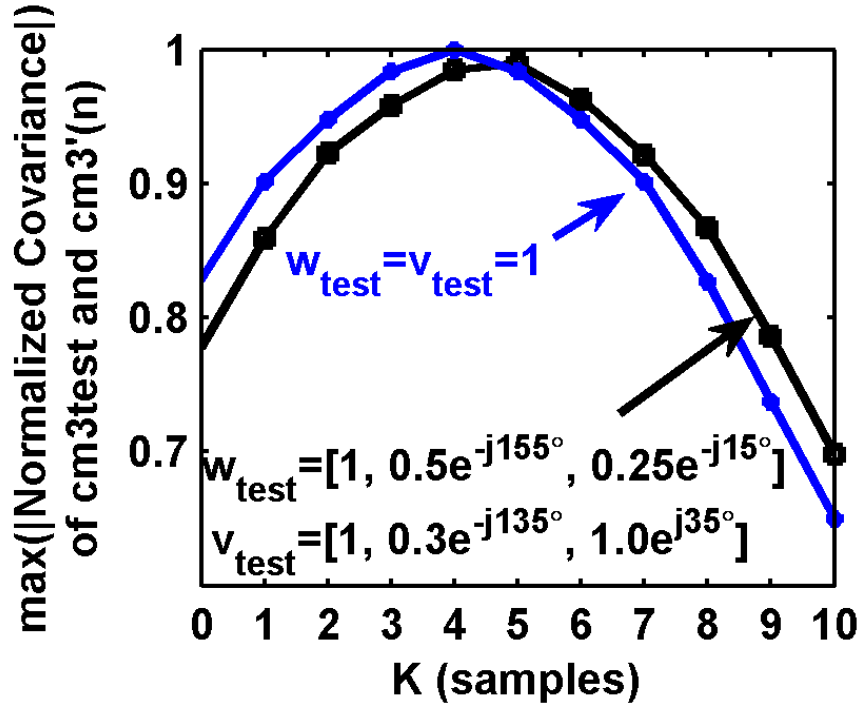


Figure 2.13: Peak of covariance of $cm3_{test}$ and for various K .

cases, the cancellation will be limited by the system noise floor before it is limited by this error in the K estimation. As described in Section 2.3.3, an additional advantage of the MISO algorithm with separate filters for $tx1^*$ and $tx2^2$ is that it can compensate for incorrect estimation of K , if the filter lengths of w and v are increased. Fig. 2.15 shows the cancellation results for K equal to 3 and filter lengths of 3 and 4 for both filters. For the former, the cancellation is not perfect, since the adaptive filters are not long enough to compensate for the estimation error in K and to compensate for w_{test} and v_{test} at the same time. With the extended filter length the cancellation is perfect, but it comes at a slightly higher computational expense due to the higher filter order. Fig. 2.16 shows the amount of obtainable cancellation for the case that M and N are 3 for the MISO filter and M is 3 for the SISO filter when time misalignment errors are introduced.

Two types of time misalignment errors are studied. First, if the group

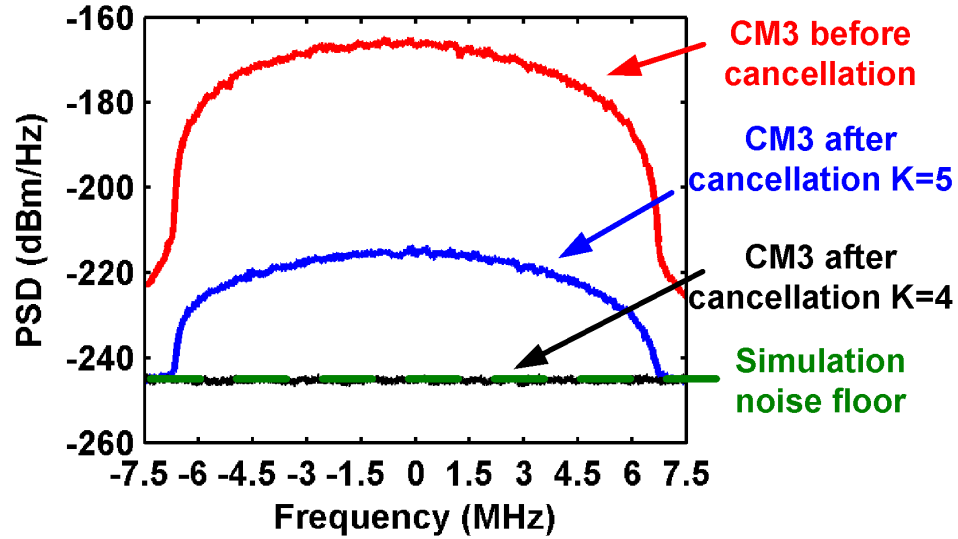


Figure 2.14: Simulated distortion before and after the MISO canceller using correct and incorrect group delay adjustment of K .

delay between $cm3_{test}$ and $cm3'(n)$ is misadjusted, i.e., J is incorrectly estimated. The second error introduces a misalignment in the group-delay difference between TX1" and TX2', i.e., an error in the estimation of K . From Fig. 2.16, one can clearly see that to obtain a cancellation in the order of 20 dB, the MISO filter can tolerate significant errors in J and K estimation. The SISO filter is relatively forgiving for underestimation of J by three samples, since the SISO filter is three taps long. However, since the SISO filter cannot compensate for errors in K , it is fairly sensitive to errors in K .

2.5 Experimental Results

2.5.1 Measurement Setup

The UL CA handset transmitter, as shown in Fig. 2.1, is implemented using B5 and B13 commercial handset PAs, handset duplexers, and a high linearity

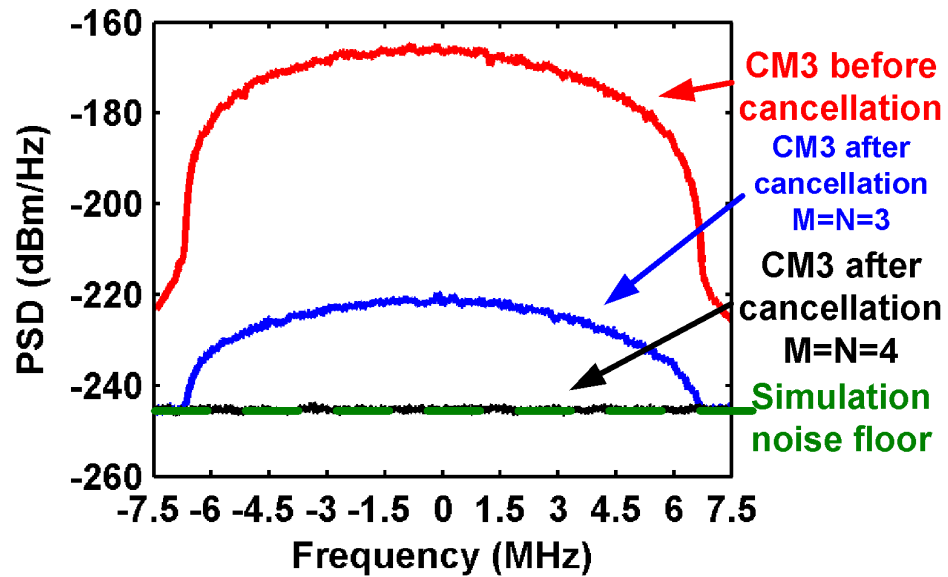


Figure 2.15: Cancellation performance for different filter lengths when K is underestimated ($K = 3$) for the MISO canceller.

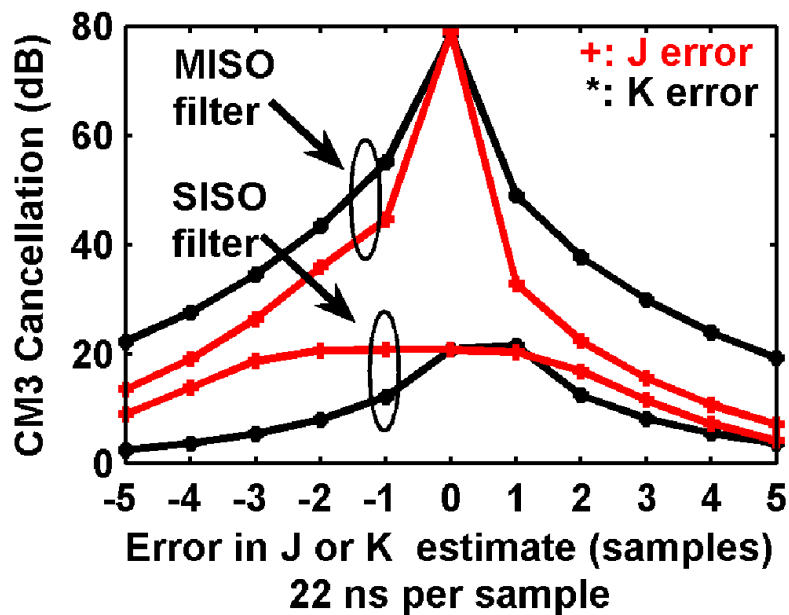


Figure 2.16: Sensitivity of MISO and SISO filter to errors in time alignment.

handset switch. Fig. 2.17 shows the complete measurement setup. The transmit signals are generated by Agilent RF signal generators. A third signal generator injects a signal in the receive band of B5 to emulate the desired received signal from the base station. The transmitters were coupled with two cell phone antennas. The antennas were placed such that the coupled transmit power of B13 (TX1”) is approximately 12 dBm, corresponding to approximately 10 dB of isolation between the two antennas. The TX2’ power at the switch is approximately 24 dBm. An instrumentation LNA followed by an Agilent signal analyzer is used as a high sensitivity receiver at the RX port of the B5 duplexer. The LNA has an IIP3 of 9 dBm. High end handset LNAs can achieve comparable performance [20, 21]. Assuming 50-dB attenuation of TX1 and TX2 through the duplexer, one can compute with (2) that the CM3 generated in the LNA has a power of -106 dBm, which is significantly lower than the CM3 of the switch and only slightly desensitizes the receiver. Furthermore, our proposed algorithm is expected to cancel CM3 generated from multiple sources. The sampling rate of the system is 45MHz, and one sample is approximately 22 ns. The measurement noise floor of this setup is -170 dBm/Hz. TX1, TX2, and RX2 are 4.5-MHz-wide LTE signals.

2.5.2 Cancellation Results

Throughout this section, the RLS algorithm was used to adapt the filter coefficients. Since the filter response of the system is relatively flat, only a total number of four parameters were required ($N = M = 2$) for the MISO filter and two parameters for ($M = 2$) SISO filter.

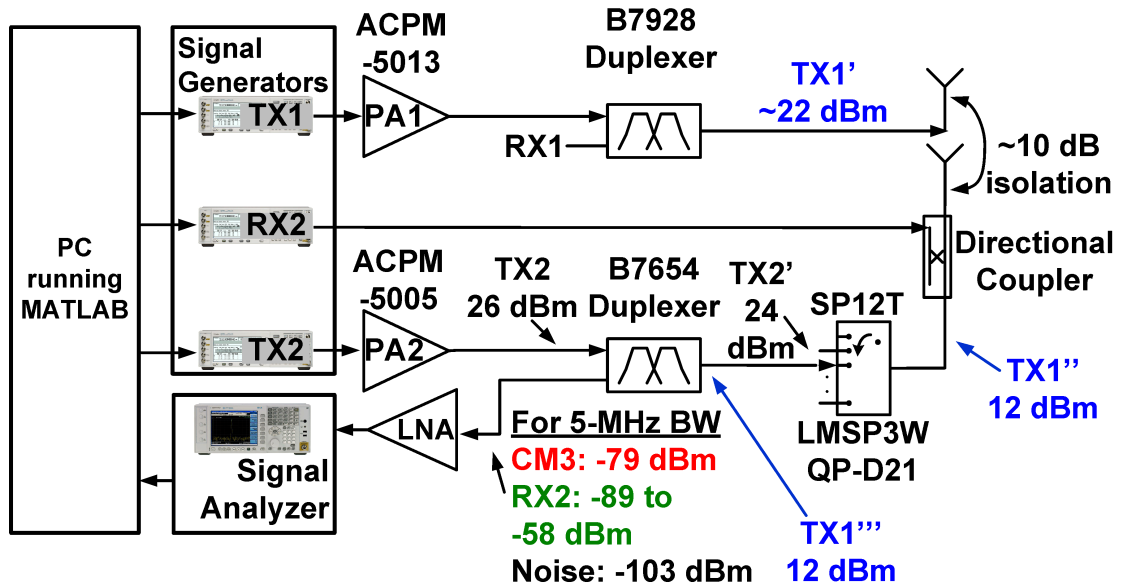


Figure 2.17: Measurement setup mimicking an UL CA handset.

Duplexer and Switch Distortion:

It is important to note that when two transmitters are strongly coupled, the duplexer alone creates a noticeable distortion component. In a first experiment, the switch was removed and a distortion from the duplexer of -96 dBm was observed in the 5-MHz receive band (-163 dBm/Hz), which is 7 dB above the measurement noise floor. This corresponds to an approximate duplexer IIP3 of 78 dBm.

Fig. 2.18 shows the complex baseband representation of the captured signal before and after cancellation using the SISO or the MISO filter. One can see that the distortion is cancelled almost down to the noise floor using either filter. Note that the spurs are part of the system noise floor, since antennas are picking up signals from the adjacent instruments. If both the switch and the duplexer are in the system, the total distortion power is -77 dBm over 13.5 MHz of distortion bandwidth. The power of 5-MHz in band distortion is approximately -79 dBm, corresponding to -146 dBm/Hz. Fig. 2.19 shows the distortion before and after cancellation using either the SISO or the MISO filter. Both algorithms achieve

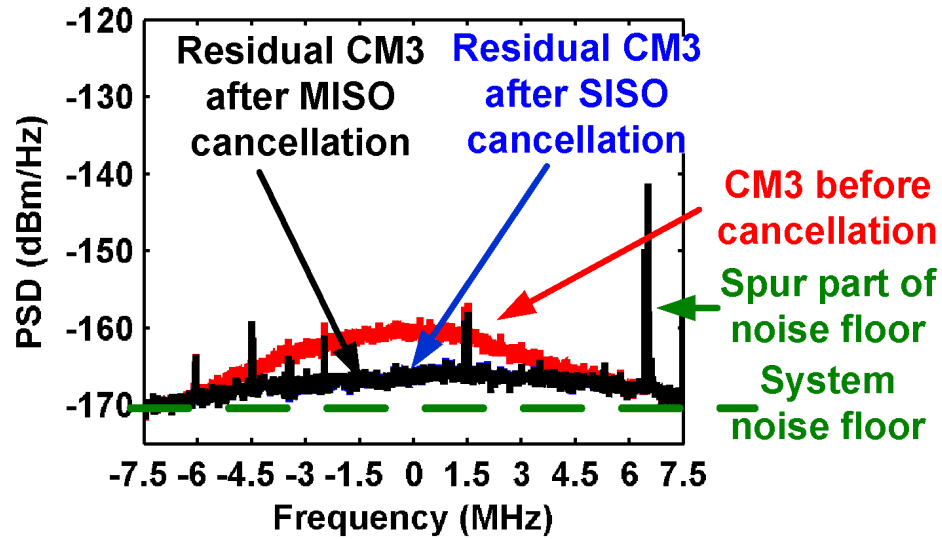


Figure 2.18: Measured duplexer distortion before and after cancellation using MISO or SISO filter.

virtually the same cancellation of approximately 20 dB. This shows that either algorithm cancels most of the distortion from switch and duplexer. However, it is known from Sections 2.3 and 2.4 that the SISO filter is more sensitive to errors in the time alignment. If K is incorrectly estimated only by one sample, 22 ns, the SISO filter performs slightly worse than the MISO filter. Fig. 2.20 shows the SISO and MISO cancellation results for that case. The MISO filter still achieves approximately 20 dB of cancellation, while the SISO filter achieves 15 dB cancellation.

Distortion and Desired Received Signal:

To demonstrate the effectiveness of the cancellation algorithm, a desired received signal at various power levels is added to the distortion of -79 dBm. For brevity, the performance of only the MISO filter is discussed in this chapter. The performance of the SISO filter has been presented in [16]. In a first experiment, a desired received signal was injected 10 dB below the distortion. Fig. 2.21 shows

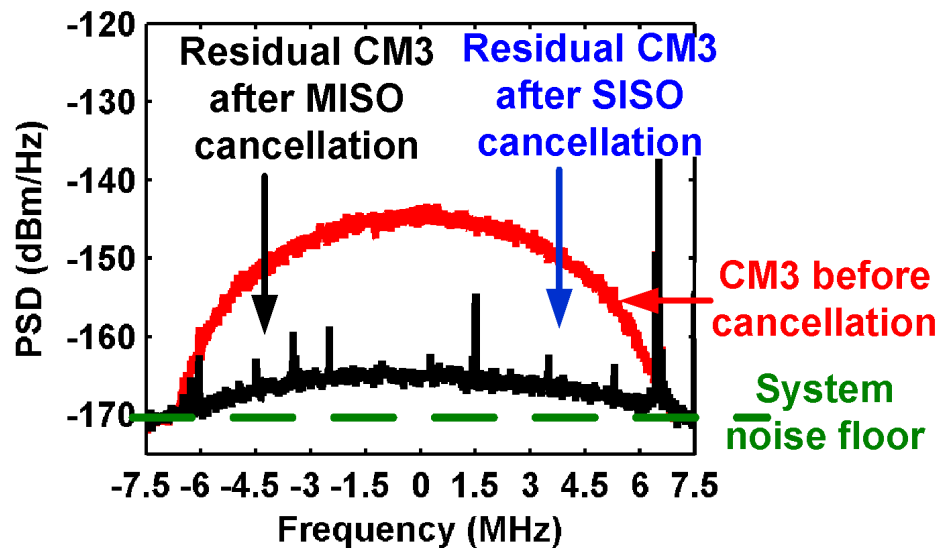


Figure 2.19: Measured switch and duplexer distortion before and after cancellation using either the MISO or the SISO filter.

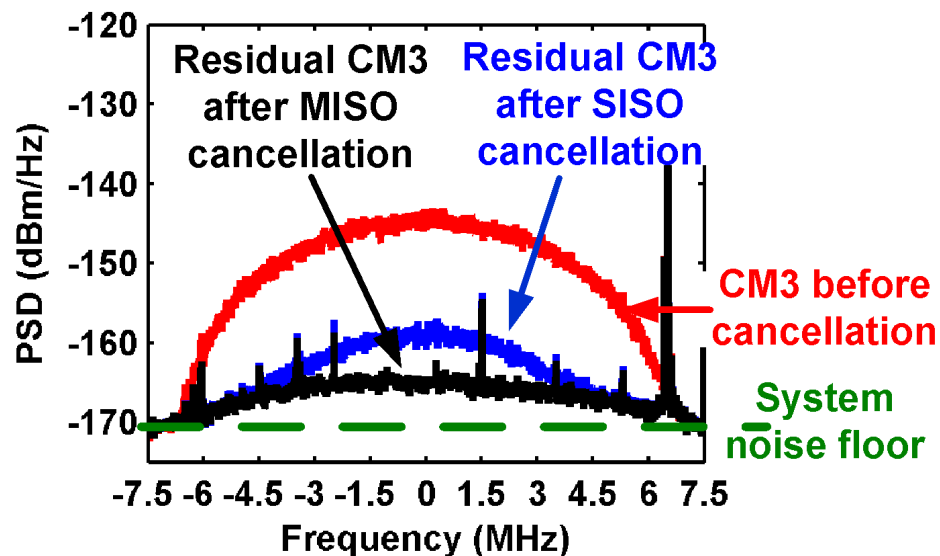


Figure 2.20: Measured switch and duplexer distortion before and after cancellation using the SISO or the MISO filter with K alignment error by minus one sample.

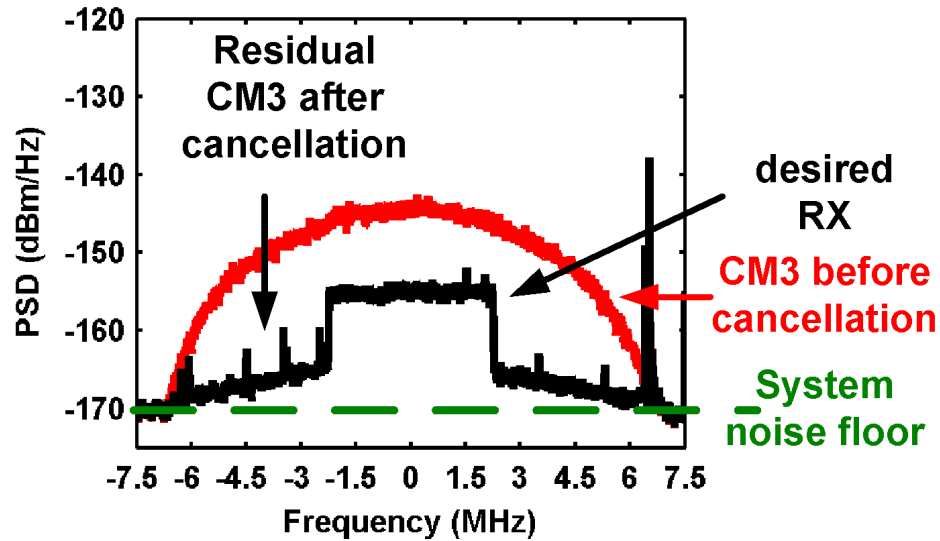


Figure 2.21: Low-power received signal with distortion before and after cancellation (SINR before/after cancellation ≈ -10 dB/ $+9.4$ dB).

the baseband spectral responses of signals before and after cancellation using the MISO filter. Before cancellation, the desired received signal was not even visible in the spectral plot. However, after cancellation, one can clearly see the desired signal.

In a second experiment, the desired received signal was 20 dB stronger than the distortion. Fig. 2.22 shows the spectral responses of the desired signal and the distortion before and after cancellation using a forgetting factor of λ of 0.99 and 0.9999. As discussed in Section 2.3.4, λ of 0.99 will not achieve any cancellation and λ was increased to 0.9999 to improve the cancellation. The in-band cancellation response is not immediately apparent from the spectral plot. However, comparing the sidelobes, one can see that the algorithm attenuates the distortion by approximately 15 dB in both cases if an appropriate λ is chosen. To evaluate the in-band response of the cancellation algorithm, the EVM and SINR of the desired received signal before and after cancellation are computed.

Table 2.1 summarizes the results for various power levels of the desired

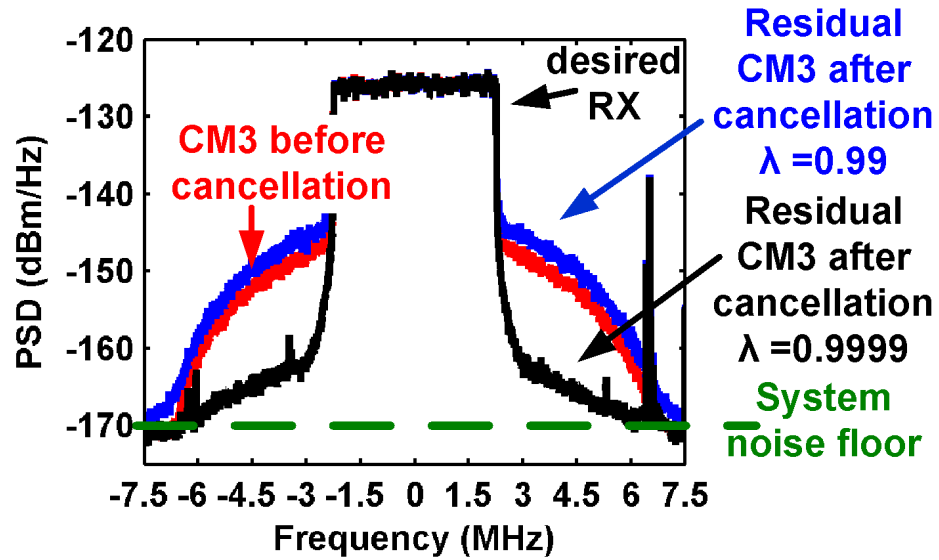


Figure 2.22: High-power received signal with distortion before and after cancellation (SINR before/after cancellation ≈ 20 dB/32 dB).

received signal. Three scenarios are studied. In the first one, no distortion is present, and the results give a baseline of our measurement setup. In the second scenario, the distortion is present, but the cancellation has not been applied yet. The third scenario evaluates the received signal after cancellation. As can be seen from the results in Table 2.1, the algorithm almost completely cancels the effect of the distortion, independent of the power of the desired signal. When distortion is noticeably stronger than the received signal, the EVM improves from 118% to 34%. When the received signal is 20 dB stronger than the distortion, the received signal acts as noise on the RLS adaptation mechanism. Nonetheless, the EVM still improves from 10% to 2.5%. The SINR after cancellation is estimated based on the EVM and (2.22) and shows 12 to 20 dB improvement. Similar results for the SISO filter were obtained and reported in [16] for a directly coupled system.

$$SINR \approx -20 \log_{10} EVM. \quad (2.22)$$

Table 2.1: Cancellation Performance of MISO Filter

RX Power (dBm)	No CM3		Before Cancellation		After MISO Cancellation	
	SINR (dB)	EVM (%)	SINR (dB)	EVM (%)	$\tilde{\text{SINR}}$ (dB)	EVM (%)
-89 ¹	15	17	-10	118	9.4	34
-79 ²	25	5	0	77	16.5	15
-69 ³	35	2	10	31	24.4	6
-59 ⁴	45	1	20	10.3	32	2.5
-89 ⁵	15	17	15	17	13.2	22
-89 ⁶	11	27	-15	119	8.2	39

¹ $\lambda = 0.99$; ² $\lambda = 0.995$; ³ $\lambda = 0.999$; ⁴ $\lambda = 0.9999$;

⁵ $\lambda = 0.99$ and no CM3 present before or after MISO cancellation

⁶ $\lambda = 0.995$ and a -45dBm jammer is added at a 6.25 MHz offset

The fifth experiment listed in Table 2.1 applies the MISO cancellation algorithm to a measurement data without any CM3 present. In this case, the SINR and EVM after cancellation are slightly degraded, since the algorithm adds a little bit of noise to the received signal. Therefore, it is desirable to disable the cancellation algorithm for high SINR and when TX1 and TX2 are both not operating close to peak output power.

Distortion and an Out-of-Band Jammer:

Today's cell phone receivers are expected to tolerate -43 dBm strong jammers at the antenna according to the 3GPP standard, which reduces to approximately -45 dBm at the receiver input after insertion loss of the duplexer. From the previous section, it is clear that such a strong jammer would significantly degrade the adaptation performance of the RLS unless one increases λ even closer to 1. However, in the case of out-of-band jammers, one can benefit from a dig-

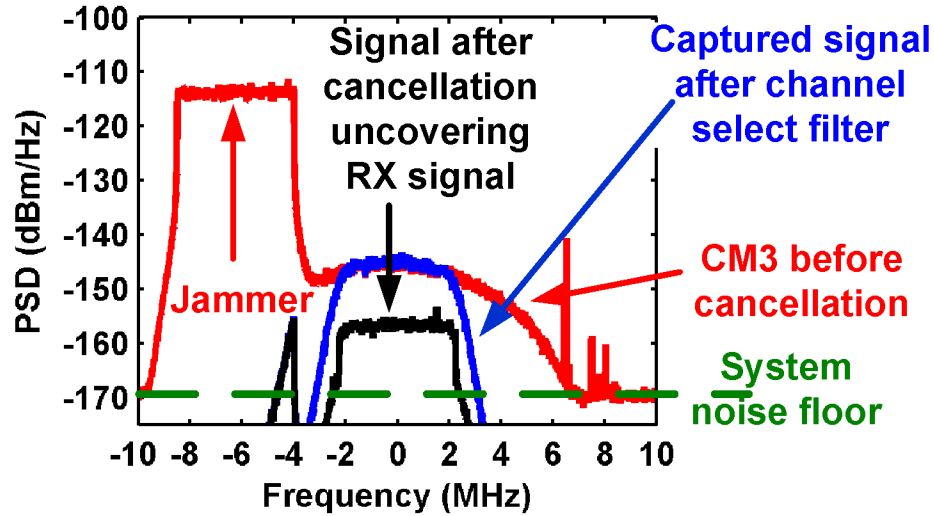


Figure 2.23: Captured signal with out-of-band jammer before and after channel select filtering and after adaptive distortion cancellation.

ital channel select filter, as discussed in Section 2.3.5. This reduces the jammer power to acceptable levels. This has been experimentally verified, by injecting a -43 dBm jammer at a 6.25-MHz offset in addition to the weak received signal of -89 dBm, which is 10 dB below the -79 -dBm distortion. Note that the addition of the jammer already decreased the EVM of the system baseline, even before the addition of the distortion, as shown in Table 2.1. Fig. 2.23 shows the distortion and the out-of-band jammer, before and after the channel select filter. The desired received signal is not apparent in the spectral plot since it is 10 dB below the distortion. The out-of-band jammer has been sufficiently attenuated by a 5th order Butterworth filter with a corner frequency of 2.5 MHz. The desired received signal is uncovered from the remaining in-band signal after the adaptive distortion cancellation. The in-band response is evaluated and summarized in Table 2.1. Before cancellation, the EVM was 119%, and after cancellation the EVM was reduced to 39%. The latter corresponds to an SINR of 8.2 dB, which implies the algorithm cancelled approximately 18 dB of the distortion.

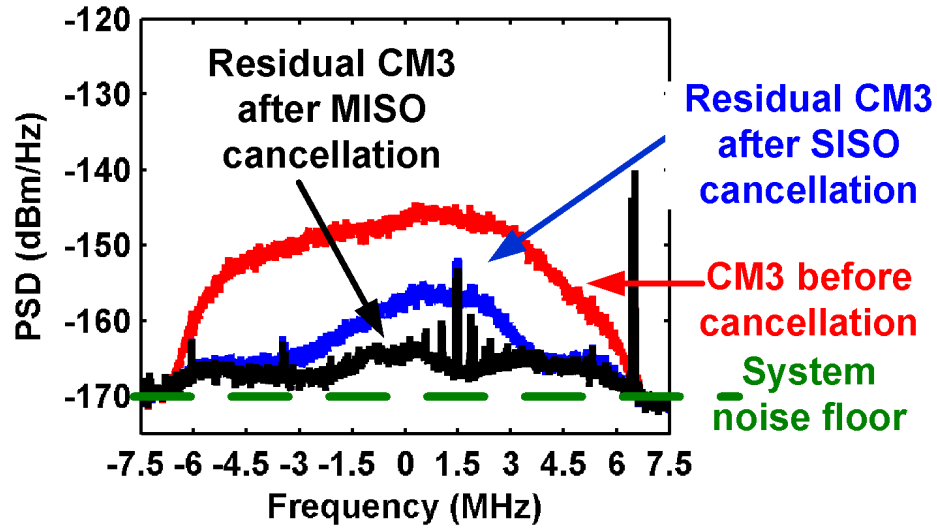


Figure 2.24: Captured CM3 in experiment using different antennas. Reflections change distortion shape. SISO filter length is 4. MISO filter length is 16.

Using Different Antennas:

The previous experiments were carried out using cell phone antennas with little spectral shaping or multipath effects. As shown in Fig. 2.19, the SISO algorithm with appropriate time alignment performed similarly to the MISO algorithm. In an alternative setup, using different antennas, the distortion was noticeably affected by the reflection of TX2 from its transmit antenna and the multipath effects of TX1. Fig. 2.24 shows the CM3 before cancellation. One can clearly see a hump around 1.5 MHz, which was not present when multipath and reflections were absent in the previous setup. In this experiment, one can see that the SISO algorithm only cancels 11 dB of the distortion, whereas the MISO algorithm cancels 18 dB of the distortion.

2.6 Rate of Convergence

In cases when a strong received signal is present, we increased the forgetting factor λ from 0.99 to 0.9999. This allows the adaptation to average more samples and therefore suppresses the effect of the strong received signal. Unfortunately, this also increases the convergence time and reduces the tracking capability of the RLS. This is illustrated for the case of no RX2 signal in Fig. 2.25, which shows the settling of the filter output over time for various λ . For a λ of 0.99, the filter settles within 2000 iterations. For a λ equal to 0.9999, the filter settles within 100 000 iterations, which is outside of the displayed range of Fig. 2.25. As derived in [22], the cancellation shows exponential convergence behavior, which slows down for higher λ . Given ADC sampling speeds in the range of 100 MHz, the algorithm settles within 20 μ s to 1 ms. The latter seems prohibitively slow. However, the RX power level changes, and the changes in the coupling between the two antennas due to hand movements are in the tens of milliseconds range. Furthermore, for strong received signal powers, the EVM is reasonably low so that one can accept the long convergence time in those cases. If required, the convergence rate likely can be increased by incorporating improved algorithms presented in [13,23] to dynamically change λ to ensure fast convergence rates and small residual distortion (EMSE).

2.7 Computational Effort

Since the cancellation needs to be performed in real time on a handset, it is critical to keep the computational cost and its associated power consumption low. The algorithm consists of two parts: the time alignment and adaptive filtering. The general time alignment, i.e., the search for J , is a common problem, and real time implementations have been demonstrated, e.g., using the early-late algorithm [15].

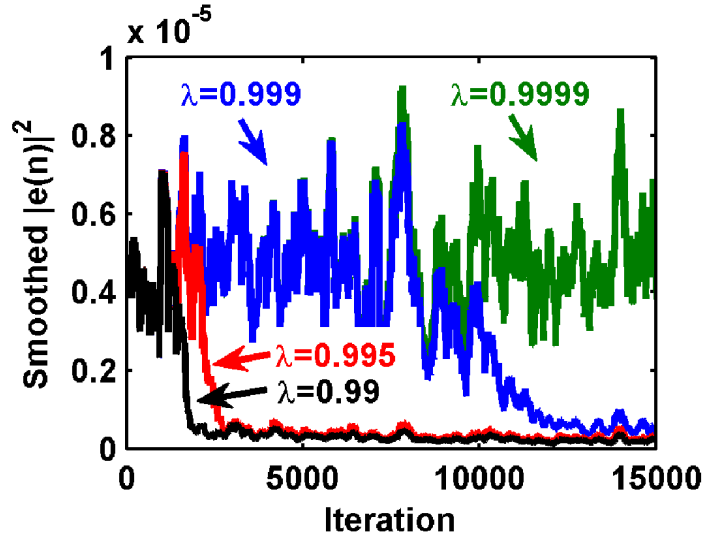


Figure 2.25: Convergence of MISO filter output for various λ .

Typically, the more computationally intense part of the algorithm is the adaptive distortion canceller, in particular, the RLS algorithm. It is commonly known that the RLS has faster convergence time than other adaptation algorithms, such as least-mean-square (LMS) or normalized-least-mean-square (NLMS) at the cost of higher number of multiplications. The RLS algorithm adapting for parameters has a complexity in the order of $O(L^2)$. In comparison, LMS and NLMS have the complexity of the order of $O(L)$. The majority of the multiplications and additions in the RLS algorithm are required to compute the P_i matrix. These computations can be reduced in cases where $x(n)$ the input vector to the RLS has a transversal structure such as

$$\phi_p = [x(n), x(n-1), \dots, x(n-L+1)] \quad (2.23)$$

where $x(n)$ is the current value at the reference input of the filter. When the data has this structure, one can take advantage of the fact that most entries of the P_i matrix are repeated at a shifted row and column, and the algorithm only needs to

update one row of the P_i matrix [24]. This reduces the computational effort from $O(L^2)$ to $O(L)$. These results can be directly applied to the SISO filter.

The MISO adaptive distortion canceller has $M \cdot N$ unknowns, which would imply that the computational expense is of the order of $O((M \cdot N)^2)$. However, one can take advantage of the structure of the input vector to the MISO distortion canceller shown in Fig. 2.9. One can rewrite the input vector ϕ_p and ϕ_{p+1} from (8) as

$$\phi_p = \begin{bmatrix} tx1^*(p) \\ \vdots \\ tx1^*(p - M + 1) \end{bmatrix} \cdot [tx2^2(p) \cdots tx2^2(p - N + 1)] \quad (2.24)$$

$$\phi_{p+1} = \begin{bmatrix} tx1^*(p + 1) \\ \vdots \\ tx1^*(p - M) \end{bmatrix} \cdot [tx2^2(p + 1) \cdots tx2^2(p - N)]. \quad (2.25)$$

If one compares (2.24) and (2.25), it becomes clear that $(N - 1) \cdot (M - 1)$ values are identical, which implies that one only needs to update $N + M - 1$ rows when computing P_i the matrix. Therefore, the complexity of the structured MISO RLS algorithm for this case can be reduced to $O([N \cdot M] \cdot [N + M - 1])$. Table 2.2 lists the order of complexity of the SISO and MISO distortion canceller for various filter lengths. In the case of the MISO filter, N equals M . For low orders of N and M , the complexity difference between the LMS and structure MISO RLS seems acceptable. For various antennas and antenna positions, N and M ranged from two to four in our experiments.

Table 2.2: Order of Complexity of SISO and MISO Filter

N=M	Complexity Order of SISO Filter			Complexity Order of MISO Filter		
	LMS	Structured RLS	RLS	LMS	Structured RLS	RLS
1	1	1	1	1	1	1
2	2	2	4	4	12	16
3	3	3	9	9	45	81
4	4	4	16	16	112	256
5	5	5	25	25	225	625

2.8 Conclusions

In transmitter RF front ends employing uplink carrier aggregation (UL CA), the nonlinearity of passive components such as switches and duplexers will create cross-modulation products. For certain band pairs, those products land in the receive band and significantly degrade the receiver sensitivity. An MISO adaptive distortion canceller is proposed and allows compensation of transfer functions for both transmitted signals and is robust against time alignment errors. The MISO algorithm has been successfully used in a realistic experiment setup to cancel the measured interference in the digital baseband without additional analog hardware by up to 20 dB. This allows the application of UL CA at full transmit power in problematic band pairs without the added expense of extremely linear passive components (antenna switches and duplexers).

Acknowledgments

Chapter 2 is mostly a reprint of the material as it appears in “All-Digital Cancellation Technique to Mitigate Receiver Desensitization in Uplink Carrier Ag-

gregation in Cellular Handsets”, *Transactions on Microwave Theory and Techniques*, Dec. 2013. This dissertation author was the primary author of this material.

Chapter 3

Analysis and Design of Stacked-FET Millimeter-Wave Power Amplifiers

3.1 Background: mm-wave Silicon PAs

The previous chapter focused on near term solutions for broader bandwidth communication by aggregating multiple smaller bands. The planned roll-out for carrier aggregation allows the combination of up to five 20-MHz channels providing a maximum bandwidth of 100 MHz [2]. However, to achieve that theoretical maximum is very challenging in practice. Recently, there has been a growing interest in using the millimeter-wave (mm-wave) bands in applications such as wideband terrestrial wireless communication, satellite radio, and automotive radar. The mm-wave band allow instantaneous bandwidth hundreds of MHz even for small fractional bandwidth systems. This has led to a research focus on the development of efficient mm-wave power amplifiers (PAs). Traditionally, compound semiconductor

MMICs were the preferred choice for such amplifiers. However, future applications are motivating the investigation of lower cost technologies. SiGe BiCMOS HBTs are attractive candidates for replacing III-V devices, since they have moderate breakdown voltages and thus can provide moderate amounts of power. CMOS technologies have traditionally not been favored for PA applications, despite the high cutoff frequencies of modern scaled CMOS devices. The main drawback of CMOS is the inability of field-effect transistors (FETs) to tolerate high voltage levels. Therefore, the power that one can obtain from a single FET is limited unless very wide FETs with low load impedances are used. However, this approach is highly sensitive to inherent device and interconnect parasitics and their resulting high losses. Furthermore, the required impedance-matching networks significantly lower the efficiency and obtainable bandwidth, and they are often not realizable with on-chip components. One strategy to overcome the problem of limited CMOS voltage range is based on series-connected (stacked) FETs. With appropriate biasing and loading, uniform voltage distributions can be obtained across the transistors. In principle, with K transistors, the overall structure can tolerate $K \cdot V_{max}$, where V_{max} can be chosen to be near the drain-source breakdown voltage of a single device. Variations of the stacking technique were applied in various technologies and frequencies over the last two decades [25–31].

In this chapter, the analysis of CMOS stacked-FET PAs is given with a particular focus on design considerations for mm-wave operation. An emphasis is set on the appropriate impedance between the stacked transistors. A theoretical framework based on phase detuning at the intermediate nodes is introduced and its effect on the output power and efficiency is studied. Furthermore, three tuning techniques are discussed to compensate for the phase detuning caused by the device parasitics.

In addition an updated sizing rule for the gate capacitances is derived, which considers the gate-drain device capacitance. The impact of the gate-drain capacitance has been ignored in the theory of previous publications [25–28,30–32]. In older technologies the gate-drain capacitance is relatively small compared to the gate-source capacitance. However, for mm-wave PAs in CMOS it is preferable to use 45-nm or smaller gate-length devices due to their fast switching speed. In these processes the relative size of the gate-drain capacitance to the gate-source capacitance is significantly larger, which makes it critical to include the gate-drain capacitance in the study of stacked-FET PAs for mm-wave operation.

After the sections focusing on the extension of stacked-FET theory, three Q -band PAs with two, three, and four stacked FETs implemented in a 45-nm silicon-on-insulator (SOI) process are presented, for which the combination of output power and efficiency are among the highest reported for Si FET technologies.

This chapter is organized in six sections. In Section 3.2, a brief summary of prior stacked-FET PAs is provided, the stacking concept is reviewed, and design tradeoffs studied with a 3-stack PA example. In Section 3.3, the appropriate intermediate node impedance is analyzed and different matching networks are compared. In section 3.4, the CMOS technology used and the design of 2-, 3-, and 4-stack amplifiers are discussed. Section 3.5 presents small- and large-signal measurement results of the amplifiers. Furthermore, a 2-stack PA with different intermediate node matching configurations is reported to highlight the importance of the intermediate node tuning at mm-waves. Section 3.6 summarizes the key results of this chapter.

3.2 FET Stacking Concept

Fig. 3.1 shows the schematic of a stacked-FET PA. The circuit is based on a series interconnection of a common-source (CS) transistor cascaded with common-gate-like transistors. The stacked FET configuration differs from a cascode, in which the gate of the common-gate transistor is grounded at the frequency of operation. Here, the gate of the common-gate-like transistor is connected to a finite impedance and experiences a voltage swing. Ideally, the drain voltages of the transistors add in phase while the drain current is constant through each transistor. The gate voltage swing in many stacked-FET PAs is controlled by introducing appropriate capacitances C_k at the gates of stacked transistors [27, 32, 33]. The series combination of C_k and the corresponding FET gate-source capacitance ($C_{gs,k}$) form voltage dividers that determine the gate voltages. In contrast with cascode amplifiers, this approach reduces the drain-gate and drain-source swings under large-signal conditions allowing reliable transistor operation under large aggregate voltage swings. However, the gain of the stacked-FET PA is lower than the gain of a cascode. Given the higher saturated output power and drain efficiency of stacked-FET PAs, especially if many transistors are connected in series, the reduced gain is an acceptable tradeoff. This technique has been demonstrated and discussed in detail for low frequency amplifiers [25–27, 30, 31, 33, 34].

3.2.1 Prior Work on Stacked-FET PAs

This subsection provides a short overview of some of the early and recent published stacked-FET amplifiers.

Shifrin et al. reported one of the first stacked-FET PAs in [25]. The authors

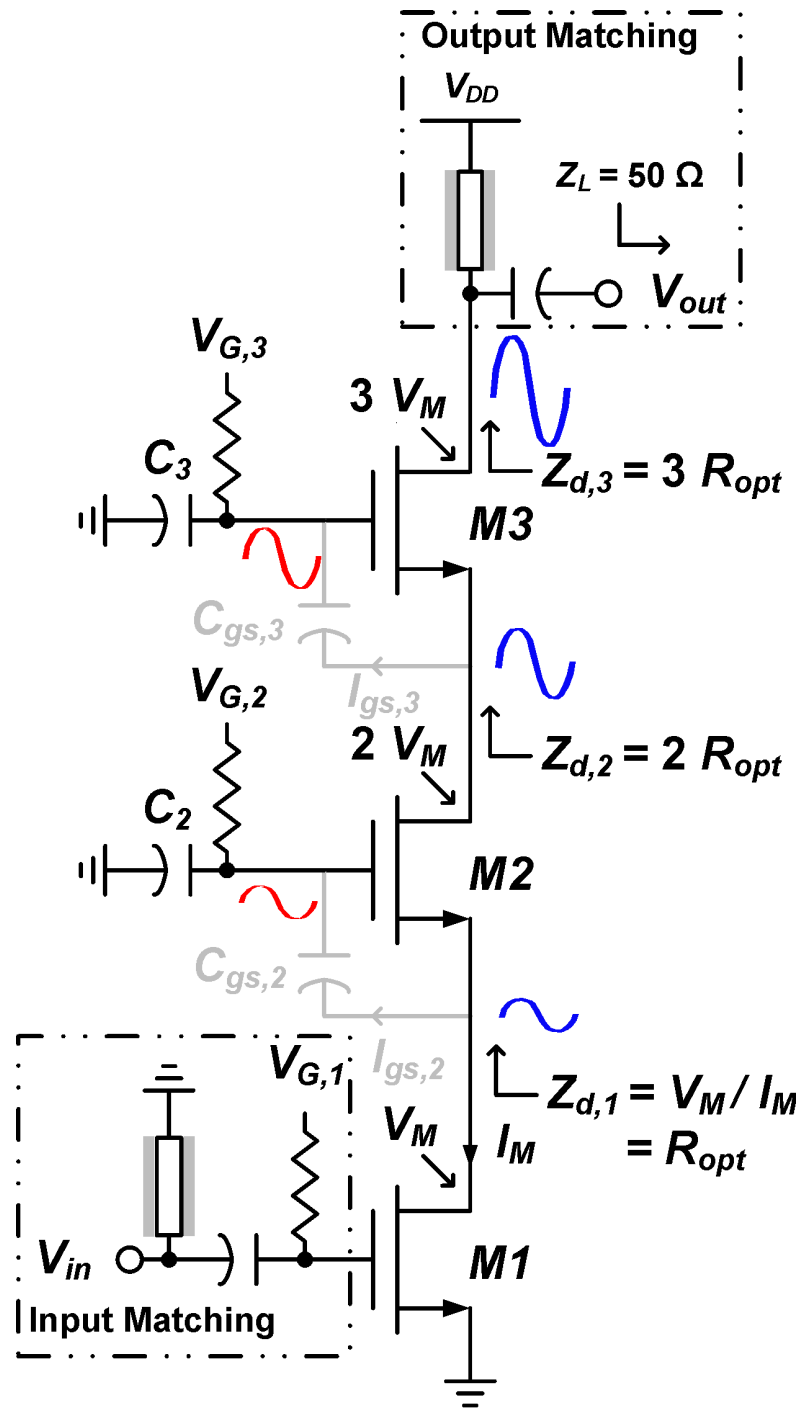


Figure 3.1: 3-stack PA schematic. The rectangular boxes used in the input and output matching network are coplanar waveguides (CPWs).

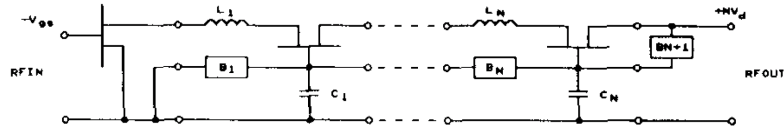


Figure 3.2: Hittite high power amplifier [25]

implemented 3-stack PA in a GaAs MESFET process, where each FET had a gate width of 8 mm. Fig. 3.2 shows the schematic of their amplifier. It is very similar to the amplifier illustrated in Fig. 3.1. It differs that it has DC biasing network for the gates derived from the power supply. More significant are the series inductors between the FETs. The authors did not discuss the purpose and the sizing of the series inductors; however, their relevance especially for mm-wave design is derived in detail in Subsection 3.3.1 and a sizing rule is provided in Subsection 3.3.2.

Another early adapter to the stacked-FET technique, although he refers to the structure as “High-Voltage/High-Power device (HiVP)”, was A. Ezzeddine. In [27] he presented a MESFET based stacked-FET amplifier similar to the one shown in Fig. 3.1. More recently Ezzeddine et al. presented an Universal High-Impedance, High-Voltage FET (UHiFET) suitable for microwave and millimeter wave operation [32]. Fig. 3.3(a) shows the schematic of the UHiFET. The main difference to the amplifier shown in Fig. 3.1 are the feedback capacitances in parallel with the series connected transistors. These capacitors act as negative capacitance looking up into the source of the stacked transistors [32]. This can be used to tune out the parasitic device capacitances present in the stack to ensure proper alignment of the voltage and current waveforms. Ezzeddine et al. derive sizing rules for $C_{d,k}$ and $C_{g,k}$; however, they ignore the effect of $C_{gd,k}$.

The previously discussed stacked-FET PAs require appropriately sized gate capacitances to ensure the proper operation. Fig. 3.3(b) shows an alternative approach, which uses transformers to feed the input signal to all the gates of

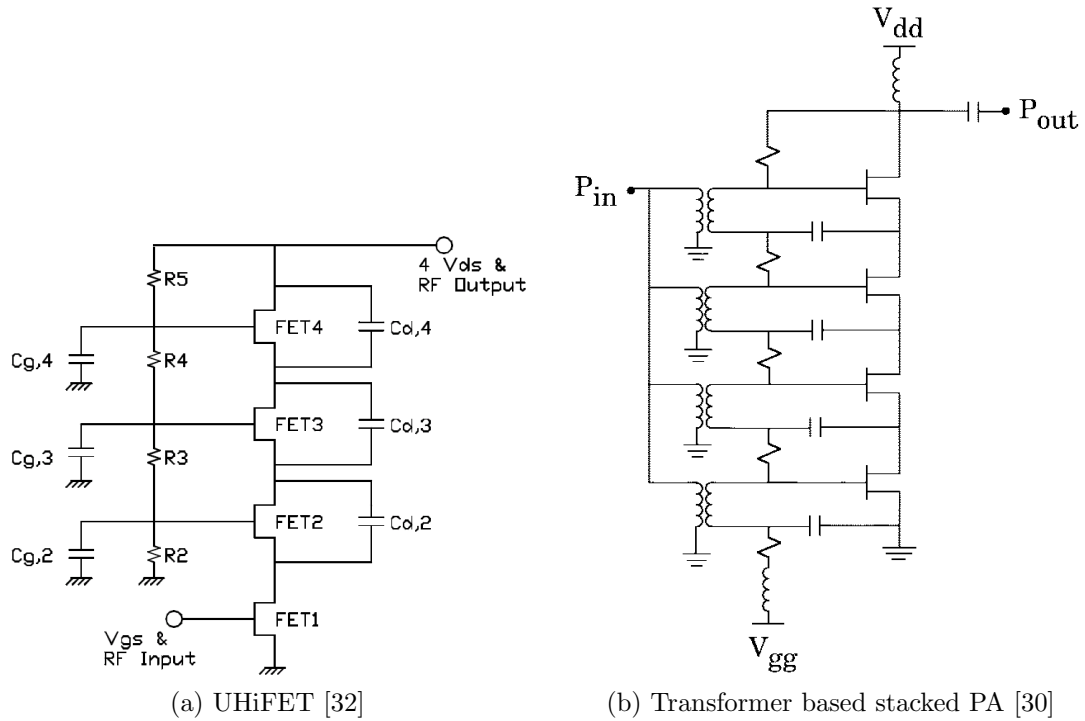


Figure 3.3: Prior stacked-FET PAs

the stacked-FETs. The main disadvantage of this technique is the large area requirements of the transformers [30].

3.2.2 Sizing of the Gate Capacitance C_k

One of the key design choices in the stacked-FET architectures similar to the one shown in Fig. 3.1 is the sizing of the gate capacitance. The impedance $Z_{d,k-1}$ seen at the drain of transistor $M(k-1)$ in Fig. 3.1, assuming linear operation and neglecting the FET small-signal output resistance and drain-source capacitance, is

$$Z_{d,k-1} = \frac{C_{gs,k} + C_k + C_{gd,k} (1 + g_{m,k} Z_{d,k})}{(g_{m,k} + sC_{gs,k}) (C_{gd,k} + C_k)} \quad (3.1)$$

where $C_{gd,k}$ is the gate-drain device capacitance and $g_{m,k}$ is the transconductance of the k th transistor in the stacked-FET PA [34].

In III-V technologies as well as in 180 nm or older CMOS processes the gate-drain capacitance is relatively small compared to gate-source capacitances and can be reasonably neglected. Ignoring C_{gd} and assuming the frequency of operation is a lot smaller than f_T (i.e. $g_m \gg C_{gs}$) (3.1) can be simplified to (3.2) [27, 33, 34].

$$Z_{d,k-1} \approx \frac{1}{g_{m,k}} \cdot \left(1 + \frac{C_{gs,k}}{C_k} \right) \quad (3.2)$$

To provide the optimum load line impedance to each of the transistors and ensure that the drain-source voltages are equally distributed among the stacked devices, the impedance $Z_{d,k}$ should be adjusted to $k \cdot R_{opt}$, where R_{opt} is the loadline impedance of a single device. This leads to the following sizing rule for the C_k 's

$$C_k = \frac{C_{gs,k}}{(k-1)g_{m,k}R_{opt} - 1}. \quad (3.3)$$

As previously mentioned, the gate-drain capacitance should not be ignored in scaled CMOS technologies. Since $sC_{gs,k}$ is still significantly smaller than $g_{m,k}$ even at mm-waves, (3.1) can be simplified to

$$\begin{aligned} Z_{d,k-1} &\approx \frac{C_{gs,k} + C_k + C_{gd,k} (1 + g_{m,k}Z_{d,k})}{g_{m,k}^2 (C_{gd,k} + C_k)} g_{m,k} \\ &- \frac{C_{gs,k} + C_k + C_{gd,k} (1 + g_{m,k}Z_{d,k})}{g_{m,k}^2 (C_{gd,k} + C_k)} sC_{gs,k}, \quad k = 2, 3, \dots, K. \end{aligned} \quad (3.4)$$

Assuming $Z_{d,k}$ is primarily real and chosen to be equal to $k \cdot R_{opt}$, the gate

Table 3.1: Evaluation of Load Impedances in Stacked-FET PA

Case #	k	C_{gd}	C_k (fF)	$Z_{d,k-1}$ desired	$Z_{d,k-1}$ sim.
1	2	0	66	13.75	13.8-j1.7
	3	0	26	27.50	27.3-j3.3
2	2	$0.2C_{gs}$	66	13.75	19.7-j6.2
	3	$0.2C_{gs}$	26	27.50	35.4-j7.2
3	2	$0.2C_{gs}$	1.79x66	13.75	12.9-j3.5
	3	$0.2C_{gs}$	1.79x26	27.50	26.7-j5.9

$$g_m = 216 \text{ mS}, C_{gs} = 130 \text{ fF}, C_{ds} = 0.1 C_{gs}, V_m = 1.1 \text{ V}, I_m = 80 \text{ mA}$$

capacitance C_k is determined by setting the real part of $Z_{d,k-1}$ to $(k-1) \cdot R_{opt}$.

$$C_k = \frac{C_{gs,k} + C_{gd,k} (1 + g_{m,k} R_{opt})}{(k-1)g_{m,k} R_{opt} - 1}, \quad k = 2, 3, \dots, K \quad (3.5)$$

Note that by setting $C_{gd,k}$ to zero in (3.5) one obtains the sizing rule for the C_k 's stated in (3.3) based on (3.2).

To verify the theoretical framework, calculations have been done for a 160- μm nMOS, which has been modeled as the capacitances C_{gs} , C_{gd} , C_{ds} , and a transconductance g_m , as specified in Table 3.1. Three cases are studied. First, if C_{gd} is zero, (3.5) leads to the nominal C_k calculated in Table 3.1. The simulation confirms that the impedance at the intermediate nodes is primarily resistive. In the second case, C_{gd} is $0.2 \cdot C_{gs}$ in the transistor model, but the C_k is not recalculated according to (3.5). Now, the resulting resistances at the drains have an error of approximately 45%. Finally, when the effect of C_{gd} is included in (3.5), the resulting C_k is approximately 1.79 larger. The resulting resistances are very close to the desired values.

Fig. 3.4 shows C_k normalized to $C_{gs,k}$ for various ratios of $C_{gd,k}/C_{gs,k}$ for a fixed gain $g_{m,k} R_{opt}$ of 3. The graph illustrates the importance of including $C_{gd,k}$

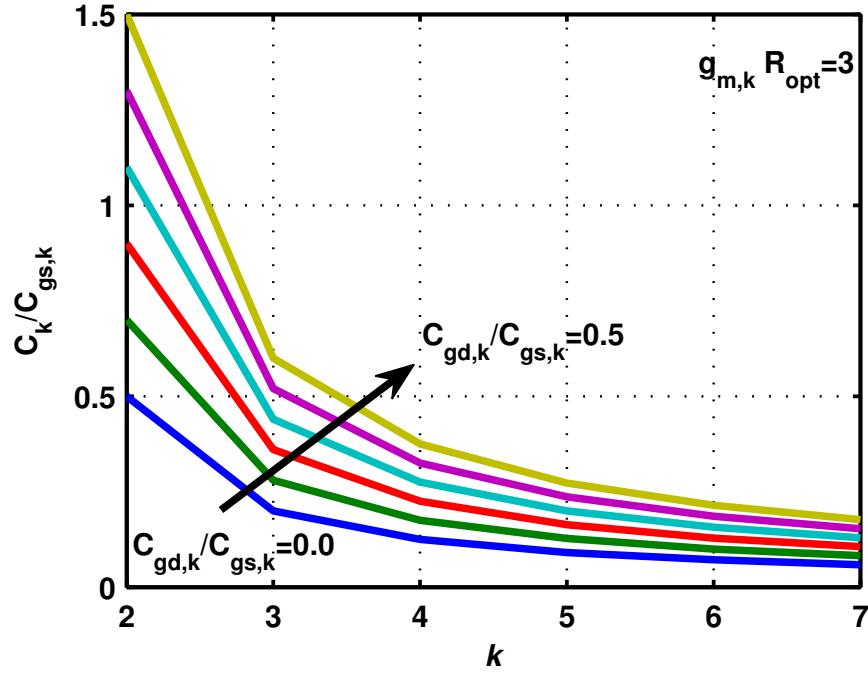


Figure 3.4: $C_k/C_{gs,k}$ for various $C_{gd,k}/C_{gs,k}$ for $g_{m,k} \cdot R_{opt} = 3$.

in the computation of the C_k 's. In the 45-nm CMOS SOI process used for the mm-wave PAs presented later in the chapter, the ratio of C_{gd} to C_{gs} is three, this means for example that C_2 would be incorrectly estimated by a factor of three if C_{gd} is ignored. Fig. 3.4 also highlights that the gate capacitance C_k become very small for high k , which makes its determination very sensitive to modeling errors. This poses one of the design challenges for stacking many transistors.

3.2.3 Voltage Distribution

A critical design consideration is the proper adjustment of the dc gate voltages for efficient and reliable operation. With a supply voltage much greater than the breakdown levels of the FETs, the gates of the stacked devices must be biased such that the dc and RF V_{gs} , V_{gd} , and V_{ds} voltages of each transistor are less than their respective breakdown voltages. In class-AB operation, the dc current

will increase with increasing RF input power P_{in} . When the dc gate voltages are fixed independent of P_{in} , the gate voltages are set considering the current levels for maximum P_{in} . If this is not considered properly, the source voltages of the top stacked FETs will experience too little voltage swing, which causes early breakdown of these devices and early compression of the CS device. As discussed in [33], the dc gate voltages for K -stacked amplifier should be set to

$$V_{G,k} = \left(\frac{k-1}{K} V_{DD} + V_{GS,k-sat} \right), \quad k = 2, 3, \dots, K \quad (3.6)$$

where $V_{GS,k-sat}$ is the dc value of the k th FET at saturation power level. As a result, the appropriate choices of C_k , $V_{G,k}$, and R_{opt} ensure that the drain-source breakdown voltage is not exceeded. However, this does not ensure that the gate-drain voltage swing does not exceed its breakdown limit. An additional constraint on the size of the transistors is required. In Appendix 3.A it is shown that, for optimal conditions, the gate-drain voltage is

$$V_{gd,k} = -\frac{1 + g_{m,k} R_{opt}}{g_{m,k} R_{opt}} V_{opt}, \quad k = 1, 2, \dots, K. \quad (3.7)$$

From (3.7), one can deduce that the peak gate-drain voltage magnitude equals $V_{opt} + |V_{gs}|$, where V_{gs} is the gate-source voltage needed to drive the RF current. If the transistor is too small for a given current, the sum of the two voltages may exceed the gate-drain breakdown voltage.

3.2.4 Benefits and Limitations of Stacking

For mm-wave PAs based on CS amplifiers in a scaled CMOS process, three factors limit the saturated output power: the transistor breakdown voltage; the

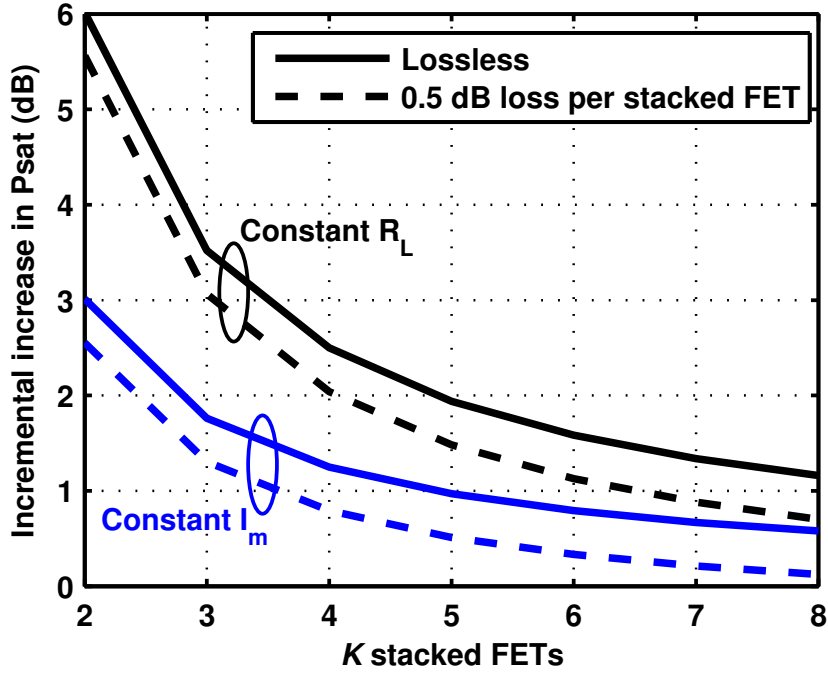


Figure 3.5: Incremental increase in P_{sat} of k th stacked FET.

maximum gate width for which reasonable gain is achieved; and the existence of a realizable matching network to the appropriate loadline impedance. For a CS amplifier in a scaled CMOS process, the third is often the primary limit. By stacking K FETs at a constant drain current, the output power increases by a factor of K and the required load impedance also increases by K . For a fixed loadline impedance R_L , stacking K FETs increases the current in each FET by K and the overall power increases by K^2 . However, the maximum transistor width is limited since the additional device parasitics decrease the transistor gain too severely. Fig. 3.5 shows the incremental increase of output power with K when the transistor width is scaled for constant R_L and when the transistor width is fixed for constant I_m .

Under constant R_L scaling, the increase in output power is substantial for the first few transistors and stacking to eight or more FETs is ideally worthwhile (solid lines in Fig. 3.5). However, the transistor gain is reduced significantly by

the parasitic resistance, capacitance, and inductance for large geometry FET layouts. Due to resistive losses in the transistor and phase misalignment of the drain voltages (discussed below), ideal power combining is limited. Assuming a constant combining efficiency of 90% for each transistor (i.e. 0.5 dB loss introduced by each successive stage) one obtains the “dashed” lines in Fig. 3.5. For this case of lossy combining, the fourth transistor still increases the output power by 2 dB for constant R_L scaling. For constant I_m , the power improvement is 1 dB. A fifth transistor increases the output power by only 0.5 dB. If one considers the marginal improvement of 2 dB as a criterion for adding an additional stage, stacking beyond four FETs is not worthwhile for the constant R_L case or stacking beyond three FETs is not worthwhile for the constant I_m case.

3.2.5 Comparison of Stacking to Other Power Combining Techniques

An appropriately tuned K -stack PA can be approximated as a single transistor with an input capacitance of

$$C_{in} = C_{gs,1} + C_{gd,1} (1 + g_{m,1} R_{opt}), \quad (3.8)$$

and from (3.36) one can show that the output matching network needs to appear as a negative capacitance equal to

$$\frac{Im\{Y_{load,opt}\}}{\omega} \approx -\frac{C_{ds,K}}{K} - \frac{C_{gd,K}}{K}, \quad (3.9)$$

which is not necessarily the same as C_{out} of the stacked-FET structure. The on-resistance of the K -stacked-FETs amplifier is also K times larger than the on-

resistance of each transistor. An alternative to stacking FETs would be to use CMOS devices with higher breakdown voltages. Good results have been demonstrated using extended drain devices [35]. These FETs have reduced f_T and are currently available in only a few processes. An alternative in most standard CMOS processes is the thick-oxide FET. These devices have higher breakdown voltages at the expense of higher intrinsic parasitic capacitance, and therefore, lower f_T and f_{MAX} . Fig. 3.6 shows the simulated f_{MAX} of two, three, and four stacked FETs. The peak f_{MAX} remains approximately constant in all three cases showing that stacking multiple devices does not reduce the effective gain of the composite FET. In the 45-nm SOI technology considered here, the thick-oxide FET has a breakdown voltage 1.5 times higher than the thin-oxide FET. A reasonable comparison for constant dc voltage supply suggests comparing three stacked thin-oxide FETs and two stacked thick-oxide FETs. In Fig. 3.6, the f_{MAX} of the stacked thick-oxide FETs is 30% lower than stacked thin-oxide FETs.

Other power combining techniques include radial combiners [36], combiners based on transformers [37], or Wilkinson combiners [38]. For example, K Wilkinson combiners are used to combine $K + 1$ equally sized amplifiers (typically with $\log_2 K$ stages). Fig. 3.7 shows the incremental increase in saturated output power per Wilkinson combiner for a lossless case and also when assuming a constant loss of 0.5 dB for each Wilkinson. Note that the increase in power per Wilkinson is identical to the increase in power per stacked transistor under constant I_m scaling.

These considerations indicate that for the highest output power one should use stacking in conjunction with other power-combining techniques. From Figs. 3.5 and 3.7, once the power added per additional stacked transistor is less than the power added per Wilkinson stage it is more efficient to use power combining rather than stacking. Other considerations such as the higher area requirements

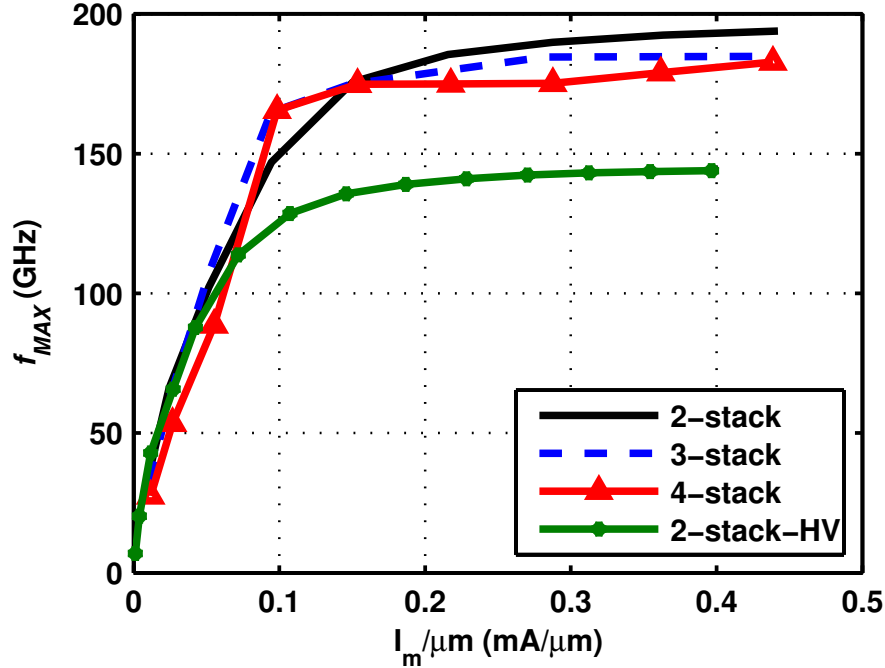


Figure 3.6: Comparison of f_{MAX} of two, three, and four stacked FETs using thin-oxide FETs and two stacked thick oxide / high-voltage (HV) FETs.

of Wilkinsons and other passive power combining techniques may also play a role in this design decision.

3.3 Complex Intermediate Node Matching

In Section 3.2, the theoretical framework was presented to explain how the intermediate node impedance was chosen to provide an appropriate loadline resistance. At low frequencies, $Z_{d,k}$ can be approximated as a resistance. However, the intermediate node impedances have a significant reactance at mm-wave frequencies caused by the transistor capacitances, as listed in Table 3.1. This reduces the efficiency for two reasons: (1) as illustrated in Fig. 3.1, part of the transistor RF current is flowing out through $C_{gs,2}$ [32] and other capacitances at the drain of $M1$ and does not reach the load; (2) the voltage waveforms are not phase aligned for

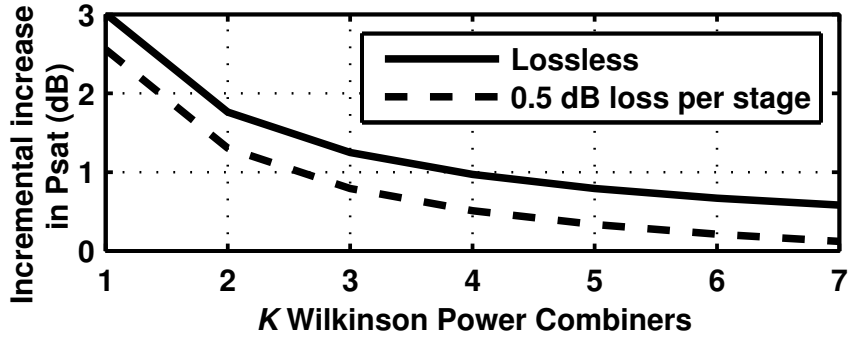


Figure 3.7: Incremental increase in P_{sat} of the k th Wilkinson combiner.

the highest swing at the top drain.

3.3.1 Optimal Complex Intermediate Node Impedance

The simplified small-signal model of stacked transistors shown in Fig. 3.8 is used to derive the optimal impedances at the drain of the FETs.

$$\begin{aligned}
 Y_{opt,k} &\approx \frac{1}{kR_{opt}} - \frac{s}{k}(C_{ds,k} + kC_{dsub,k} + C_{gd,k}) = \frac{1}{kR_{opt}} - \frac{s}{k}(C_{equiv,k}), \\
 k &= 1, 2, \dots, K
 \end{aligned} \tag{3.10}$$

Details of this derivation are presented in Appendix 3.A. This condition ensures that all drain-source voltages, as well as drain currents, are aligned, leading to highest output power and best efficiency.

In the previous section, the gate capacitance C_k was chosen to ensure that the stacked transistor presents the optimal load resistance. However, the optimal susceptance presented to the k th transistor should be inductive to tune out the capacitances at the drain of that transistor. However, the $(k + 1)$ th transistor instead represents a capacitive load. From Fig. 3.8, one can derive the admittance

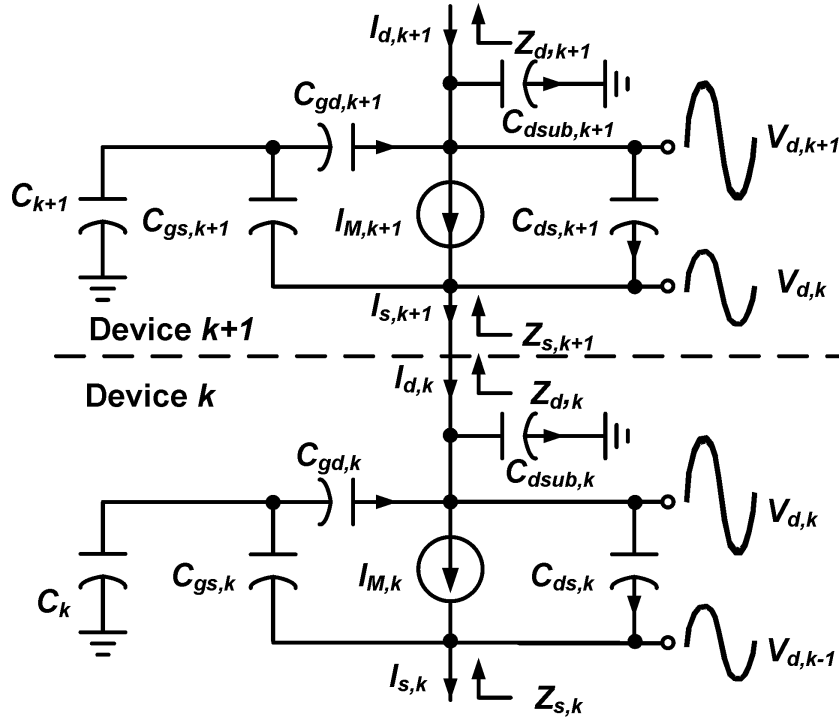


Figure 3.8: Simplified small-signal model of stacked transistors.

looking into the source of the $(k + 1)$ th stacked transistor as follows:

$$I_{s,k+1} = (g_{m,k+1} + sC_{gs,k+1}) V_{gs,k+1} + sC_{ds,k+1} V_{opt}, \quad (3.11)$$

$$Y_{s,k+1} = -\frac{g_{m,k+1} V_{gs,k+1} + sC_{ds,k+1} V_{opt}}{kV_{opt}} - \frac{sC_{gs,k+1} V_{gs,k+1}}{kV_{opt}},$$

$$k = 1, 2, \dots, K - 1 \quad (3.12)$$

where V_{opt} is the optimal voltage at the drain of $M1$. With (3.34) and (3.35), one

can simplify (3.12) to

$$Y_{s,k+1} = \frac{1}{kR_{opt}} - \frac{sC_{ds,k+1}}{k} + \frac{sC_{gs,k+1}}{kg_{m,k+1}R_{opt}}, \quad (3.13)$$

$$\begin{aligned} Z_{s,k+1} &\approx kR_{opt} - k \frac{R_{opt}}{g_{m,k+1}} s (C_{gs,k+1} - g_{m,k+1}R_{opt}C_{ds,k+1}) \\ k &= 1, 2, \dots, K-1. \end{aligned} \quad (3.14)$$

From (3.10) and (3.13), one can show that the phase angle of the impedance presented by the $(k+1)$ th transistor to k th transistor is

$$\Phi_{s,k+1} = \arctan \left(\omega \left(\frac{C_{gs,k+1}}{g_{m,k+1}} - C_{ds,k+1}R_{opt} \right) \right). \quad (3.15)$$

However, the optimal load at the k th drain should have a phase of

$$\Phi_{opt,k} = \arctan(-\omega(C_{eqv,k})R_{opt}) \quad (3.16)$$

to tune out the effect of capacitances at that node with $C_{eqv,k}$ defined in (3.10). Therefore, additional matching components are required to phase rotate $Y_{s,k+1}$ by

$$\Phi_k = -\Phi_{opt,k} + \Phi_{s,k+1}. \quad (3.17)$$

This ensures that the drain voltage and current from the k th transconductance are phase aligned. If the phase error is corrected at each intermediate node, the drain voltages optimally add for the highest power. In Appendix 3.B, the relationship between phase errors Φ_k , the *power combining efficiency of stacking* and

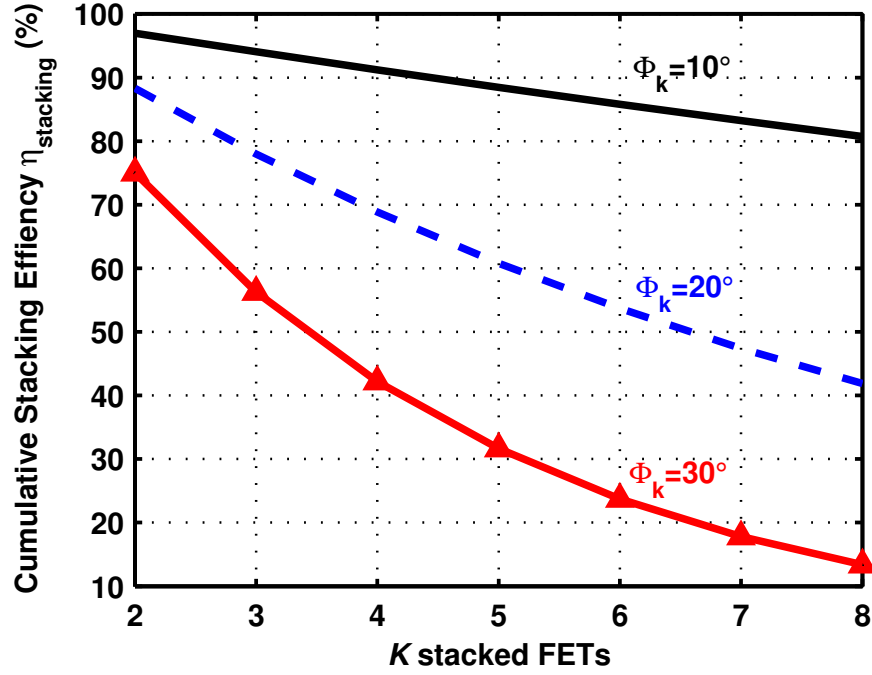


Figure 3.9: Cumulative stacking efficiency for various phase misalignments.

the output power is derived for the case where there are phase errors Φ_k

$$\eta_{stacking} \approx \left(\prod_{k=1}^K \cos(\Phi_k) \right)^2 \quad (3.18)$$

$$P_{out\ K-stack} \approx \eta_{stacking} P_{out\ ideal\ K-stack} \quad (3.19)$$

where K is the number of stacked transistors and $P_{out\ ideal\ K-stack}$ is the output power of the K -stack PA if all currents and voltage are optimally aligned. Even modest degrees of misalignment cause an appreciable degradation of the efficiency as seen in Fig. 3.9. This phase alignment effect presents another obstacle to the effectiveness of stacking many FETs.

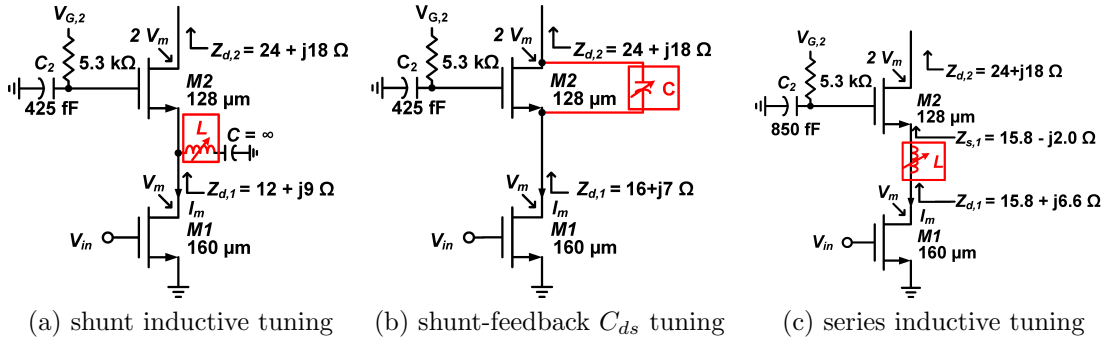


Figure 3.10: 2-stack PA schematic with different intermediate node tuning techniques.

3.3.2 Optimal Intermediate Node Impedance Matching

Without any reactive tuning, the efficiency and power reduction are significant at mm-wave frequencies. To achieve the proper complex impedance between the transistors, additional tuning elements are needed for optimal performance. In recent literature, three different circuit approaches to implement the proper complex intermediate node have been presented. Fig. 3.10(a) illustrates a shunt L tuning technique [39]. The 2-stack PA in Fig. 3.10(b) shows a shunt-feedback drain-source capacitor tuning technique [32] and Fig. 3.10(c) shows a 2-stack PA using a series inductance between the two transistors [40]. The inductances directly tune out the parasitic capacitances at the intermediate node either in a series LC or parallel LC sense. The shunt-feedback C_{ds} approach achieves the same effect because the capacitance across the transistor effectively appears as a negative capacitance, as seen in (3.13). In this section, the required shunt-feedback C_{ds} , series L , and shunt inductance are derived and their performances are compared.

Shunt Inductance

By appropriate choice of C_k , it is ensured that $Re\{Y_{s,k+1}\}$ equals the desired $Re\{Y_{opt,k}\}$. By adding a shunt inductance L_k between the two transistors, one can

ensure optimal phase alignment [39, 41]. Solving

$$\text{Im}\{Y_{s,k+1}\} + \frac{1}{sL_k} = \text{Im}\{Y_{opt,k}\}, \quad k = 1, 2, \dots, K-1 \quad (3.20)$$

for L_k one finds

$$\begin{aligned} \frac{1}{L_k} &= \frac{\omega^2 (C_{ds,k} - C_{ds,k+1})}{k} + \frac{\omega^2 C_{gs,k+1}}{kg_{m,k+1}R_{opt}} + \omega^2 \left(\frac{C_{gd,k} + kC_{dsub,k}}{k} \right) \\ k &= 1, 2, \dots, K-1. \end{aligned} \quad (3.21)$$

The first term shows that for equally sized transistors the drain-sources capacitances cancel. The second term is the capacitive load of the $(k+1)$ th transistor. Its effect is reduced by the voltage gain across the transistor. The third term relates to the gate-drain and drain-substrate capacitance of the k th transistor.

Shunt-feedback Drain-Source Capacitance

The second tuning technique using a capacitance C_d in parallel to C_{ds} has been proposed by A. Ezzeddine *et al.* [32]. From (3.22), one can observe that increasing the drain-source capacitance along the stacked transistors ensures proper phase alignment

$$\begin{aligned} (C_{ds,k+1} + C_{d,k+1}) &= (C_{ds,k} + C_{d,k}) + \frac{C_{gs,k+1}}{g_{m,k+1}R_{opt}} + kC_{dsub,k} \\ k &= 1, \dots, K-1. \end{aligned} \quad (3.22)$$

Since $C_{d,1} = 0$ one can deduce that $C_{d,k+1}$ equals:

$$\begin{aligned} C_{d,k+1} &= C_{ds,k} - C_{ds,k+1} + k \frac{C_{gs,k+1}}{g_{m,k+1}R_{opt}} + kC_{gd,k} + k^2C_{dsub,k}, \\ k &= 1, 2, \dots, K-1. \end{aligned} \quad (3.23)$$

Since the effective drain-source capacitance is increased by this technique, the required inductive tuning at the top drain needs to be higher than when the shunt inductive tuning technique is used. This may make the output match more challenging.

Series Inductance

From (3.10), the desired series impedance is

$$Z_{opt,k} = kR_{opt} \frac{(1 + sC_{eqv,k}R_{opt})}{1 + (\omega C_{eqv,k}R_{opt})^2}, \quad k = 1, 2, \dots, K. \quad (3.24)$$

It is noteworthy that the term $(\omega C_{eqv,k}R_{opt})^2$ at mm-wave frequencies may not be much smaller than 1 and should not be neglected. In such a case the desired resistance is a multiple of a scaled version of R_{opt} .

$$R'_{opt,k} = \frac{R_{opt}}{1 + (\omega C_{eqv,k}R_{opt})^2} \quad (3.25)$$

This leads to an updated equation for the gate capacitance.

$$C'_{k+1} = \frac{C_{gs,k+1} + C_{gd,k+1} (1 + g_{m,k+1}R'_{opt,k})}{kg_{m,k+1}R'_{opt,k} - 1} \quad k = 1, 2, \dots, K - 1 \quad (3.26)$$

C'_k for the series inductive tuning case is larger than C_k in the shunt inductive tuning case since $R'_{opt,k}$ is smaller than R_{opt} . From (3.14), (3.24) and (3.25) one can determine the required series inductance as

$$\begin{aligned} L_k &\approx kR'_{opt,k} \frac{C_{gs,k+1}}{g_{m,k+1}} - kR'_{opt,k}R'_{opt,k}C_{ds,k+1} + kR'_{opt,k}R_{opt}C_{eqv,k} \\ k &= 1, 2, \dots, K - 1. \end{aligned} \quad (3.27)$$

Table 3.2: Reactive Intermediate Node Tuning

Case #	k	C_k (fF)	$L_{k-1} / C_{d,k}$ (pH) / (fF)	$Z_{d,k-1}$ desired	$Z_{d,k-1}$ sim.
shunt L	2	118.1	159	13.75	13.78
	3	47.1	318	27.50	27.54
shunt C_{ds}	2	118.1	78	13.75	13.77
	3	47.1	157	27.50	27.51
series L	2	118.1	14	13.75	14.10
	3	47.1	29	27.50	28.02
	2	123.3	14	13.75	13.75
	3	48.6	29	27.50	27.59

$$g_m = 216 \text{ mS}, C_{gs} = 130 \text{ fF}, C_{ds} = 0.1 C_{gs}, V_m = 1.1 \text{ V}, I_m = 80 \text{ mA}$$

The first two terms represent a series inductance tuning out the capacitive loading caused by the $(k + 1)$ th transistor. The third term tunes out the effective drain capacitance of the k th transistor.

3.3.3 Verification of Intermediate Node Matching Analysis

The analytical expressions have been verified by simulation using a 3-stack PA with a linearized transistor model with parameters specified in Table 3.2. The values for the shunt inductance, shunt-feedback drain-source capacitance, and series inductance are selected based on (3.21), (3.23), and (3.27), respectively. The admittance at the top drain is set according to (3.10). Using the shunt tuning elements ensures that the first and second transistors are appropriately loaded. However, for the series inductive case, one should note that even though the reactance is tuned out appropriately, the resistance is not. To correctly tune the circuit, the gate capacitance values C_k are adjusted based on (3.26), leading to appropriate load resistance for $M1$ and $M2$, as shown in the last two rows of Table 3.2.

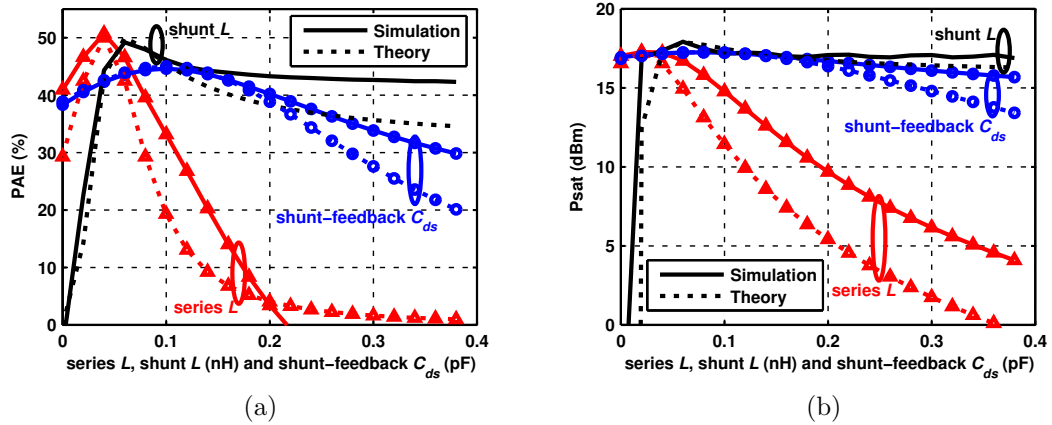


Figure 3.11: PAE (a) and P_{out} (b) for $P_{in} = 9$ dBm using series L , shunt L , and shunt-feedback C_{ds} intermediate node tuning.

Fig. 3.11 show the simulated PAE and P_{out} of a 2-stack using the different lossless tuning elements for a 2-stack PA (solid lines). It also shows the theoretically predicted degradation of PAE and P_{out} based on the misaligned phase by inappropriate tuning as described in (3.18) and (3.19) (dotted lines). The theoretical prediction of PAE and P_{sat} in the shunt L and shunt-feedback C_{ds} case are in very good agreement with simulation. The theoretical prediction of PAE and P_{sat} in the series L case provides reasonable agreement as well. When the series inductance is used as tuning element, both the phase and admittance presented to the current generator changes. However, (3.18) and (3.19) only consider the PAE and P_{out} degradation due to phase misalignment. In Section 3.5, this significant efficiency difference for different inductance values is confirmed in measurements for the shunt inductance technique.

3.3.4 Comparison of Intermediate Node Matching Techniques

As shown in Fig. 3.11, all three circuits achieve approximately the same peak output power. The series L and shunt L approach achieve approximately the same peak PAE of 50%, while the shunt-feedback C_{ds} technique achieved a slightly lower peak PAE of 45%. It is important that each technique be appropriately tuned. For example, if the shunt inductance is too small, it effectively creates a short to ground decreasing the efficiency significantly. If the shunt inductance is too large, it effectively acts as an open and does not tune out the losses through the capacitances $C_{gs,2}$, $C_{ds,1}$, and $C_{gd,1}$.

While the simulation indicates comparable performance when each technique is tuned appropriately, there are some practical differences between the tuning techniques. First, the required gate capacitance C_k is larger when the series inductance is used, which reduces the gate swing and may slightly increase the gate-drain voltage. This would slightly reduce P_{sat} and potentially the peak PAE in cases where the gate-drain breakdown voltage is limiting the reliable operation range. Second, the capacitive tuning technique requires a larger inductive impedance at the top drain to compensate for the additional capacitive loading. This might make the output matching more challenging. Furthermore, the efficiency benefits are sensitive to model accuracy, as one can see in Fig. 3.11. The shunt inductive tuning seems the least sensitive to mistuning. However, both the series L and shunt-feedback C_{ds} techniques have one advantage over the shunt L tuning technique. These tuning elements according to (3.23) and (3.27) are frequency independent, making them suitable for broadband amplifiers. It is noteworthy that the different tuning techniques present different harmonic terminations.

However, in simulation, no significant differences in output power and efficiency were observed.

3.4 Technology and Amplifier Implementation

3.4.1 45-nm CMOS SOI Technology

The stacked-FET PAs were implemented in a 45-nm CMOS SOI process with 11 metal layers. The top metal layer is a $2.2\text{-}\mu\text{m}$ thick aluminum metal layer. The capacitance of the floating source and drain nodes are reduced by the buried oxide, which reduces the losses due to capacitive coupling to the $13\ \Omega\text{-cm}$ silicon substrate [28]. The floating bodies of the transistors are particularly beneficial for stacked-FET PAs since these devices do not “suffer” from the body effect as occurs in bulk CMOS. The body effect would significantly degrade the transconductance of the transistors, especially in cases where three or more devices are stacked [42].

3.4.2 PA Implementation

Figs. 3.12-3.14 show the schematics of the designed 2-, 3-, 4-stack PAs. The inputs are matched to $50\ \Omega$ using an L -match consisting of a series CPW and shunt metal-finger capacitor. The latter offers roughly $1.3\ \text{fF}/\mu\text{m}^2$ with Q ranging from 20 to 30 at 45 GHz. The load impedances are chosen to maximize efficiency and allow feasible output matching networks using on-chip components. The resistance at the fundamental frequency of the load impedance Z_L should approximately correspond to the loadline resistance while the imaginary part of Z_L tunes out the capacitance of the top transistor as stated in (3.10) for optimal phase alignment. If the device sizes are chosen appropriately, one can ensure that the optimum loadline impedance is close to the highlighted area in Fig. 3.15 [43]. By

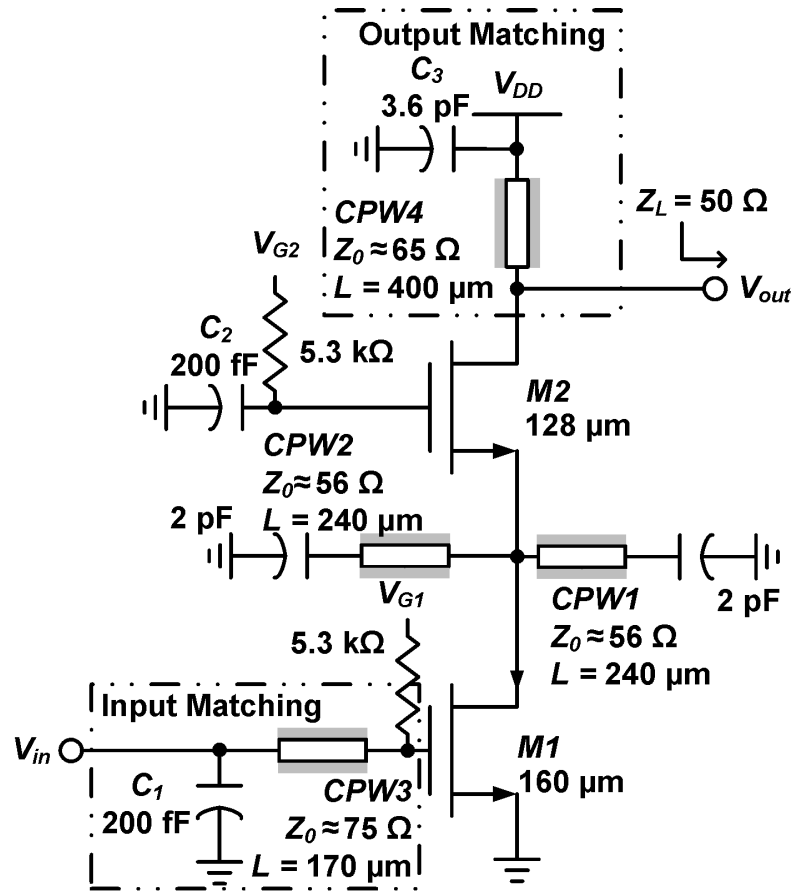


Figure 3.12: Schematic of 2-stack PA with shunt tuning elements between the two transistors.

changing the length of a shorted shunt stub one can transform a 50- Ω impedance to anywhere on the line. At the same time, this transmission line connects the amplifier to the supply and minimizes the losses at the output of the amplifier.

A shunt tuning element between the first two transistors is used in all amplifiers. However, due to layout area constraints, a series tuning element was used between the second and third device in the 3- and 4-stack PA. Fig. 3.16 shows the chip micrograph of the 3-stack amplifier. It occupies an area of 600 μm x 500 μm including pads. The 2- and 4-stack amplifiers have a similar layout and occupy approximately the same area.

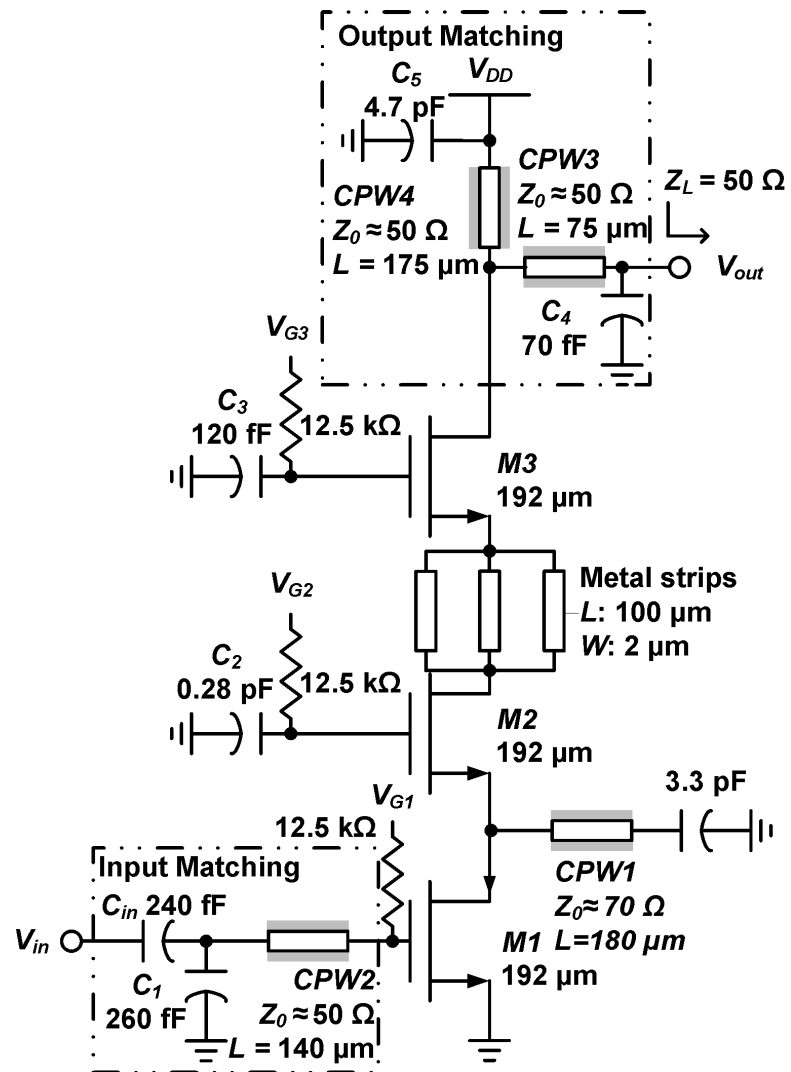


Figure 3.13: Schematic of 3-stack PA with shunt tuning element between $M1$ and $M2$ and series tuning inductance between $M2$ and $M3$.

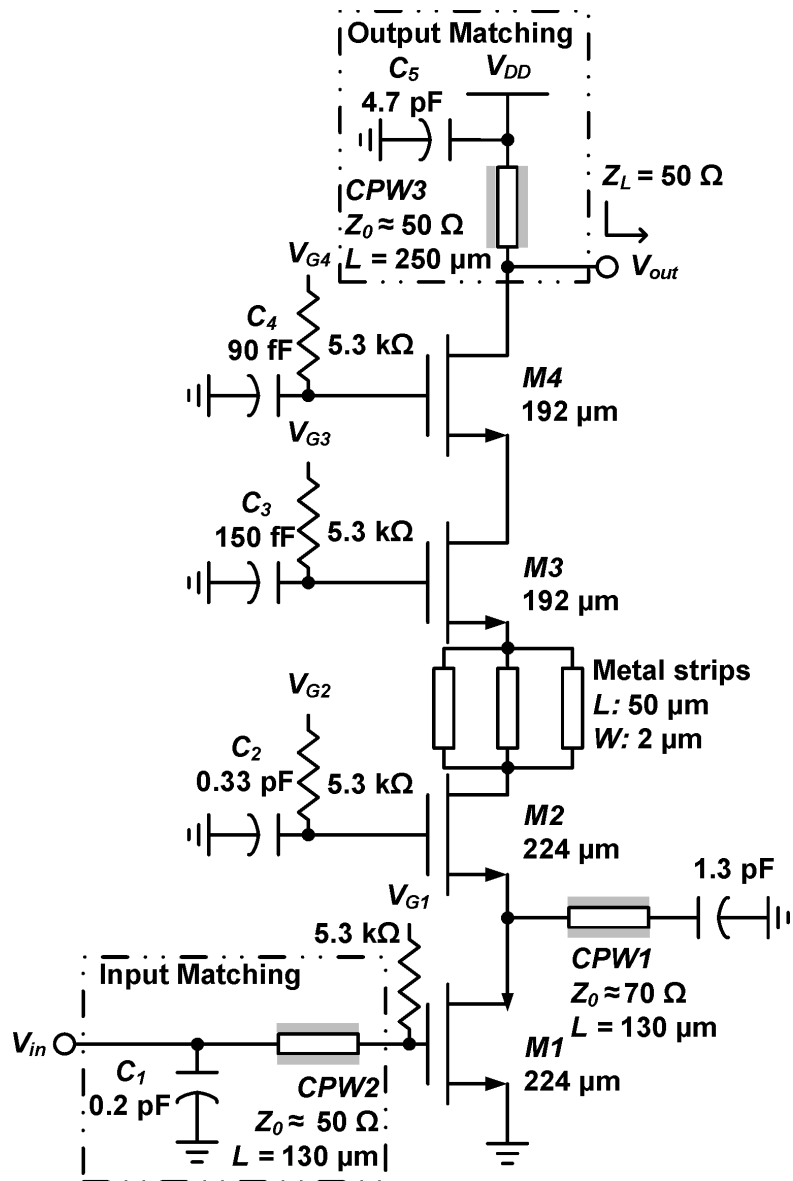


Figure 3.14: Schematic of 4-stack PA with shunt tuning element between $M1$ and $M2$ and series tuning inductance between $M2$ and $M3$.

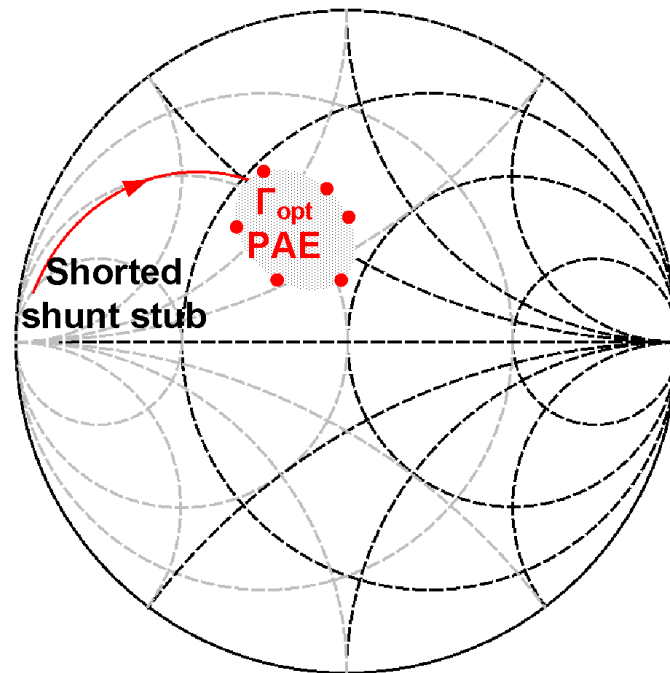


Figure 3.15: 50-Ω load and pad capacitance are transformed by a shunt stub (solid line) to a load impedance for optimal PAE inside the highlighted region.

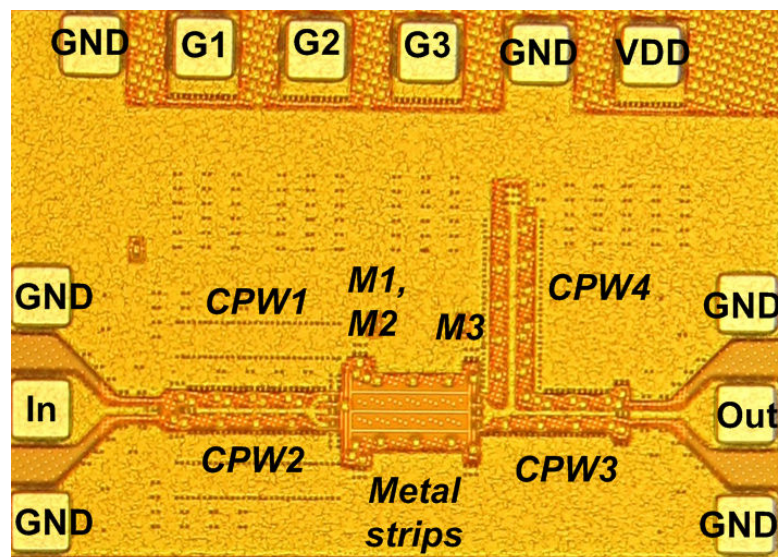


Figure 3.16: Photomicrograph of 3-stack PA occupying 0.6 mm x 0.5 mm including pads.

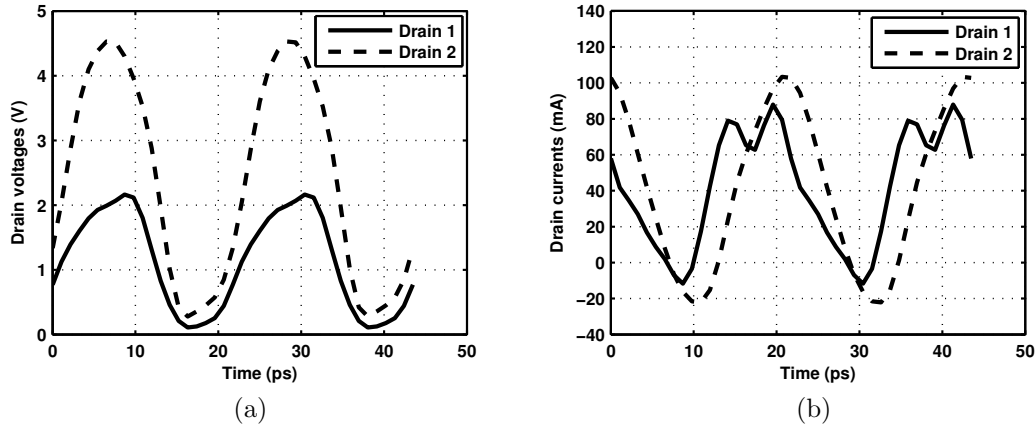


Figure 3.17: Simulated drain voltages (a) and drain currents (b) of 2-stack PA from Fig. 3.12 without *CPW2*.

Harmonic Terminations

In conventional PA design, waveform engineering using appropriate harmonic loads provides the highest performance. However, our simulations showed that there is very little benefit to add any intentional harmonic impedance control at mm-waves for two reasons. First, the output capacitance of the transistors already present low impedances at 90 GHz and 135 GHz. Therefore, it is very difficult to build appropriate harmonic terminations. Second, the losses in these matching elements would negate the efficiency improvements. Nonetheless, appropriate biasing of the transistors shapes the current waveforms to contain higher harmonic content and slightly improve efficiency. Fig. 3.17(a) shows the simulated drain voltage waveforms of a 2-stack amplifier; the waveforms are almost sinusoidal. Fig. 3.17(b) shows the current waveform at the drains. The drain current of the bottom transistor has a significant third harmonic due to the device non-linearity. However, the current at the top drain is almost sinusoidal since the high frequency components have been filtered out by the parasitic capacitances of the two transistors.

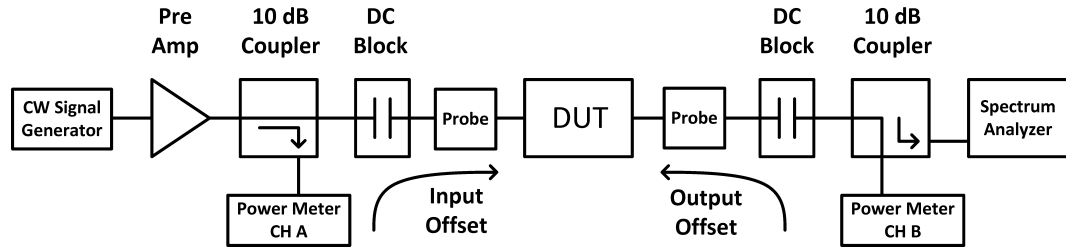


Figure 3.18: Large-signal measurement setup.

3.5 Experimental Results

3.5.1 Measurement Setups

S-parameters were measured on-wafer with ground-signal-ground probes and an Agilent E8361A network analyzer. Off-chip calibration has been performed up to the probe tips.

Fig. 3.18 shows a diagram of the large-signal measurement setup. The input power and output power are separately measured by Agilent N8487A power meters and losses of the input- and output-fixture, as well as the probes, are experimentally determined and de-embedded. An Agilent PSA 4448A spectrum analyzer monitors the output of the device-under-test (DUT) to ensure that no in-band or out-of band oscillations are present.

3.5.2 Intermediate Node Matching

The relevance of intermediate node matching has been experimentally studied on three versions of a 2-stack PA, as shown in Fig. 3.12: first with both shunt *CPW*s 1 and 2, second with one shunt *CPW*, and third without *CPW*1 and *CPW*2. The first case corresponds to a small shunt inductance, the second case to a larger shunt inductance and the third to no shunt inductance.

The large-signal response has been measured for comparable bias condi-

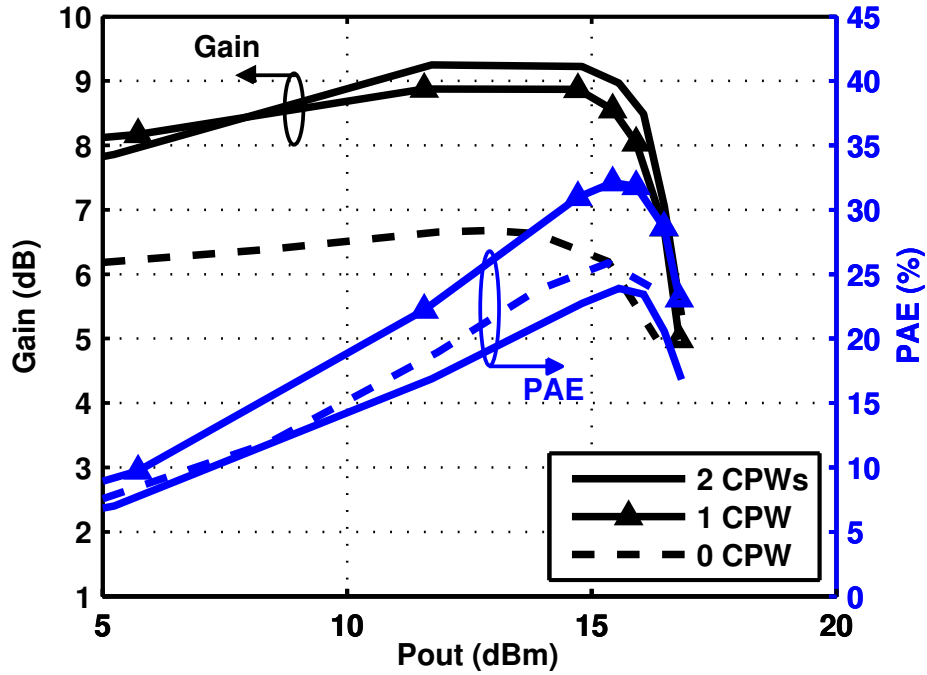


Figure 3.19: Measured gain and PAE as a function of output power at 46 GHz for the 2-stack PA with two shunt *CPW*s, with one shunt *CPW*, and no shunt *CPW* biased at: $V_{G,1} \sim 0.2$ V, $V_{G,2} = 1.8$ V, $V_{DD} = 2.8$ V, $I_{DC} = 8$ mA.

tions. Fig. 3.19 shows the gain and PAE response of the three amplifiers at 46 GHz. By comparing no tuning element to one *CPW*, one can see an increase of PAE from 26% to 32%, which conclusively shows the effectiveness of the shunt tuning technique. Applying two *CPW*s reduced the efficiency from 26% to 24%. The decrease in performance is related to additional losses in the *CPW* and mistuning of the impedance at that node. In Fig. 3.20, the efficiency and saturated power are relatively constant over frequency. It is noteworthy that the relationship between the amplifiers remains the same regardless of frequency, where the best performance is consistently achieved by the amplifier with one shunt *CPW*.

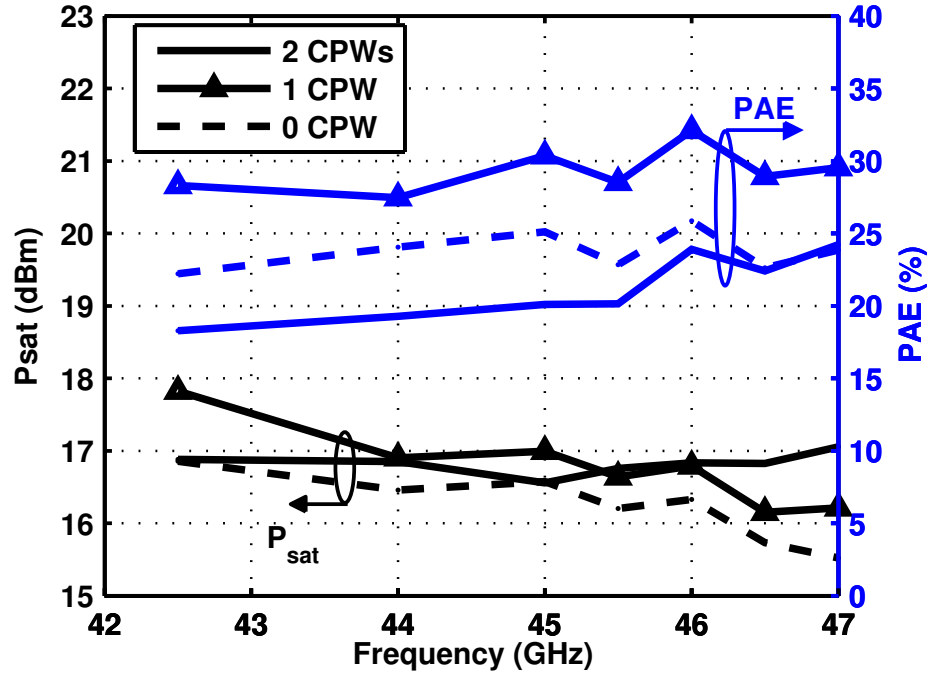


Figure 3.20: Measured PAE and P_{sat} over frequency for 2-stack PA with two shunt CPWs, with one shunt CPW and no shunt CPW.

3.5.3 Comparing 2-, 3-, and 4-Stack PAs

Small-signal measurement results of the amplifiers of Fig. 3.12-3.14 are shown in Fig. 3.21. The amplifiers are biased such that the quiescent currents are equal for the three PAs. The 2-stack and 4-stack amplifiers are slightly mistuned at input and output. The 3-stack amplifier mistuning is more pronounced and is probably related to output capacitances modeling. The 2-, 3- and 4-stack amplifiers respectively have gains of 9.6 dB at 47 GHz, 8.6 dB at 53 GHz, and 10.6 dB at 45 GHz.

Large-signal measurement results versus output power are shown in Fig. 3.22. The best performance for the 2-stack and 4-stack PAs were measured in class-A / AB regime. The peak gain is approximately 9.4 dB for both amplifiers. The best PAE for the 3-stack amplifier was observed when operating closer to the class B regime. Its peak gain is 8.9 dB. The 2-stack has a peak PAE of 32.7% at

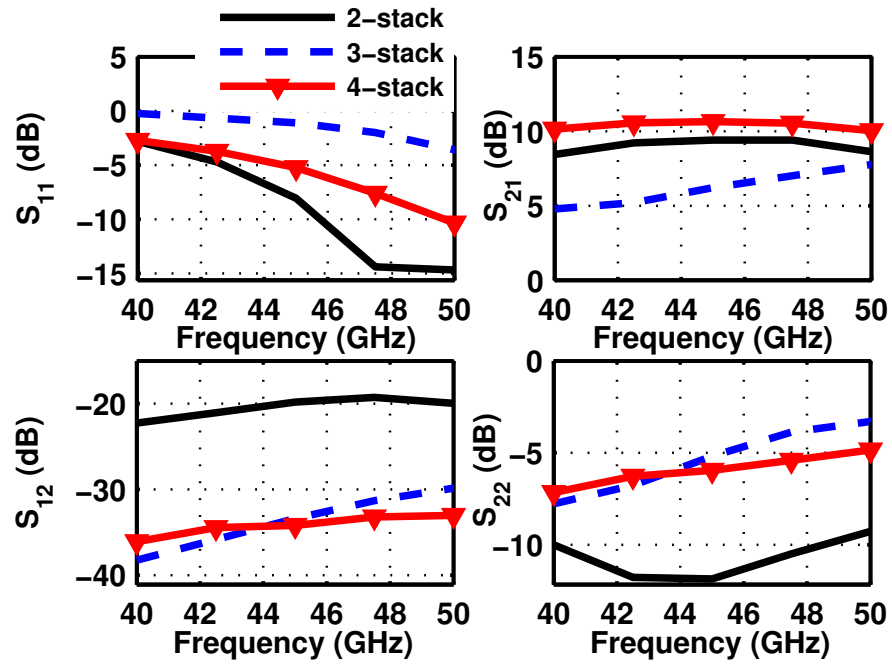


Figure 3.21: Measured S-parameter for 2-stack, 3-stack, 4-stack PA;
 2-stack: $V_{G,1}=0.3$ V, $V_{G,2}=1.6$ V, $V_{DD}=2.5$ V;
 3-stack: $V_{G,1}=0.2$ V, $V_{G,2}=1.7$ V, $V_{G,3}=2.5$ V, $V_{DD}=3.5$ V;
 4-stack: $V_{G,1}=0.3$ V, $V_{G,2}=1.7$ V, $V_{G,3}=2.7$ V, $V_{G,4}=4$ V, $V_{DD}=5$ V.

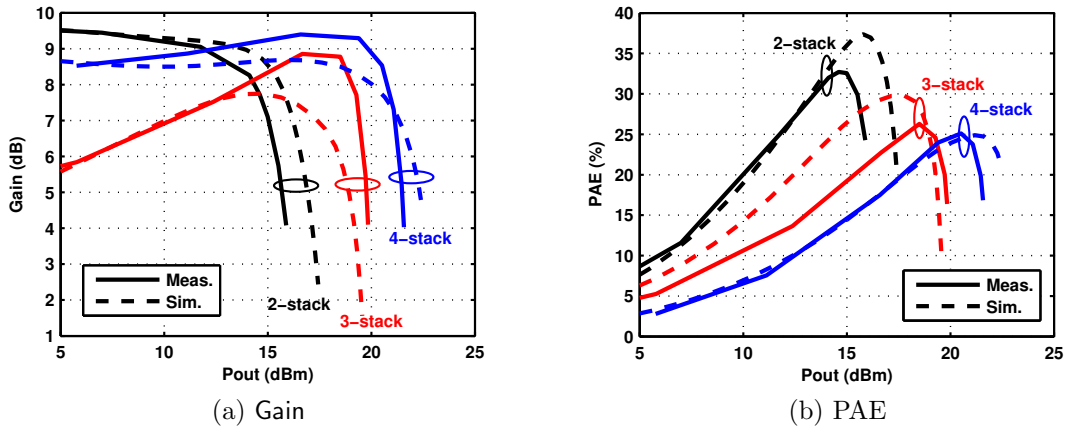


Figure 3.22: Measured gain and PAE versus Pout for 2- and 3-stack PA at 46 GHz and 4-stack PA at 41 GHz; 2-stack PA: $V_{G,1}=0.3$ V, $V_{G,2}=1.6$ V, $V_{DD}=2.5$ V; 3-stack PA: $V_{G,1}=0.2$ V, $V_{G,2}=1.7$ V, $V_{G,3}=2.5$ V, $V_{DD}=3.5$ V; 4-stack PA: $V_{G,1}=0.3$ V, $V_{G,2}=1.7$ V, $V_{G,3}=2.7$ V, $V_{G,4}=4.0$ V, $V_{DD}=5.0$ V.

14.6 dBm, the 3-stack has a PAE of 26.3% at 18.5 dBm and the 4-stack achieves a peak PAE of 25.1% at 20.5 dBm. The 2- and 3-stack PA were measured at 46 GHz and the 4-stack at 41 GHz. The measurement results are in good agreement with simulations. Accurate modeling of the source and drain interconnects and associated inductance was found to be critical for this agreement, as well as EM simulation of interconnects between stages. The saturated output powers of the 2-, 3-, and 4-stack PA were 15.9, 19.8, and 21.6 dBm. As highlighted in Fig. 3.23, the saturated output power of the amplifiers increases with each added transistor as theory predicts, with a 5-6 dB increase in output power from the 2-stack relative to the 4-stack PA. The theoretical prediction assumes a peak current of 1.05 mA per μm gate width, a knee voltage of 0.15 V, and a drain voltage swing of 2.45 V per device.

As mentioned in Section 3.2, the stacking concept enables these high output powers without requiring low load impedances. The 2-, 3-, and 4-stack PA case have loadline impedances of approximately 15, 18.5, and 21 Ω . This allows a rela-

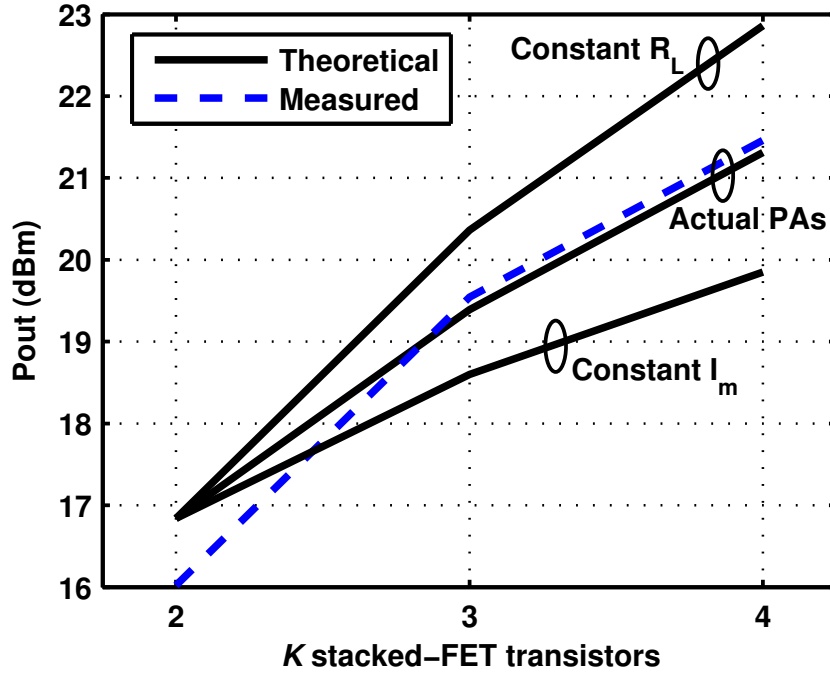


Figure 3.23: Pout versus number of stacked transistors.

tively low impedance transformation with quality factors of 1.1 to 1.5 and enables a wideband on-chip matching network. When increasing the number of stacked transistors, one can trade off the bandwidth and saturated power by changing the current i.e. changing the device sizes. The constant current approach, corresponding to no device size change, would provide the widest bandwidth, but low output power. The constant R_L case, corresponding to a linear increase device size with number of stacked transistors, would provide the highest output power, but a relatively small bandwidth. Furthermore, a small reduction in efficiency is expected due to the additional losses in the device parasitics. In this work, a moderate increase in device size was chosen for the 3- and 4-stack amplifiers, balancing the increase in saturated output power and bandwidth. Fig. 3.24 shows the PAE and saturated power of the three amplifiers as a function of frequency. While reduction of efficiency is expected, the 3- and 4-stack amplifiers nevertheless still achieve PAE around to 22%-25%. As can be seen in Fig. 3.24, the 3- and 4-stack ampli-

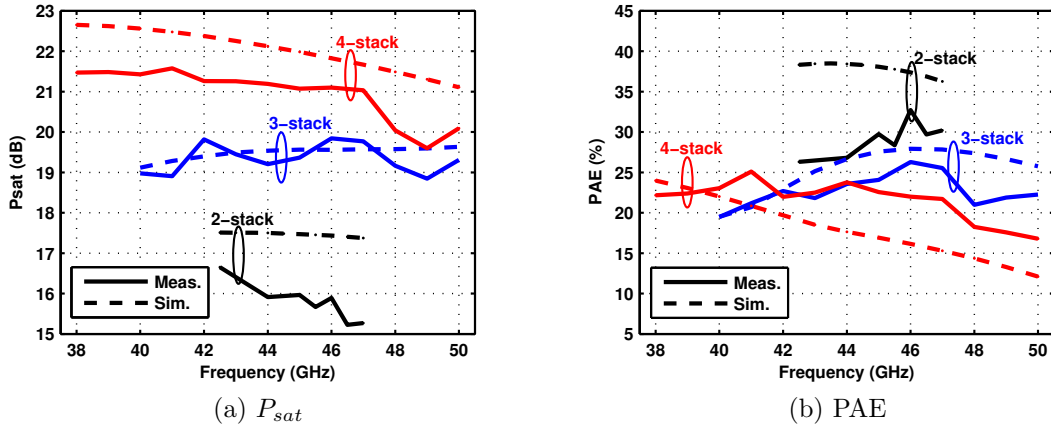


Figure 3.24: Measured peak PAE and P_{sat} versus frequency for 2-, 3-, 4-stack PA.

fiers maintain good performance over a wide frequency range from approximately 40 GHz to 48 GHz.

Table 3.3 summarizes the measured results and compares them with prior work. The 4-stack amplifier achieves the highest reported power in CMOS amplifiers, while maintaining good efficiency. The stacking technique is a promising alternative to passive power combining to achieve powers in the 100-200 mW range. For power levels closer to the watt level, stacking can be used in conjunction with on-chip and free space power combining. This may allow reducing the maximum power gap for CMOS relative to III-V amplifiers at mm-waves.

3.6 Conclusions

This chapter presents design guidelines for stacked-FET amplifiers at mm-wave frequencies. An updated theoretical discussion is presented including the gate-drain capacitance, which becomes significant in highly scaled CMOS processes. Furthermore, the importance of intermediate node matching is shown in theory and measurement, which result in significant efficiency improvements. The

Table 3.3: Comparison To Previously Reported Silicon mm-Wave PAs

Reference	Process	Architecture	Freq. (GHz)	Supply (V)	Psat (dBm)	PAE (%)
This Work	45-nm CMOS SOI	2-stack PA	46	2.5	15.9	32.7
This Work	45-nm CMOS SOI	3-stack PA	46	3.5	19.8	26.3
This Work	45-nm CMOS SOI	4-stack PA	41	5	21.6	25.1
GaAsIC 1999 [44]	GaAs pHEMT	4-way power combining	40	6	27.9	26.6
RFIC 2012 [45]	45-nm CMOS SOI	2-stack PA	42.5	2.7	18.6	34
T-MTT 2012 [46]	45-nm CMOS SOI	Push pull PA	45	2	15	27.5
CICC 2012 [47]	45-nm CMOS SOI	2-stack PA	47	2.4	17.6	34.6
CICC 2012 [47]	45-nm CMOS SOI	4-stack PA	47.5	5	20.3	19.4

measured efficiency increased from 26% to 32%, which is among the highest reported PAE values at mm-wave frequencies for silicon amplifiers. The saturated output power of 16 dBm of the 2-stack PA is comparable to previously reported results. However, higher output powers of approximately 19-20 dBm and 21-22 dBm were obtained by stacking three and four transistors respectively. This demonstrates the effectiveness of stacking FETs in an SOI process as an alternative to passive power combining.

3.7 Appendix 3.A: Optimal Drain Impedance

For optimal performance, the drain voltages of the transistors should be time aligned to each other and time aligned to the current from the transconductances. This condition can be expressed as:

$$\frac{V_{d,k+1}}{V_{d,k}} = \frac{k+1}{k}, \quad k = 1, 2, \dots, K-1 \quad (3.28)$$

$$V_{ds,k} = V_{d,1} = V_{opt}, \quad k = 1, 2, \dots, K \quad (3.29)$$

where V_{opt} is the optimal voltage waveform across each transistor. Using the small signal model shown in Fig. 3.8 one can derive the drain current of the $(k+1)$ th transistor using Kirchhoff's current law (KCL) as:

$$\begin{aligned} I_{d,k} &= I_{M,k} + I_{C_{ds,k}} + I_{C_{dsub,k}} - I_{C_{gd,k}}, \\ &= g_{m,k}V_{gs,k} + sC_{ds,k}V_{opt} + sC_{dsub,k}kV_{opt} - sC_{gd,k}V_{gd,k}, \\ k &= 1, 2, \dots, K. \end{aligned} \quad (3.30)$$

From (3.28), (3.29), and (3.30) one can solve for $Y_{opt,k} = 1/Z_{opt,k}$ as:

$$\begin{aligned} Y_{opt,k} &= \frac{g_{m,k}V_{gs,k}}{kV_{opt}} + \frac{sC_{ds,k}}{k} + sC_{dsub,k} - \frac{sC_{gd,k}V_{gd,k}}{kV_{opt}}, \\ k &= 1, 2, \dots, K. \end{aligned} \quad (3.31)$$

Using Kirchoff's voltage law (KVL) and KCL one can derive the gate-source voltages and gate-drain voltages:

$$V_{gs,k} = \frac{C_{gd,k} - (k-1)C_k}{C_{gs,k} + C_k + C_{gd,k}} V_{opt}, \quad (3.32)$$

$$V_{gd,k} = -\frac{kC_k + C_{gs,k}}{C_{gs,k} + C_k + C_{gd,k}} V_{opt}, \quad k = 2, 3, \dots, K. \quad (3.33)$$

If C_k is chosen according to (3.5) such that $Re\{Y_{opt,i}\} = 1/(k \cdot R_{opt})$, $V_{gs,k}$, and $V_{gd,k}$ can be expressed as:

$$V_{gs,k} = -\frac{V_{opt}}{g_{m,k}R_{opt}}, \quad (3.34)$$

$$V_{gd,k} = -\frac{1 + g_{m,k}R_{opt}}{g_{m,k}R_{opt}} V_{opt}, \quad k = 1, 2, \dots, K. \quad (3.35)$$

Using (3.31), (3.34), and (3.35) one can represent the desired admittance presented to the drain of the $(k+1)$ th transistor as

$$\begin{aligned} Y_{opt,k} &\approx \frac{1}{kR_{opt}} - \frac{s}{k} (C_{ds,k} + kC_{dsub,k}) - \frac{s}{k} \left(1 + \frac{1}{g_{m,k}R_{opt}} \right) C_{gd,k}, \\ &= \frac{1}{kR_{opt}} - \frac{s}{k} (C_{eqv,k}), \quad k = 1, 2, \dots, K. \end{aligned} \quad (3.36)$$

3.8 Appendix 3.B: Stacking Efficiency

Referring to Fig. 3.8 one can represent the load $Z_{M,k}$ of $(k + 1)$ th current generator $I_{M,k}$ as a shunt combination of an equivalent load capacitance $C_{load,k}$ and $k \cdot R_{opt}$. This can be expressed as

$$Z_{M,k} = kR_{opt} \cos(\Phi_k) e^{j\Phi_k}, \quad (3.37)$$

where $Z_{M,k}$ is the impedance presented to the transconductance and Φ_k is

$$\Phi_k = \arctan(-\omega C_{load,k} k R_{opt}). \quad (3.38)$$

For highest output power and efficiency, the drain voltages of the stacked transistors should be in phase with each other and in phase with the current i.e., the phases Φ_k should be 0° , $Z_{M,k} = k \cdot R_{opt}$. However, the power delivered from the $(k + 1)$ th current generator into $Z_{M,k}$ is

$$P_{out I_{M,k}} = \frac{1}{2} \operatorname{Re} \{Z_{M,k}\} I_{M,k}^2, \quad (3.39)$$

$$= \cos^2(\Phi_k) P_{out ideal I_{M,k}}. \quad (3.40)$$

This means that the output power and efficiency of each current generator in the stack is reduced by a factor of $\cos^2(\Phi_k)$. This is a worse case estimate and some of this reduction can be counteracted by changing bias voltages and input drive levels.

In the stacked-FET PA, a phase mistuning at a higher level will also mistune lower levels due to the feedback through the drain-source and gate-source capac-

itances. However, ignoring that effect, one can approximate the *power combining efficiency* of stacking and the output power of a stacked-FET PA as:

$$\eta_{stacking} \approx \left(\prod_{k=1}^K \cos(\Phi_k) \right)^2, \quad (3.41)$$

$$P_{out\ N-stack} \approx \eta_{stacking} P_{out\ ideal\ N-stack}, \quad (3.42)$$

where K is the number of stacked transistors and $P_{out\ ideal\ K-stack}$ is the output power of K -stacked PA if all currents and voltages are optimally aligned.

Acknowledgments

Chapter 3 is mostly a reprint of the material as it appears in “Analysis and Design of Stacked-FET Millimeter-Wave Power Amplifier”, *Transactions on Microwave Theory and Techniques*, Apr. 2013. This dissertation author was the primary author of this material.

Chapter 4

High Data Rate mm-Wave Wireless Transmission

In the quest for higher data rates for wireless handsets as well as satellite communication the mm-wave frequency bands are very compelling due to the availability of very wide channels. Consequently, research efforts have attempted to develop wireless transmitters and receivers in CMOS for mm-wave applications e.g. the authors of [48] achieved high data rates in a CMOS transmitter, by employing low complexity modulation schemes over very wide bandwidths. Low complexity modulation has several advantages such as the low PAPR of the signal and relaxed linearity requirements of the system, which implies that the PAs can be operated closer to compression. Unfortunately, lower complexity signals such as 16-QAM, QPSK, and BPSK also have lower spectral efficiency and inherently require the use of large bandwidths for high data rate communication. This would limit the number of users in a given band. A few groups focused on higher complexity modulation schemes such as 64-QAM, 256-QAM, or 1024-QAM in the mm-wave regime to achieve higher spectral efficiency [4, 49]. However, the achieved output

power levels are relatively low.

In this thesis, the viability of high complexity modulation of CMOS PAs in the mm-wave region is demonstrated. By a combination of various techniques such as spatial power combining, on-chip power combining, and stacking of transistors radiated output powers of 600 mW have been reported by B. Hanafi [50]. However, the nonlinearity of the amplifiers particularly when operated near compression distorts the modulated signal. This can be counteracted by employing digital predistortion (DPD). In order to predistort an array of PAs after spatial power combining one needs to feedback the output signal of the PAs to a DPD system. In this work we show that after DPD, an EVM of 1.3% for a 1024-QAM signal is achieved for a 98-MS/s signal enabling transmission of approximately 1 Gb/s.

This chapter is divided into four sections. In the next section the DPD system and DPD algorithms are described. In Section 4.2 an array with four antennas, combining the output power of eight stacked-FET PAs, is described. In Section 4.3 the radiate output signal quality of the array is evaluated before and after DPD. Section 4.4 summarizes the results.

4.1 Mark E Predistortion System

One of the advantages of operating transmitters at mm-waves is that PAs have much wider bandwidth compared to their RF counterparts. In order to leverage the wider channels at mm-waves, a wideband DPD system has been assembled as part of this dissertation.

Fig. 4.1 shows a block diagram of the DPD system (referred to as “Mark E”). The signal processing is performed on a PC running Matlab. The transmit data is uploaded to the block RAM of a Xilinx Virtex 6 FPGA via Ethernet, which

feeds the data to an I-Q DAC pair from Analog Devices (ADI), the AD9122. The signal is sampled at 368.64 MS/s. The DAC output signals are upconverted to 2.29 GHz with a quadrature modulator (QMOD), the ADL5735 from ADI and amplified. The 2 GHz signal is upconverted by a Qinstar mixer to 44.74 GHz, where it is amplified by a Centellax TA2U50HA driver and feed to the DUT. The DUT output is downconverted to 2.14 GHz with a Qinstar mixer, amplified, and downconverted once more with the ADL5365 to an IF of 368.64 MHz. There, it is filtered and sampled by a 12-bit, 491.52-MS/s ADC from ADI (AD9343). All the signal generators and sampling clocks are frequency locked from a common 10-MHz reference. The captured ADC data is stored on the DDR of the FPGA board and passed to a PC via Ethernet for post-processing. The Mark E system has a DPD observation bandwidth of approximately 250 MHz, which can be used to predistort desired signals with modulations in the order of 50 to 100 MHz, depending on the ACP requirements and the availability of filters at the PA output.

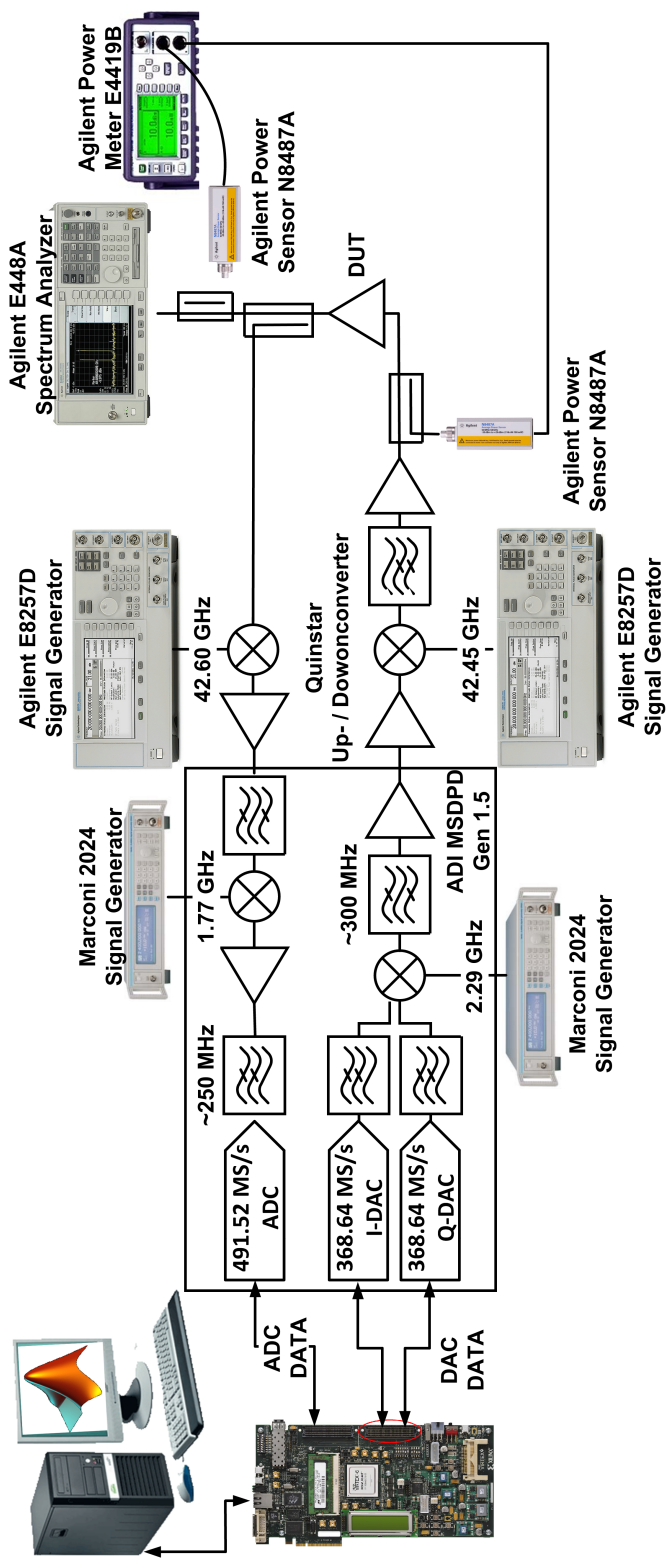


Figure 4.1: Simplified block diagram of mm-wave predistortion system

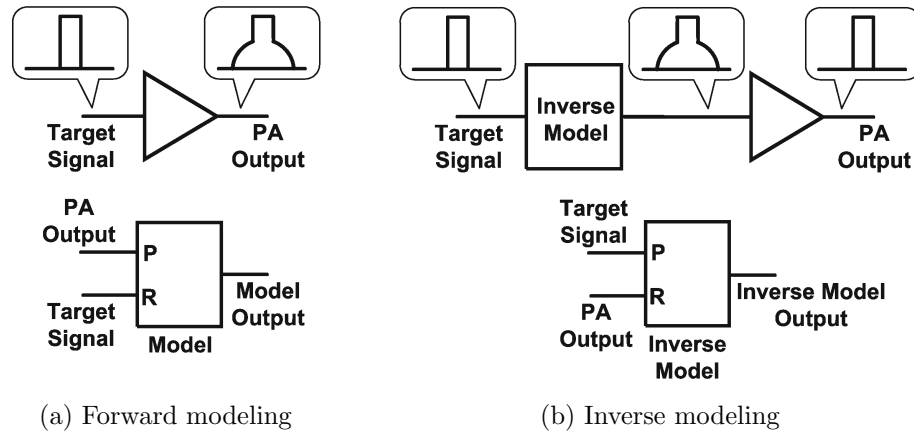


Figure 4.2: Modeling and inverse modeling of PA for DPD

4.1.1 DPD Algorithms

In Matlab the captured output signal is time aligned to the desired signal. From those two signals an AM-AM/AM-PM model is generated. In addition various polynomial models can be used to correct the nonlinearity as well as the memory effects. The polynomial models can be used to “model” the PA, to predict the output of the PA to a given input, as shown in Fig. 4.2(a). Alternatively, they can be used to build an inverse model, which computes a predistorted input signal for the PA, as shown in Fig. 4.2(b). The former is referred to as “forward modeling” and the latter is referred to as “inverse modeling”. The underlying principle is similar to the noise/distortion cancellation scheme discussed in Chapter 2 and the model parameters are extracted in a similar way. The difference between the “forward modeling” and “inverse modeling” is the swap of the primary (P) and reference (R) input. However, the extraction of the “inverse modeling” parameters is less reliable and often multiple iterations are needed, where the forward modeling generally leads to good results in the first iteration [51].

Many PA models are based on a subset of the Volterra-series, two of those are the memory polynomial (MP) [52] and the generalized memory polynomial

(GMP) [51]. Equation (4.1) shows the MP and GMP model, where the first sum is shared in the MP and GMP model and second and third sum are “cross memory kernels” added in the GMP model and the fourth term is added to model a dc offset. The number of model coefficients varies widely on the signal bandwidth, system nonlinearity and system memory. The coefficients are estimated similar to the least squares method described in Chapter 2 and [51].

$$\begin{aligned}
y_{GMP}(n) = & \sum_{k=0}^{K_a-1} \sum_{l=0}^{L_a-1} a_{kl} x(n-l) |x(n-l)|^k \\
& + \sum_{k=1}^{K_b} \sum_{l=0}^{L_b-1} \sum_{m=1}^{M_b} b_{klm} x(n-l) |x(n-l-m)|^k \\
& + \sum_{k=1}^{K_c} \sum_{l=0}^{L_c-1} \sum_{m=1}^{M_c} c_{klm} x(n-l) |x(n-l+m)|^k \\
& + d
\end{aligned} \tag{4.1}$$

For the experiments described in this Chapter a modified version of the GMP model is used, shown in (4.2) and referred to as RGMP. In this version all the cross memory terms up to a maximum memory length L_k are included in the model and only up to that memory depth. For example if $K = 1$ and $L_k = 1$ the model shown in (4.2) would include the terms $x(n)|x(n-1)|$ and $x(n-1)|x(n)|$. However, to include terms such as $x(n)|x(n-1)|$ in the conventional GMP, one would also include additional terms like $x(n-1)|x(n-2)|$, which may or may not be advantageous depending on the PA.

$$\begin{aligned}
y_{RGMP}(n) &= \sum_{k=0}^{K_a-1} \sum_{l=0}^{L_a-1} a_{kl} x(n-l) |x(n-l)|^k \\
&+ \sum_{k=1}^{K_b} \sum_{l=0}^{L_{b,k}-1} \sum_{m=1}^{L_{b,k}-l} b_{klm} x(n-l) |x(n-l-m)|^k \\
&+ \sum_{k=1}^{K_c} \sum_{l=0}^{L_{c,k}-1} \sum_{m=1}^l c_{klm} x(n-l) |x(n-l+m)|^k \\
&+ d
\end{aligned} \tag{4.2}$$

In addition the MP and RGMP algorithms have been extended with the band-limited DPD technique presented in [19]. This improves the DPD performance in cases where the observation bandwidth is less than three times the modulation bandwidth.

Furthermore, iterative approaches such as the memory mitigation algorithm (MM) [53] can be used to determine the best achievable performance. However, MM is targeted for the lab environment, since it relies on repeating the identical target signal for each iteration and is therefore not suitable for real-time implementations.

For the experiments described in this chapter MM DPD is used to establish a bound on the achievable performance and the MP or RGMP algorithms are used to show the achievable performance using real-time implementable DPD algorithms.

In addition, it is proposed in this dissertation to use the optimal PA input signal generated by MM DPD to calculate the inverse model. For this, the MM DPD signal is used as the primary input and the target signal as the reference signal as shown in Fig. 4.3. Mathematically this operation is similar to “forward modeling”, but the extracted parameters correspond to the inverse model. The

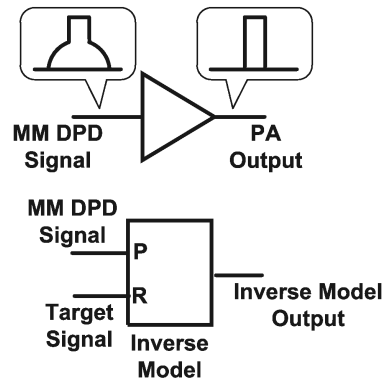


Figure 4.3: Inverse modeling of PA using the MM signal as primary input

closest approximation of the MM DPD signal will yield the best achievable performance using one of the models. Even though this approach is not suitable for a real-time implementation, it has the advantage that the accuracy of the different models can be compared offline once the MM DPD input signal is determined.

4.1.2 M-QAM Test Signals

A variety of modulated signals have been developed and are used in cellular or other communication systems. Generally, there is a tradeoff between spectral efficiency, peak to average ratio, occupied bandwidth, and robustness to system nonidealities. In this chapter, M-QAM modulation is used. M-QAM can have fairly high bandwidth efficiency, good energy efficiency, and the PAPR levels are acceptable for today's applications [54]. However, the achieved results are independent of the used modulation or coding scheme and similar results could be achieved with other signal types.

Without filtering, the bandwidths of the M-QAM signals are very wide. Therefore, they are commonly filtered using a root-raised-cosine filter (RRC) before transmission and filtered once more after reception. The actually occupied bandwidth after filtering slightly extends beyond the symbol rate and depends on

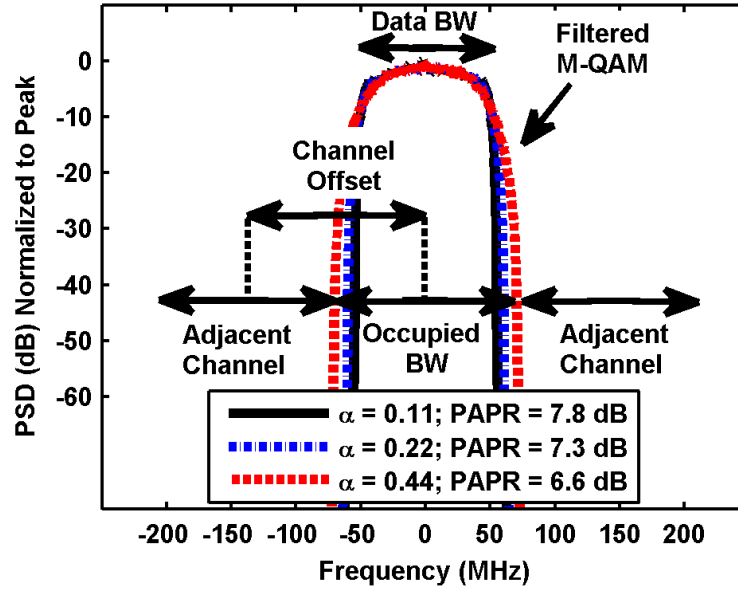


Figure 4.4: Spectrum of M-QAM signal after RRC filtering with different α

α according to (4.3) [54]. A smaller α leads to a smaller occupied bandwidth, but it increases the PAPR of the signal after filtering as shown in Fig. 4.4.

$$\text{Occupied bandwidth} = (1 + \alpha) \cdot \text{Symbol rate}. \quad (4.3)$$

Four different M-QAM signals were used to evaluate the system performance with and without predistortion. An α of 0.22 was used for the RRC filters, since it provides a reasonable tradeoff between occupied bandwidth and PAPR. Table 4.1 list the modulation order, symbol rate, the adjacent channel offset, peak to average power ratio (PAPR), and target EVM of the four signals. As discussed in Chapter 1, for a 1024-QAM signal one should achieve an EVM of approximately 1.2% for a BER of 10^{-6} and for 256 QAM an EVM of 2.3% is sufficient, assuming white noise is the dominant error source. However, depending on application higher BERs are acceptable e.g. the 3GPP standard allows BERs up to 10^{-3} [1, 55]. This significantly relaxes the EVM requirements as listed in Table 4.1. The ACPR

Table 4.1: Specifications of used modulated signals

Modulation	Symbol rate (MS/s)	Adjacent channel offset (MHz)	PAPR (dB)	Target EVM (%)
1024 QAM	49.15	56	7	1.2 - 2
	81.92	91	6.8	1.2 - 2
	98.3	110	7.1	1.2 - 2
256 QAM	98.3	110	7.1	2.3 - 4

requirements depend on the communication standard and the channel separation. For this series of experiments, the offsets equaled the occupied bandwidth of the signal.

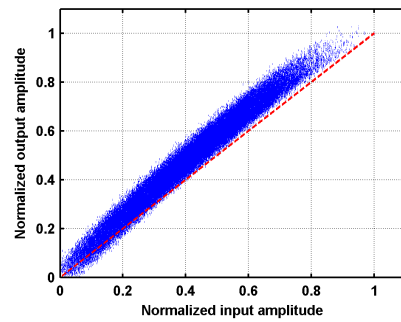
4.1.3 Predistortion of Mark E “Through” Test

To set a lower bound of the achievable performance, a “through” test at 44.74 GHz was conducted, which directly connects the transmitter to the coupler of the DPD receiver system. Even the most challenging 1024-QAM test signal, with a symbol rate of 98.3 MS/s, can be reliably demodulated after predistortion.

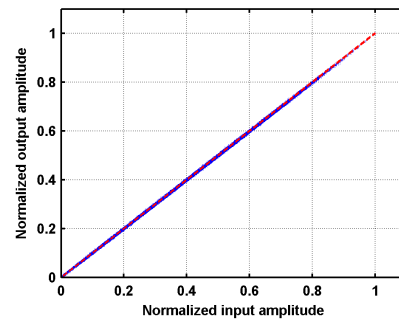
Fig. 4.5(a) and (b) show the AM-AM behavior of the Mark E system before and after predistortion using the memory mitigation (MM) algorithm. One can clearly see the memory of the system, indicated by the fuzz in the AM-AM curve, and the slight curvature of AM-AM curve indicates a slight nonlinearity of the system. However, the MM predistortion can correct for both these effects. The EVM before predistortion is 8.6% and after DPD it is 0.8%. The latter is sufficient to correctly demodulate a 1024-QAM signal as shown in Fig 4.5(d). Since, the MM algorithm is not field implementable the MP and RGMP models are used to model and predistort the system. Both models were trained on 10000 points from

the 65364 point pattern using the least squares method (pseudoinverse matrix) explained in Section 2.3.2. The MP model had 77 coefficients [$K_a = 7$ (even numbers only), $L_a = 20$] and the RGMP model used 178 coefficients [$K_a = 5$ (even numbers only), $L_a = 20$, $L_{b,1} = L_{c,1} = 10$, $L_{b,2} = L_{c,2} = 6$]. Fig. 4.5 (e) and (f) show the demodulated constellation with MP and RGMP predistortion. Note the RGMP DPD only performs slightly better besides the significantly higher number of coefficients. In this particular case the additional degrees of freedom the RGMP model provides are not required.

In the previous “through” test, one cannot distinguish between the nonlinearity of the transmitter and DPD receiver. However, ensuring the linearity of the system, in particular the DPD receiver, is critical. The DPD algorithm executed on the PC will correct for the cascaded nonlinearity of the transmitter, the DUT, and the DPD receiver. If the DPD receiver itself is nonlinear, it is not ensured that the signal at the DUT output will be correctly linearized. Similar concerns are also relevant for the system memory. The linearity of the transmitter and Quinstar downconversion receiver are evaluated separately. Fig. 4.6(a) shows the output spectrum of the driver amplifier before the DUT at the highest power level used for the experiments with stacked-FET PAs described in Section 4.3. It shows a small amount of spectral re-growth, which is expected based on the AM-AM curve of the system. The ACPR is -39 dBc (slightly degraded by the spectrum analyzer noise floor) and -40 dBc after downconversion to 2.14 GHz as shown in Fig. 4.6(b). Since the ACPR was not degraded after downconversion one can conclude that the nonlinearity of the downconversion mixer is not significant at these power levels. Even when the power levels are increased by 4 dB (the peak values needed for the experiments with the stacked-FET PAs in Section 4.3), the ACPRs at the output of the mm-wave driver and after downconversion to 2.14 GHz were -36.6 dBc.



(a) AM-AM without DPD



(b) AM-AM with DPD

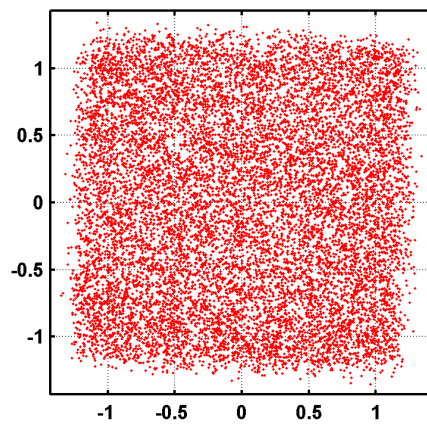
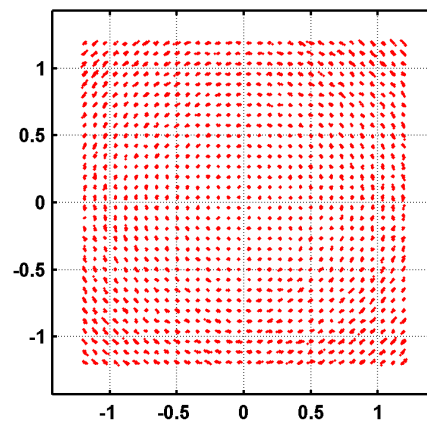
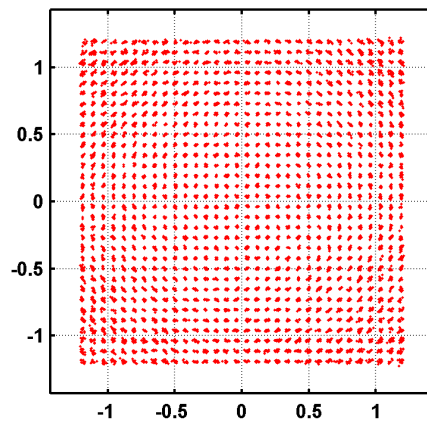
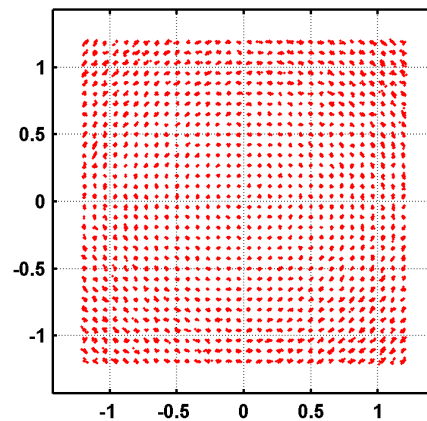
(c) Constellation without DPD
EVM=8.6%; $P_{out} \approx 3.8$ dBm(d) Constellation with MM DPD
EVM=0.8%; $P_{out} \approx 3.8$ dBm(e) Constellation with MP DPD
EVM=1%; $P_{out} \approx 3.8$ dBm(f) Constellation with RGMP DPD
EVM=0.9%; $P_{out} \approx 3.8$ dBm

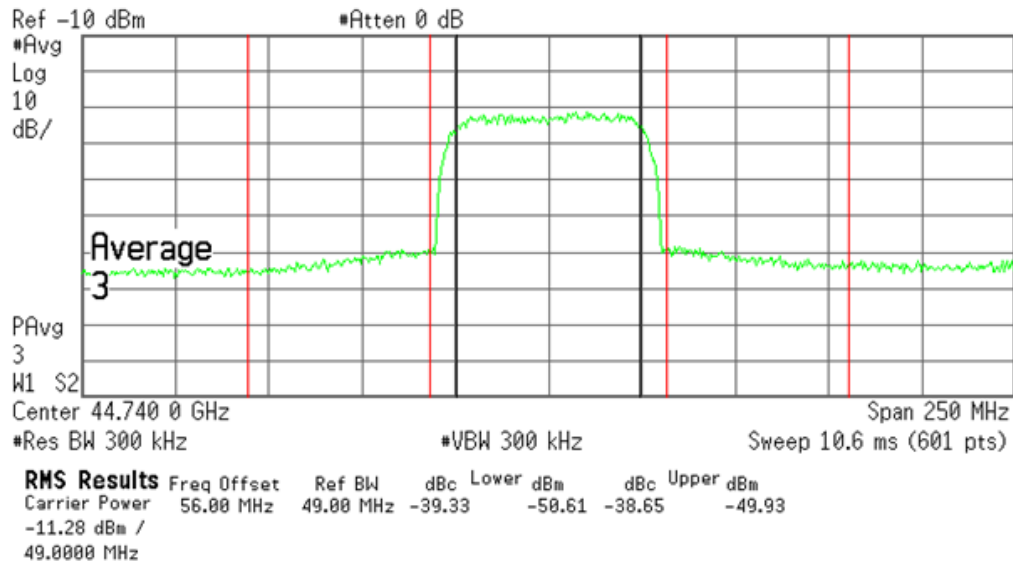
Figure 4.5: Evaluation of linearity and memory of the Mark E system in “through” test; P_{out} de-embedded to Quinstar downconverter RF input.

Since the linearity of the Quinstar downconversion mixer and the rest of the DPD receiver were sufficient, no separate model for the transmitter and DPD receiver were generate to “de-embed” the system nonlinearities. However, the DPD receiver had a significant transfer function across the band of interest due to various filters. By visual inspection it was determined that the majority of the transfer function is associated with the DPD receiver. An 32-tap FIR equalizer is generated and applied to the received ADC data to correct for the transfer function. A slightly more elaborate approach would separately equalize the transmitter and DPD receiver. However, it seemed not critical for these experiments due to the flat frequency response of the transmitter.

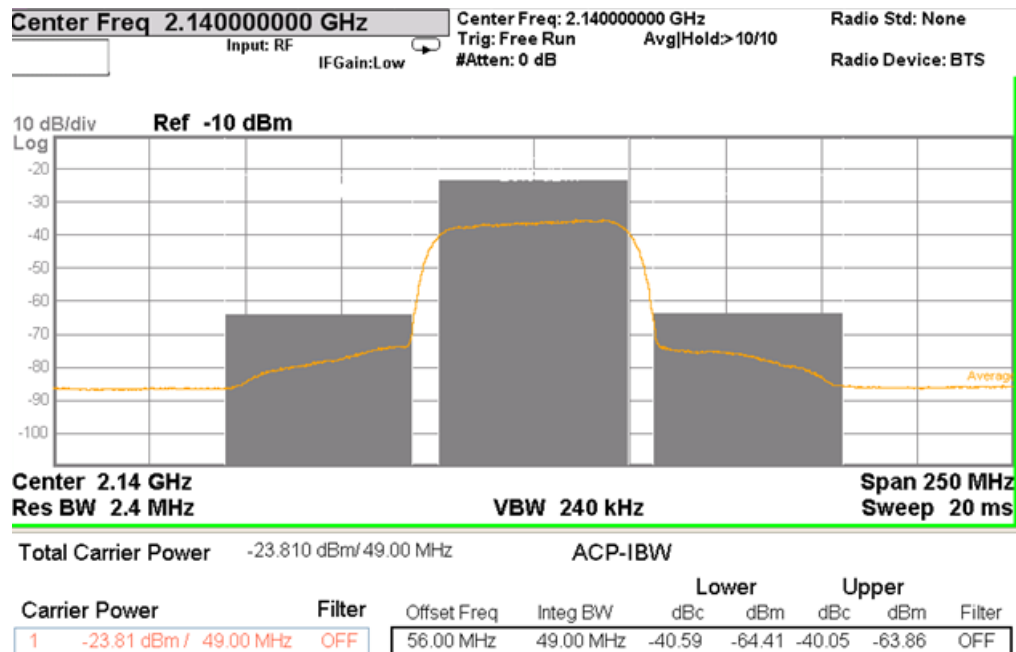
4.1.4 System Accuracy Limits

To successfully modulate and demodulate a high complexity signal a good system accuracy is required. This entails a good SNR of the transmitter and receiver as well as good repeatability of the experiments. The signal quality is influenced by the phase noise of the signal generators, the frequency locking errors of the signal generators, and thermal noise in the system. Conventionally the SNR is defined as the ratio of the signal power and the total noise power in the Nyquist band of the data converter. Note that the SNR is independent of the modulation bandwidth, assuming identical statistics of the signal.

A figure of merit for the accuracy of a signal is the normalized root means square error (NRMSE), as defined in (4.4). It is comparable to the EVM except that EVM is only evaluated at specific samples spaced by the symbol rate and the NRMSE is evaluated for every captured sample. However, the NRMSE has the advantage that it can be evaluated for any waveform including sine waves.



(a) mm-wave driver output spectrum for an output power of -3 dBm (highest used power in Section 4.3); ACPR \approx -39 dBc



(b) Output spectrum of the mm-wave downconverter for an input power of -3 dBm; output ACPR \approx -40 dBc

Figure 4.6: Spectral response of system “through” test after the mm-wave driver and after the mm-wave downconverter.

Table 4.2: Summary of NRMSE of CW signal after digital filtering with various filter corner frequencies

Filter bandwidth (MHz)	10	50	125	250
NRMSE (%)	0.95	1.05	1.1	1.25

$$NRMSE = \sqrt{\frac{\sum_{n=1}^{Record\ length} |Measured\ signal(n) - Target\ signal(n)|^2}{\sum_{n=1}^{Record\ length} |Measured\ signal(n)|^2}} \quad (4.4)$$

The NRMSE and EVM of a signals is degraded by the “in-band” as well as “out-of-band” noise. Therefore, the received signals are generally filtered reducing the degradation of the wideband white noise on the signal accuracy.

The system accuracy and sensitivity to the wideband noise floor of the system is evaluated by conducting a continuous wave (CW) test at 2.36 MHz offset from the carrier and applying a 5th order Butterworth filter with various corner frequencies in Matlab. Table 4.2 lists the achieved NRMSE results for different bandwidths of the digital receive filter. If the white noise would dominate the error, the NRMSE would grow proportional to the square root of the filter bandwidth. However, increasing the filter bandwidth and hence system noise has only a marginal effect on the NRMSE. This shows that the NRMSE is dominated by the phase noise and frequency locking of the signal generators.

To further support this observation, DPD experiments were conducted for a “through” test using 4.9-, 49- and 98-MS/s, 64-QAM signals with comparable PAPRs. Each of the signals is filtered before transmission and before demodulation with a RRC filter with an α of 0.22. Table 4.3 lists the achieved NRMSE and EVM with MM and RGMP DPD. It shows that the achievable EVM/SNR has only a

Table 4.3: Summary of NRMSE/EVM with MM and RGMP DPD for various symbol rates at a power of -3.5 dBm at the output of the mm-wave driver

Symbol rate (MS/s)	NRMSE (%)		EVM (%)	
	MM	RGMP	MM	RGMP
4.9	1.2	1.2	0.7	0.7
49	1.2	1.2	0.8	0.8
98	1.2	1.3	0.8	0.8

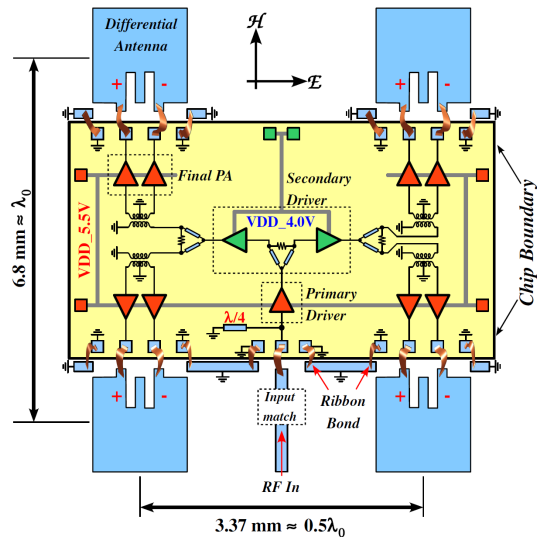
small dependency to symbol rate and the white noise of the system.

4.2 Spatially Power Combined stacked-FET PAs

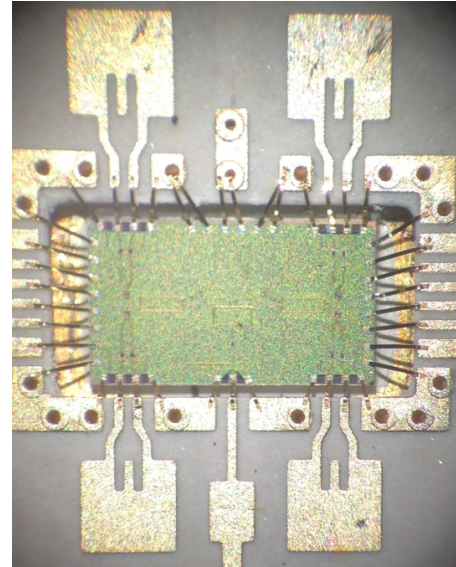
The stacked-FET PA architecture has frequently appeared in recent literature achieving very high output power at reasonable efficiencies for CMOS PAs [56,57]. In order to achieve higher radiated output powers a 2x2 power amplifier antenna array has been assembled by Bassel Hanafi. In [50] it is reported that eight 4-stack PAs achieved a radiated output power of 600 mW after spatial power combining using four differential patch antennas. The 4-stack PAs are similar to the ones presented in Chapter 3. The PA IC is mounted on a PCB and the PA output is bonded to patch antennas on the PCB.

Fig. 4.7(a) illustrates the printed circuit board (PCB) including the PA IC and antennas on the PCB. Fig. 4.7(b) shows a picture of the mounted and bonded IC on the PCB and connected to the antennas. The 2x2 array had a very broad output beam as shown in Fig. 4.7(c).

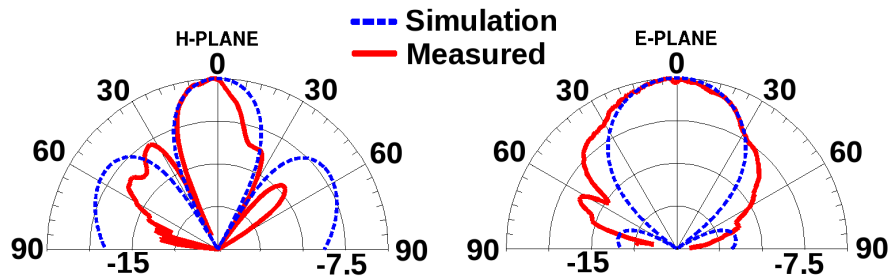
In conventional predistortion systems part of the PA output is feedback to an auxiliary receiver using a coupler. However, this was not an option in this system.



(a) Diagram of PCB for spatial power combining



(b) Picture of mounted IC connected to PCB patch antennas



(c) Measured radiated pattern of 2x2 array

Figure 4.7: Diagram of stacked-FET PA array with differential patch antennas [50]

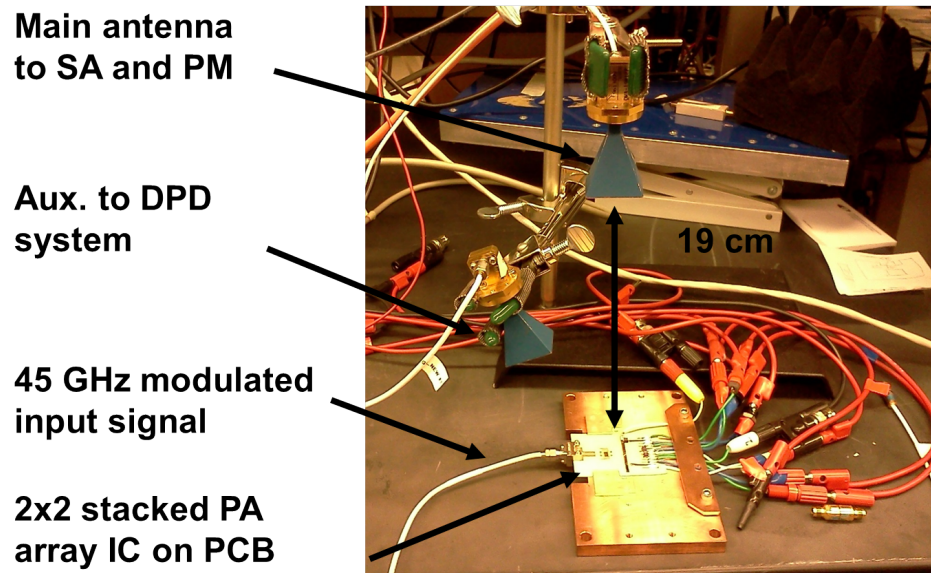


Figure 4.8: Picture of antenna assembly around the PCB with 2x2 antenna array.

Since the radiation pattern was very broad, an auxiliary antenna was placed aside from the main antenna to capture a signal for the DPD system. Fig. 4.8 shows the placements of the antennas around the PCB. Centered on top of the chip is the main antenna, which mimics the targeted receiver. The received power and output spectrum are monitored on lab instruments from the main antenna. The signal from the auxiliary antenna is fed to the DPD system. As described before, it is downconverted, filtered, and sampled by a 12-bit, 491.52-MS/s ADC.

The amplifiers are biased in class AB mode for higher efficiency. Therefore, the dc current changes with output power. Fig. 4.9 shows the measured equivalent isotropically radiated power (EIRP) received by the main antenna at 19 cm distance vs. DC current drawn by the eight 4-stack PAs for CW excitation. The coupler and cable loss are de-embedded. The peak EIRP is approximately 36.5 dBm with a ± 1.3 dB measurement uncertainty due to placement of the absorbers and the auxiliary antenna. The actual output power of the PA array is extrapolated from (4.5). The gain of the 2x2 antenna array (G_{TX}) is simulated to

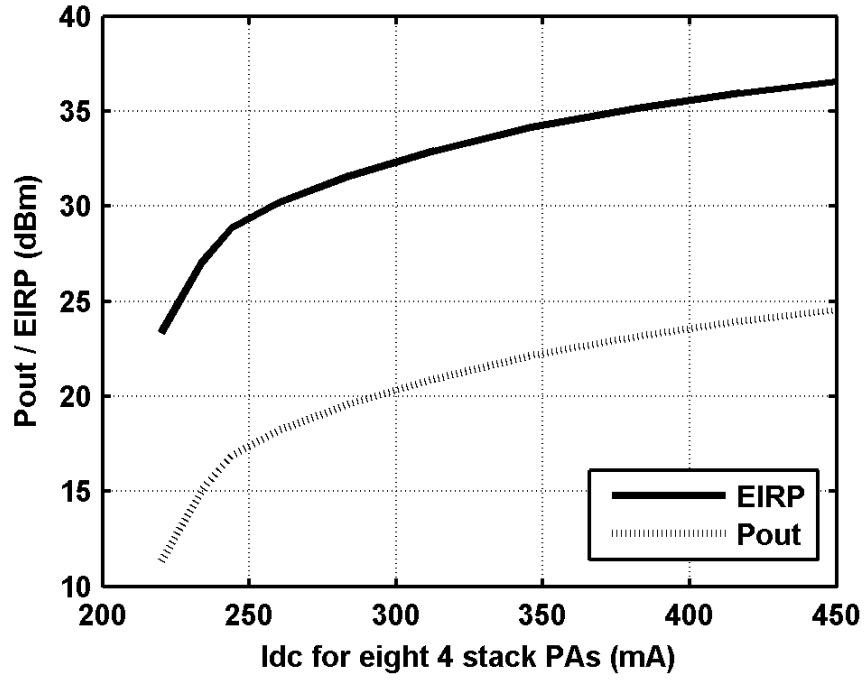


Figure 4.9: Measured EIRP at main antenna and estimated Pout vs. Idc of eight 4-stack PAs for CW excitation

be 12 dB.

$$EIRP = P_{out} \cdot G_{TX} \quad (4.5)$$

The system accuracy including the PA array is evaluated using the same CW test as in Section 4.1.4 without digital filtering. Table 4.4 summarizes the achieved NRMSE results for various output powers of the PA array. Comparing the NRMSE results with the PA array and without (as listed in Table 4.2) one can observe a small degradation of the NRMSE, due to the added noise by the PA and path loss by radiating the signal. The degrading effect on the NRMSE of the added thermal noise slightly decreases for higher output powers of the PA.

Table 4.4: Summary of NRMSE of CW signal without digital filtering for various output powers of the PA array

EIRP at receive antenna	28.02	29.4	30.7	31.2
NRMSE (%)	1.36	1.38	1.29	1.27

4.3 DPD Results of Spatially Power Combined stacked-FET PAs

The first DPD experiment on the spatially combined PA array was conducted at an EIRP power of 28.7 dBm at the “main antenna” (MA), using the 1024-QAM signal with a symbol rate of 49 MS/s. Considering a PAPR of 7 dB the peak power at the MA was approximately 35.7 dBm, which is only 0.8-1.8 dB less than the peak power achieved under CW excitation.

Following the approach explained in Section 4.1.1 the MM algorithm was used to iteratively compute the optimal input to the PA. The nonlinearity order and memory length of the MP and RGMP model was varied to find the model which best approximates this signal. The MP and RGMP models, which approximate the optimal predistorted signal most closely have 103 and 123 coefficients. The MP used the even orders up to 11 had 17 memory taps for each nonlinear kernel ($K_a = 11$, even only; $L_a = 18$) and one dc kernel. The RGMP model in addition uses “cross-memory terms” up to four past values i.e. $L_{b,1} = L_{c,1} = 5$.

Fig. 4.10 shows the frequency response of the input signal to the PA after the experimentally optimized predistortion using the MM algorithm along with the frequency response of the modeling error using the MP model and RGMP model to approximate it. Note that in this case the addition of 20 “cross memory terms” in the RGMP model significantly improved the modeling accuracy by 7-10 dB compared to the MP model. After predistortion, the MM, MP, and RGMP DPD

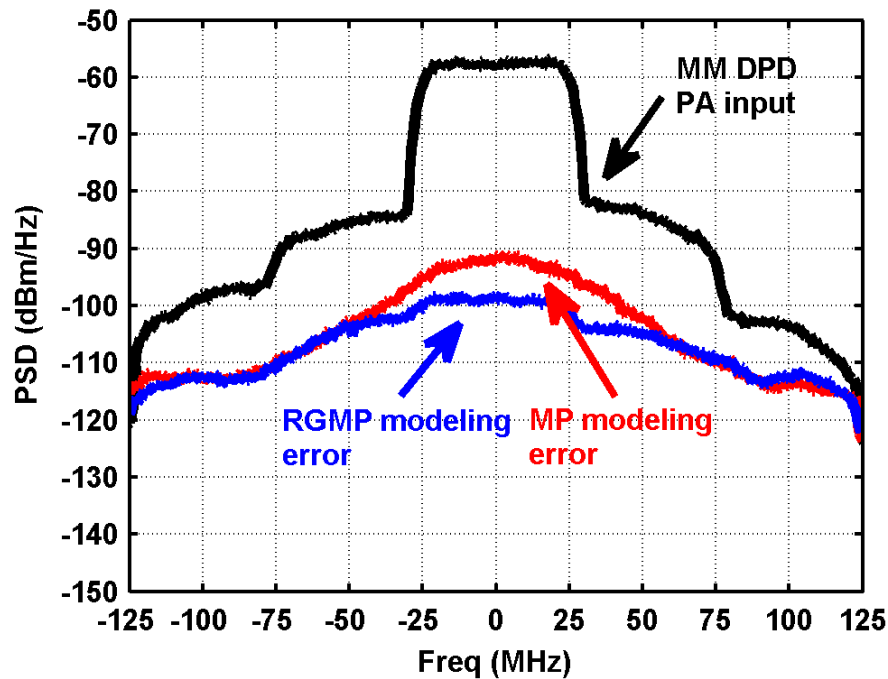


Figure 4.10: Comparison of MP and RGMP model to match MM DPD PA input signal

algorithms respectively achieved an EVM of 1%, 2%, and 1.3%. Similar results were obtained by two or three iterations of conventional “inverse modeling” using the MP and RGMP algorithms.

For the following experiments only MM and RGMP DPD is used, due to the inferior modeling accuracy achieved by the MP model. Table 4.5 summarizes EVM and ACPR without DPD, with memory mitigation (MM) and with RGMP DPD for the four signals. The EVM after MM DPD with the DUT in place is slightly worse than the system without DUT, which was expected given the difference in NRMSE results for the CW test with and without DUT. The MM algorithm achieves the same EVM regardless of output power of the PA (as long as no symbols are clipped). Even for the widest band signal, the RGMP model achieved comparable performance as the MM algorithm at low output powers.

Table 4.5: Summary of ACPR and EVM with and without DPD

Modulation	Symbol rate (MS/s)	EIRP (dBm)	ACPR (dBc)		
			No DPD	MM	RGMP
1024 QAM	49	29.8	-30.8	-37.6	-38
	82	28	-28.3	-37.5	-36
	98	26.2	-29.5	-32.9	-32.3
	98	27.7	-27	-30.6	-30.6
256 QAM	98	26.2	-29.5	-32.9	-32.3
			EVM (%)		
			No DPD	MM	RGMP
1024 QAM	49	29.8	6.2	1	1.24
	82	28	8	1.3	1.3
	98	26.2	5	1.2	1.3
	98	27.7	6.3	1.26	1.6
256 QAM	98	26.2	5	1.2	1.3

However, for the more challenging cases of higher power levels the EVM after RGMP predistortion was only 1.6%, where MM achieved an EVM of 1.26%. This is probably due to the higher order of nonlinearity, the resulting need for a higher order model, which leads to a higher numerical estimation inaccuracy of the model parameters.

Due to the limited DAC and ADC sampling rate, not all of the spectral regrowth can be corrected when wideband signals are used. This limits the achievable ACPR results in those cases as shown in Fig. 4.13-4.14(e).

Fig. 4.11-4.14(a) and (b) show the AM-AM curves of with and without DPD for the four modulation signals. Note that the input power is sufficiently backed-off to ensure that amplifier is not clipping any of the symbols. However, without DPD

one can clearly see the nonlinear behavior of the PAs in the curvature of the AM-AM curves. This is corrected with the DPD operation. As expected the amount of memory (dispersion) in the AM-AM behavior before DPD increases with the bandwidth of the modulated signal. To compensate this, the RGMP model requires an increasing number of coefficients for good modeling. Even though the RGMP model significantly reduces the amount of “memory” contained in the signals, as the bandwidth increases the dispersion in the AM-AM curves also increases after DPD for two reasons. First, the “in-band” SNR decreases with increasing modulation bandwidth, since the total power remains unchanged and spreads across a wider band of system noise. Second, the extraction of the model parameters becomes less accurate due to the limited observation bandwidth. Nonetheless, for all four signals the RGMP model achieved excellent results and signal quality of the PA array approaches the signal quality of the system.

Fig. 4.11-4.14(c) and (d) show the constellation with and without DPD. The constellations without DPD are unintelligible. However, once DPD is applied one can clearly recognize the constellation. Due to the marginal EVM even after DPD one can observe that for the 1024-QAM signals the symbols on the edges run the risk to be incorrectly demodulated. Fig. 4.14(d) shows that for 256 QAM the constellations are correctly demodulated for most symbols when DPD is applied, since the EVM is sufficient for that modulation order.

Fig. 4.11-4.14(e) shows the spectral response received by the main antenna for the four signals. Without DPD one can clearly see the spectral regrowth due to the transmitter nonlinearity. This can be significantly mitigated for the 49-MS/s and 82-MS/s signals. In the case of the 98 MS/s signals, DPD cannot completely remove the adjacent channel leakage due to the limited sampling rate of the DACs and ADC. DPD still significantly reduces the spectral regrowth close to the desired

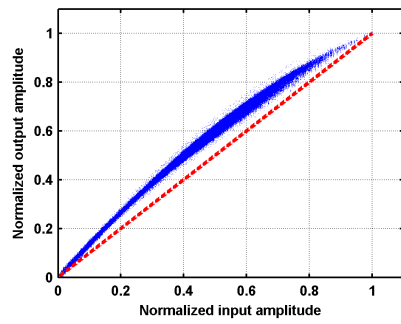
signal, which simplifies filtering specifications [19].

When reporting the linearity performance of an amplifiers e.g. by quoting their EVM or ACPR, it is relevant to note the average power and the PAPR at the PA output in comparison to the saturated output power of the amplifier. The peak EIRP under CW excitation is 36.6-37.6 dBm.

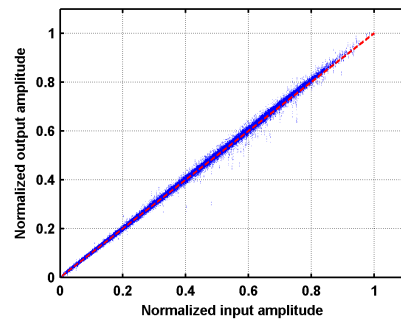
Fig. 4.15(a) shows the EVM before and after DPD versus output power. Given the saturated output power of 37.6 dBm and the PAPR of approximately 7 dB the highest power symbols will be clipped once the average power exceeds 31 dBm, which explains the sharp rise of EVM above that level.

Before predistortion the EVM is well above 2% even when the PA is backed-off very far. However, after predistortion for all four signals the EVM can be as low as 1.3%. In the case of the 49-MS/s signal this level can be maintained up to an average power of 29.8 dBm, which implies that the peak symbols have a power of 36.9 dBm, which in turn is very close to the saturated power level. However, as the bandwidth increases the inverse model extraction becomes more challenging due to the reduced signal to noise ratio. However, by reducing the average output power to approximately 26.5 dBm, the PA is slightly more linear. Therefore, some of the high power points are more easily predistorted, which improves the average EVM to 1.3% after DPD. In case when an EVM of 2.3% is sufficient the PAs can be operated up to 28 dBm for the 98-MS/s signals.

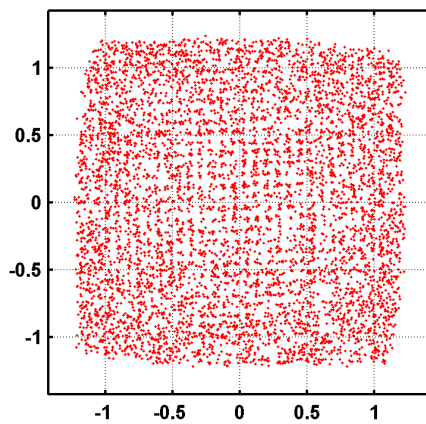
Fig. 4.15(b) shows the ACPR before and after DPD versus output power. As already shown in the captured spectra, the 49 MS/s signals will be predistorted in such a way that their adjacent channels are very clean. Therefore, their ACPR is very low after DPD. However, as the bandwidth of the modulation signal increases the adjacent channel exceeds the bandwidth of the DPD system. Therefore, the band edges cannot be corrected via DPD. However, the spectral leakage



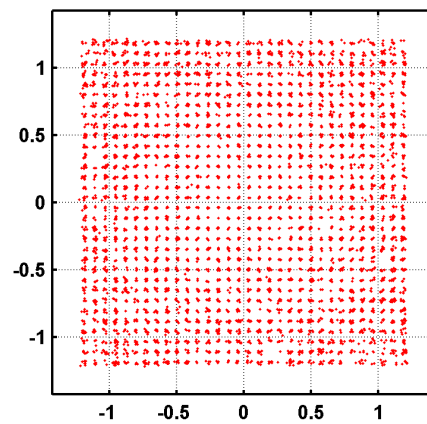
(a) AM-AM without DPD



(b) AM-AM with DPD; #coeff. = 91

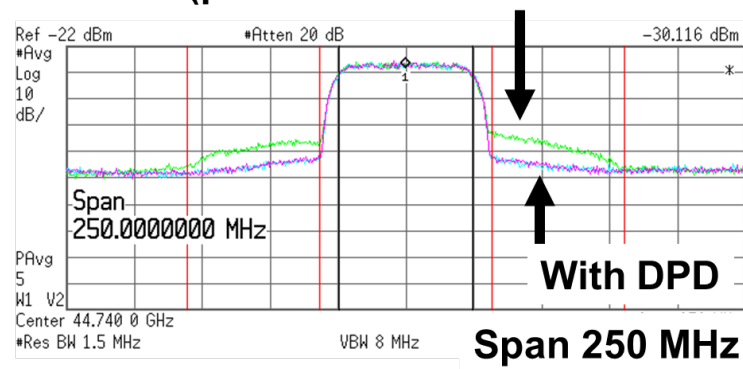


(c) Constellation without DPD
EVM = 5%; EIRP = 28.5 dBm



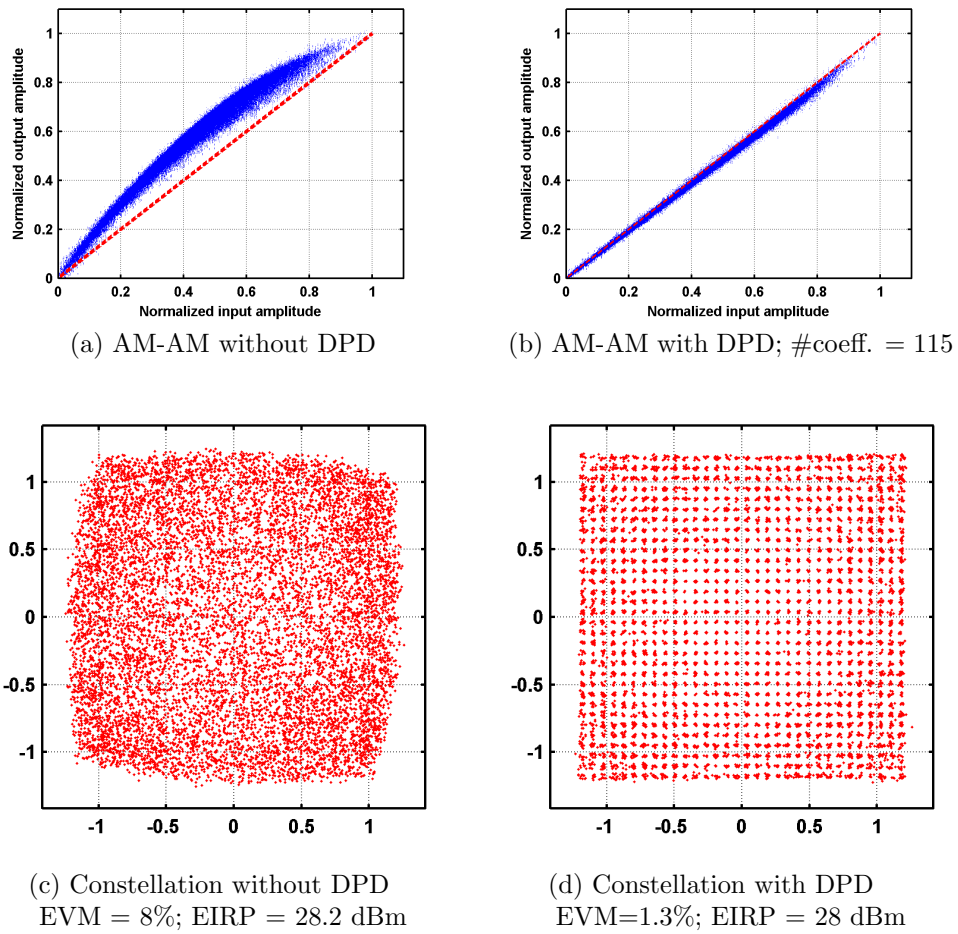
(d) Constellation with DPD
EVM = 1.24%; EIRP = 29.8 dBm

Without DPD (power matched to “with DPD” case)

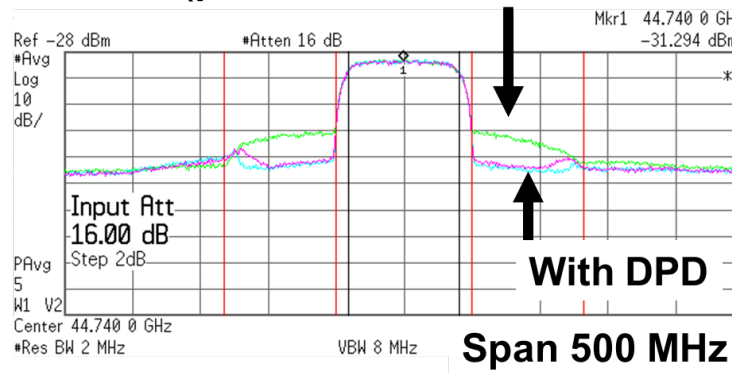


(e) Spectrum received by main antenna with and without DPD; EIRP = 29.8 dBm
Channel BW / Offset = 49 / 55 MHz; ACPR = -30.8 / -38 dBc without / with DPD;
 $K_a = 11$ (even numbers only), $L_a = 12$, $L_{b,1} = L_{c,1} = 4, L_{b,2} = L_{c,2} = 4$

Figure 4.11: PA array output before and after DPD for 49-MS/s, 1024-QAM signal

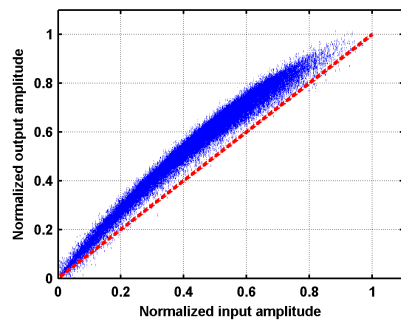


Without DPD (power matched to “with DPD” case)

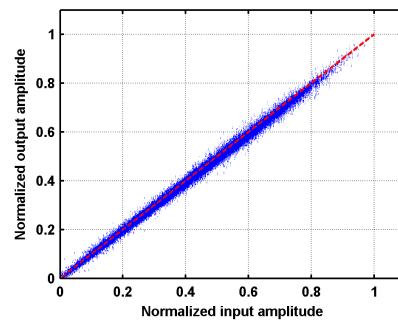


(e) Spectrum received by main antenna with and without DPD; EIRP = 28 dBm
 Channel BW / Offset = 82 / 91 MHz; ACPR = -28.3 / -36 dBc without / with DPD;
 $K_a = 11$ (even numbers only), $L_a = 16$, $L_{b,1} = L_{c,1} = 4, L_{b,2} = L_{c,2} = 4$

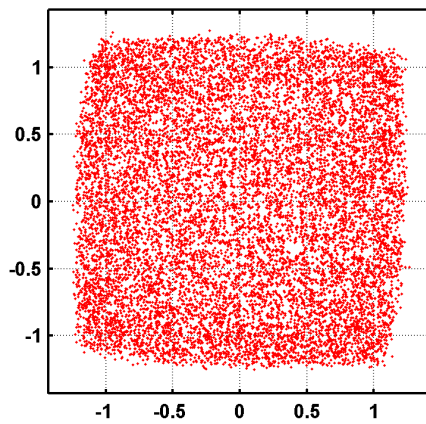
Figure 4.12: PA array output before and after DPD for 82-MS/s, 1024-QAM signal



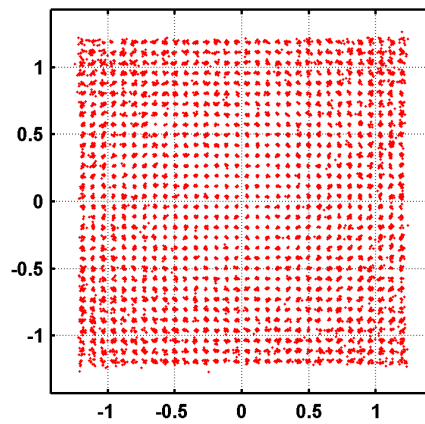
(a) AM-AM without DPD



(b) AM-AM with DPD; #coeff. = 181

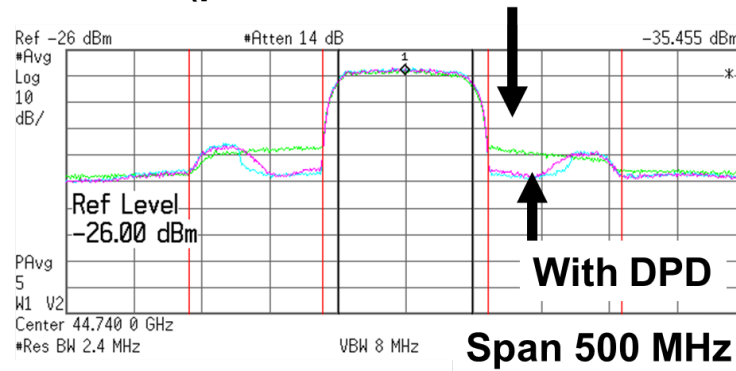


(c) Constellation without DPD
EVM = 6.4%; EIRP = 27.9 dBm



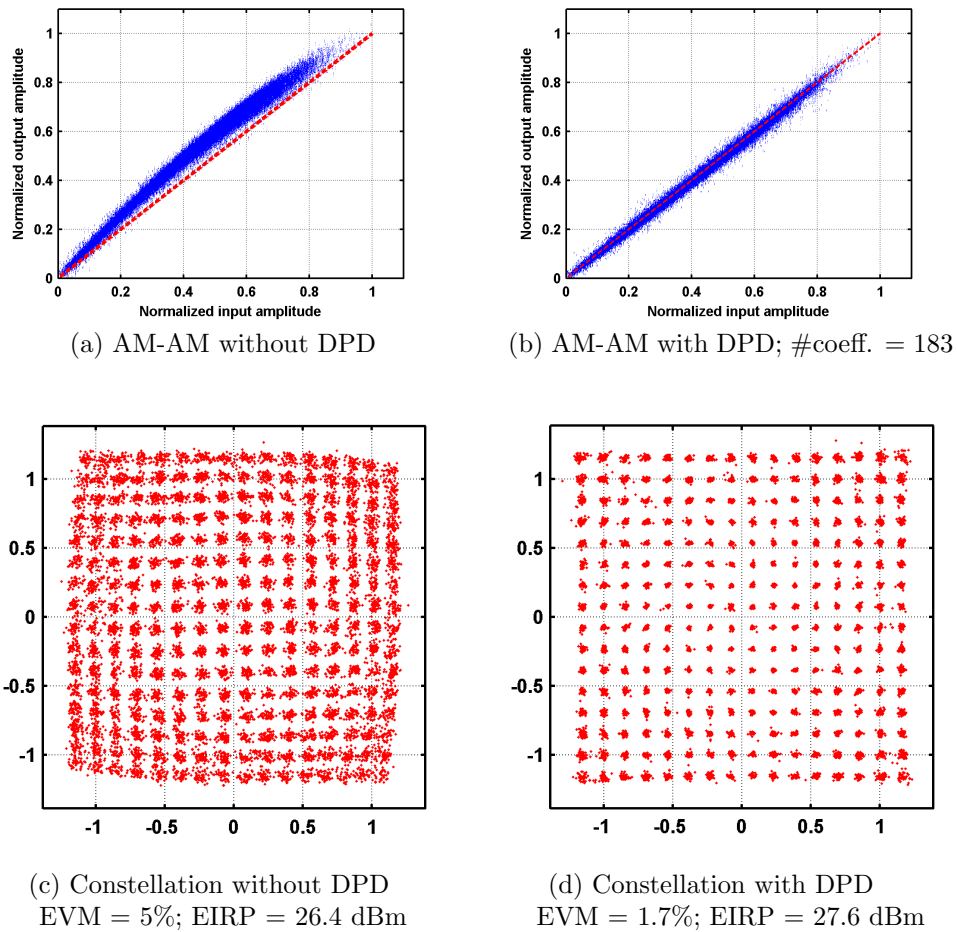
(d) Constellation With DPD
EVM = 1.3%; EIRP = 26.2 dBm

Without DPD (power matched to “with DPD” case)

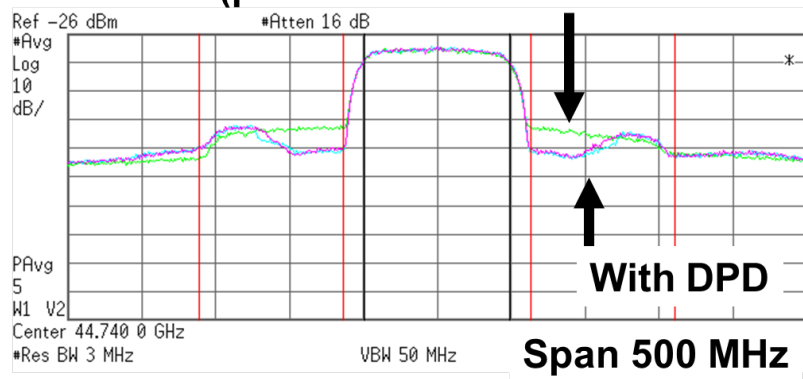


(e) Spectrum received by main antenna with and without DPD; EIRP = 26.2 dBm
Channel BW / Offset = 98 / 110 MHz; ACPR = -29.4 / -32.3 dBc without / with
DPD; $K_a = 11$ (even numbers only), $L_a = 15$, $L_{b,1} = L_{c,1} = 10$, $L_{b,2} = L_{c,2} = 6$

Figure 4.13: PA array output before and after DPD for 98-MS/s, 1024-QAM signal



Without DPD (power matched to “with DPD” case)



(e) Spectrum received by main antenna with and without DPD; EIRP = 26.2 dBm
Channel BW / Offset = 98 / 110 MHz; ACPR = -29.5 / -32.3 dBc without / with
DPD; $K_a = 7$ (even numbers only), $L_a = 18$, $L_{b,1} = L_{c,1} = 9$, $L_{b,2} = L_{c,2} = 7$

Figure 4.14: PA array output before and after DPD for 98-MS/s, 256-QAM signal

within the bandwidth of the DPD system is significantly suppressed to a level of approximately 35-36 dBc, as shown in Fig. 4.14(e).

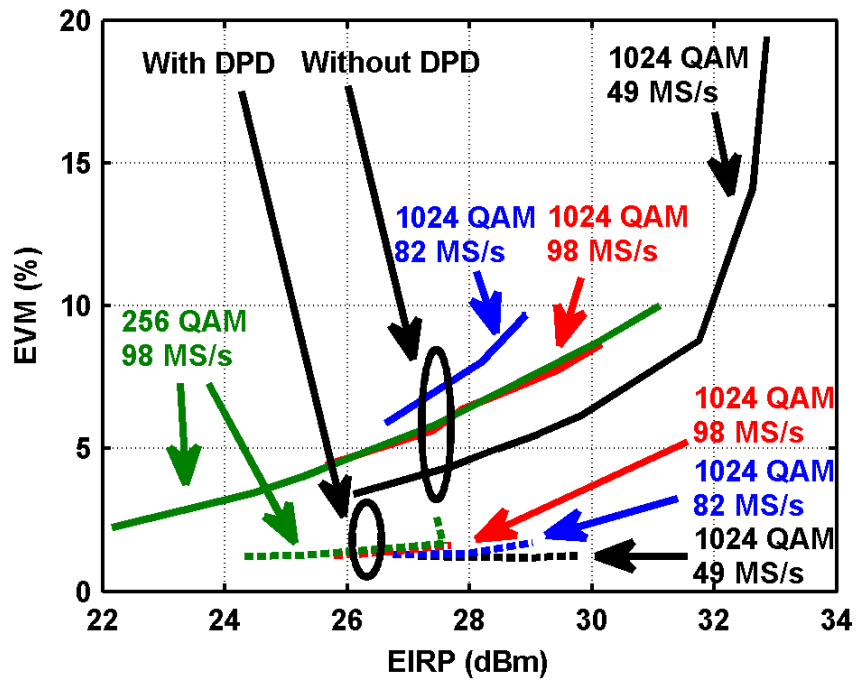
4.4 Conclusions

This chapter presents a digital predistortion system for mm-wave operation over relatively wide bandwidths. High data rates of approximately 1 Gb/s were achieved in a very spectrally efficient manner using 1024-QAM signals. The DPD system was used to predistort an array of stacked-FET PAs after spatial power combining. It was demonstrated that an auxiliary antenna can be used to capture power from a radiate “sidelobe” and use it as a the feedback signal for the DPD system. After predistorion using the RGMP model the signal achieved an average EVM of approximately 1.3% at an EIRP of 26.5 dBm. In cases where higher BER or higher power are critical the system can be used to transmit 256-QAM signals with an EVM of 1.6% at a power of 28 dBm.

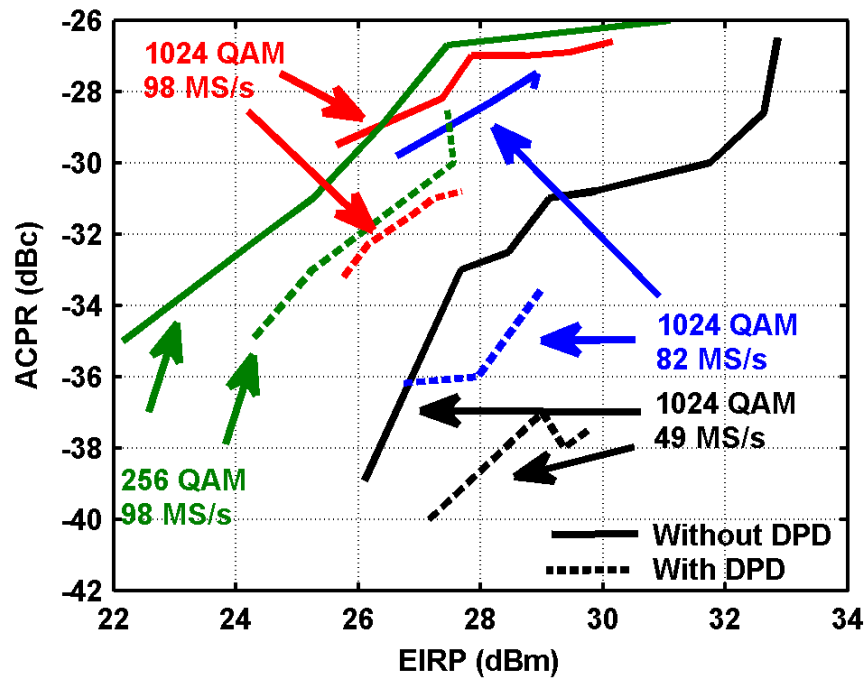
Acknowledgments

The material in Section 4.3 will in part be used for a publication in preparation with the working title “Digital Predistortion for 1024-QAM of Millimeter-wave, Free-space-combined Stacked-FET PAs”. The dissertation author was the primary author of this material.

Section 4.2 is in part a reprint of the material as it appears in “A CMOS 45 GHz Power Amplifier with Output Power >600 mW Using Spatial Power Combining”, accepted to *2014 IEEE MTT-S International Microwave Symposium (IMS)*. The dissertation author was a co-author of that work. The material is only included to explain the system and the relevance of the achieved results.



(a) EVM vs. EIRP for various signals



(b) ACPR vs. EIRP for various signals

Figure 4.15: EVM and ACPR vs. EIRP

Chapter 5

Conclusions and Future Work

5.1 Dissertation Summary

This dissertation focused on various challenges for high data rate wireless communication.

In the first part of the the dissertation, one of the issues for carrier aggregation arising when certain bands are paired is addressed. The thesis describes how this problem can be mitigated by a novel DSP based canceller.

The satellite and WiFi community partially addresses the need for higher data rates, by utilizing the wide channels available in mm-wave bands. The second part of the dissertation addresses some of the challenges for low cost mm-wave transmitters by the development of higher power and higher efficiency mm-wave CMOS PAs.

The third part of the dissertation describes the use of digital predistortion for mm-wave transmitters to enable the use of complex modulation schemes, which improves the spectral efficiency and allows high data rates in relatively narrow channels.

Chapter 2 describes the self-jamming problem, which arises in transceiver RF front-ends employing uplink carrier aggregation (UL CA). The nonlinearity of passive components such as switches and duplexers will create cross-modulation products. For certain band pairs, those products land in the receive band and significantly degrade the receiver sensitivity. A multiple input single output (MISO) adaptive distortion canceller is proposed and allows compensation of transfer functions for both transmitted signals and is robust against time alignment errors. The MISO algorithm has been successfully used in a realistic experiment setup to cancel the measured interference in the digital baseband without additional analog hardware by up to 20 dB. This allows the application of UL CA at full transmit power in problematic band pairs without the added expense of highly linear passive components (antenna switches and duplexers).

Even if all challenges related to carrier aggregation are overcome it allows at most the simultaneous transmission of five 20-MHz LTE signals. This is not enough to address the anticipated demand for bandwidth. A promising long term solution are the use of the mm-wave bands for extremely high data rates, due to the availability of continuous wideband channels of many hundreds of megahertz.

Chapter 3 describes some of the key challenges in designing low cost and efficient power amplifiers in CMOS for mm-wave operation. The low breakdown voltage issue in CMOS is overcome by adapting the stacked-FET technique to mm-waves. A theoretical framework is presented including the gate-drain capacitance, which becomes significant in highly scaled CMOS processes. Furthermore, the importance of intermediate node matching is shown in theory and measurement, which result in significant efficiency improvements. The measured efficiency increased from 26% to 32%, which is among the highest reported PAE values at mm-wave frequencies for silicon amplifiers. The saturated output power of 16 dBm

of the 2-stack PA is comparable to previously reported results. However, higher output powers of approximately 19-20 dBm and 21-22 dBm were obtained by stacking three and four transistors respectively. This demonstrates the effectiveness of stacking FETs in an SOI process as an alternative to passive power combining.

Implementation of PAs with acceptable efficiency addresses only part of the problem. Prior work achieved very high data rates by adapting low complexity modulations such as OOK, BPSK, and QPSK. These are inherently low in spectral efficiency.

In Chapter 4 a mm-wave digital predistortion (DPD) system is described, which linearizes a set of stacked-FET PAs. An auxiliary antenna was used to capture power from a radiate “sidelobe” and feed the signal to the DPD system. DPD improved the EVM of compressing PAs to 1.3%, which is sufficient to reliably demodulate 1024-QAM signals. This allowed data rates of approximately 500, 820, and 1000 Mb/s across 50, 82 and 100 MHz wide channels. This would allow simultaneous operation of up to 15-30 channels in a 3 GHz band and an aggregated data rate of 15 Gb/s.

5.2 Future Work

5.2.1 Two UL CA and Three DL CA

The uplink carrier aggregation scenario described in Chapter 2 focuses on the simultaneous transmission on two channels and the relating cross-modulation caused by the third order nonlinearity. However, carriers are also planning to deploy two uplink and three downlink CA e.g. pairing bands 8, 3, and 7. Fig. 5.1 shows a likely implementation of a transceiver for that scenario. One of the antennas is connected to a diplexer. The diplexer allows the simultaneous con-

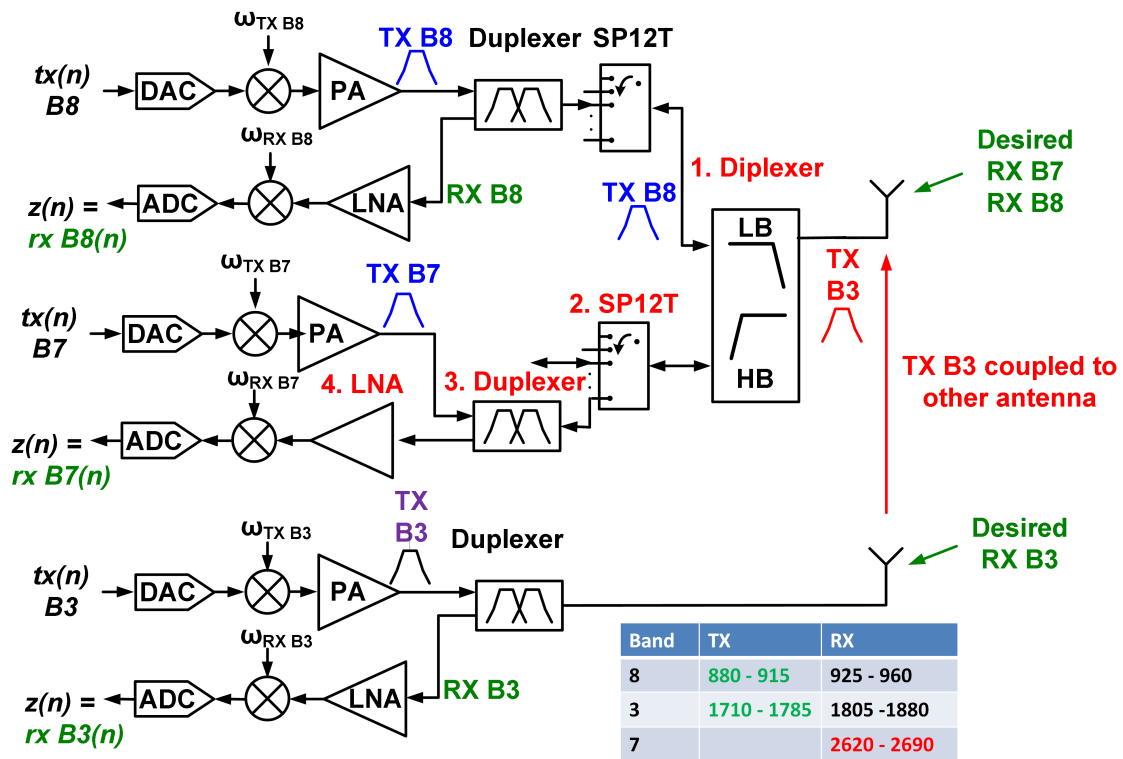


Figure 5.1: 2 UL and 3 DL CA with CM2 desensing on of the receivers

nection of two transceivers to the antenna, while providing approximately 15 dB isolation between the two transmitters. In this two UL and three DL CA scenario, signals are transmitted in band 8 (B8) and band 3 (B3). Each of the signals are radiated from a different antenna. However, due to the limited antenna isolation both the B3 and B8 transmit signals will appear at the diplexer ports. Attenuated copies of both transmit signals will also appear at the multi-throw switch, the B7 diplexer, and LNA. If all components were perfectly linear the B3 and B8 transmit signals would be sufficiently attenuate and would not desensitize the B7 receiver. However, in this particular combination of band pairs the second order cross modulation product CM2 will land on the B7 receive band.

The second order nonlinearity is specified by the second order intercept point (IIP2). Table 5.1 lists likely IIP2 values of high end components. Given

Table 5.1: B3 and B8 TX power at different components of B7 receiver and resulting CM2 power

	B7 LNA	B7 Duplex	Switch	Diplexer
P_{TXB3} (dBm)	-40	10	10	11
P_{TXB8} (dBm)	-44	6	6	21
IIP2 (dBm)	50	100	100	100
P_{CM2} (dBm)	-134	-84	-84	-68

these IIP2s and the power levels at the various front-end components the CM2 power is estimated. Since the targeted receiver sensitivity is in the order of -100 dBm, the corresponding receiver desense of the front-end passive components is severe. Future work should adapt the MISO distortion canceller described in this thesis to this scenario to mitigate the significant receiver desense for two UL and three DL CA scenario.

5.2.2 Silicon mm-Wave Transmitters

The transmitter used in the mm-wave DPD system used external components to achieve the high signal quality at these high data rates. Future work, should implement this in silicon chips. A. Gupta and A. Agah implemented mm-wave modulators in silicon which would be suitable for this purpose [58,59]. Therefore, an interesting follow up experiment uses the silicon transmitters to drive the stacked-FET PA array and evaluate the achievable EVM after DPD. This is currently under preparation by P. Wu.

Thinking even further ahead higher speed and low cost transceivers would even further benefit the mm-wave systems, given the availability of wide bandwidths at mm-waves and the fact that the bandwidth of the stacked-FETs spans

across many gigahertz.

Bibliography

- [1] G.-S. Choi, J.-H. Bae, H.-M. Park, and S.-W. Kim, "State-parallel MAP module design for turbo decoding of 3GPP," *Journal of the Korean Physical Society*, vol. 40, no. 4, pp. 677–685, April 2002.
- [2] J. Wannstrom, "Carrier aggregation explained," 3GPP, Tech. Rep. [Online]. Available: <http://www.3gpp.org/Carrier-Aggregation-explained>
- [3] M. Rumney, "3GPP lte standards update: Release 11, 12 and beyond," Agilent, Tech. Rep., 2012. [Online]. Available: http://www.home.agilent.com/upload/cmc_upload/All/25Oct12LTE.pdf?&cc=DE&lc=ger
- [4] W.-H. Lin, H.-Y. Yang, J.-H. Tsai, T.-W. Huang, and H. Wang, "1024-QAM high image rejection E-band sub-harmonic IQ modulator and transmitter in 65-nm CMOS process," *IEEE Trans. Microw. Theory Tech.*, vol. 61, no. 11, pp. 3974–3985, 2013.
- [5] *International Technology Roadmap for Semiconductors*, 2011.
- [6] S. C. Cripps, *RF Power Amplifiers for Wireless Communications, Second Edition (Artech House Microwave Library)*. Artech House, 2006.
- [7] *Sony Corporation, CXM3555ER Datasheet*. [Online]. Available: http://www.sony.net/Products/SC-HP/new_pro/july_2010/pdf/cxm3555er_e.pdf
- [8] *Epcos, Saw Duplexer B7654 Datasheet*, Dec 2011. [Online]. Available: <http://www.epcos.com/inf/40/ds/mc/B7654.pdf>
- [9] M. Kahrizi, J. Komaili, J. Vasa, and D. Agahi, "Adaptive filtering using LMS for digital TX IM2 cancellation in WCDMA receiver," in *IEEE Radio and Wireless Symp.*, 2008, pp. 519–522.
- [10] M. Omer, R. Rimini, P. Heidmann, and J. S. Kenney, "A compensation scheme to allow full duplex operation in the presence of highly nonlinear microwave components for 4G systems," in *IEEE MTT-S Int. Microw. Symp. Dig.*, 2011, pp. 1–4.

- [11] M. Omer, R. Rimini, P. Heidmann, and J. S. Kenney, "All digital compensation scheme for spur induced transmit self-jamming in multi-receiver RF front-ends," in *IEEE MTT-S Int. Microw. Symp. Dig.*, 2012, pp. 1–3.
- [12] B. Widrow, J. R. Glover, Jr., J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, and R. C. Goodlin, "Adaptive noise cancelling: Principles and applications," *Proc. IEEE*, vol. 63, no. 12, pp. 1692–1716, 1975.
- [13] S. O. Haykin, *Adaptive Filter Theory (4th Edition)*. Prentice Hall, 2001.
- [14] C. Lederer and M. Huemer, "LMS based digital cancellation of second-order TX intermodulation products in homodyne receivers," in *IEEE Radio and Wireless Symp. (RWS)*, 2011, pp. 207–210.
- [15] C. D. Presti, D. F. Kimball, and P. M. Asbeck, "Closed-loop digital predistortion system with fast real-time adaptation applied to a handset WCDMA pa module," *IEEE Trans. Microw. Theory Tech.*, vol. 60, no. 3, pp. 604–618, 2012.
- [16] H. Dabag, H. Gheidi, P. Gudem, and P. M. Asbeck, "All-digital cancellation technique to mitigate self-jamming in uplink carrier aggregation in cellular handsets," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2013, pp. 1–3.
- [17] J. C. Gomez and E. Baeyens, "Identification of multivariable hammerstein systems using rational orthonormal bases," in *Proc. of the 39th IEEE Conf. on Decision and Control*, vol. 3, 2000, pp. 2849–2854.
- [18] M. Omer, R. Rimini, P. Heidmann, and J. S. Kenney, "A PA-noise cancellation technique for next generation highly integrated RF front-ends," in *IEEE Radio Freq. Integ. Circuits Symp.*, 2012, pp. 471–474.
- [19] C. Yu, L. Guan, E. Zhu, and A. Zhu, "Band-limited Volterra series-based digital predistortion for wideband RF power amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 60, no. 12, pp. 4198–4208, 2012.
- [20] V. Aparin and L. E. Larson, "Modified derivative superposition method for linearizing FET low-noise amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 2, pp. 571–581, 2005.
- [21] N. Kim, V. Aparin, K. Barnett, and C. Persico, "A cellular-band CDMA 0.25 μm CMOS LNA linearized using active post-distortion," *IEEE J. Solid-State Circuits*, vol. 41, no. 7, pp. 1530–1534, 2006.
- [22] R. M. Johnstone, C. R. Johnson, R. R. Bitmead, and B. D. O. Anderson, "Exponential convergence of recursive least squares with exponential forgetting factor," in *21st IEEE Conf. on Decision and Control*, vol. 21, 1982, pp. 994–997.

- [23] C. Paleologu, J. Benesty, and S. Ciochina, "A robust variable forgetting factor recursive least-squares algorithm for system identification," *IEEE Signal Process. Lett.*, vol. 15, pp. 597–600, 2008.
- [24] A. H. Sayed, *Adaptive Filters*. Wiley-IEEE Press, 2008.
- [25] M. Shifrin, Y. Ayasli, and P. Katzin, "A new power amplifier topology with series biasing and power combining of transistors," in *Proc. IEEE Microw. and Millim.-Wave Monolithic Circuits Symp.*, 1992, pp. 39–41.
- [26] T. Sowlati and D. M. W. Leenaerts, "A 2.4-GHz 0.18- μ m CMOS self-biased cascode power amplifier," *IEEE J. Solid-State Circuits*, vol. 38, no. 8, pp. 1318–1324, Aug. 2003.
- [27] A. K. Ezzeddine and H. C. Huang, "The high voltage/high power FET (HiVP)," in *Proc. IEEE Radio Freq. Integr. Circuits Symp.*, 2003, pp. 215–218.
- [28] S. Pornpromlikit, H.-T. Dabag, B. Hanafi, J. Kim, L. E. Larson, J. F. Buckwalter, and P. M. Asbeck, "A Q-band amplifier implemented with stacked 45-nm CMOS FETs," in *Proc. IEEE Compound Semicond. Integr. Circuit Symp.*, 2011, pp. 175–178.
- [29] D. Fritsche, R. Wolf, and F. Ellinger, "Analysis and design of a stacked power amplifier with very high bandwidth," *IEEE Trans. Microw. Theory Tech.*, vol. 60, no. 10, pp. 3223–3231, 2012.
- [30] J. G. McRory, G. G. Rabjohn, and R. H. Johnston, "Transformer coupled stacked FET power amplifiers," *IEEE J. Solid-State Circuits*, vol. 34, no. 2, pp. 157–161, 1999.
- [31] M.-F. Lei, Z.-M. Tsai, K.-Y. Lin, and H. Wang, "Design and analysis of stacked power amplifier in series-input and series-output configuration," *IEEE Trans. Microw. Theory Tech.*, vol. 55, no. 12, pp. 2802–2812, 2007.
- [32] A. K. Ezzeddine, H. C. Huang, and J. L. Singer, "UHiFET - a new high-frequency high-voltage device," in *Proc. IEEE MTT-S Int. Microw. Symp. Dig.*, 2011, pp. 1–4.
- [33] S. Pornpromlikit, J. Jeong, C. D. Presti, A. Scuderi, and P. M. Asbeck, "A watt-level stacked-FET linear power amplifier in silicon-on-insulator CMOS," *IEEE Trans. Microw. Theory Tech.*, vol. 58, no. 1, pp. 57–64, 2010.
- [34] S. Pornpromlikit, "CMOS RF power amplifier design approaches for wireless communications," Ph.D. dissertation, University of California, San Diego, 2010.

- [35] R. Zhang, M. Acar, M. P. van der Heijden, M. Apostolidou, L. C. N. de Vreede, and D. M. W. Leenaerts, "A 550–1050 MHz +30 dBm class-E power amplifier in 65nm CMOS," in *Proc. IEEE Radio Freq. Integr. Circuits Symp.*, 2011, pp. 1–4.
- [36] A. E. Fathy, S.-W. Lee, and D. Kalokitis, "A simplified design approach for radial power combiners," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 1, pp. 247–255, 2006.
- [37] J. Kim, W. Kim, H. Jeon, Y.-Y. Huang, Y. Yoon, H. Kim, C.-H. Lee, and K. T. Kornegay, "A fully-integrated high-power linear CMOS power amplifier with a parallel-series combining transformer," *IEEE J. Solid-State Circuits*, vol. 47, no. 3, pp. 599–614, 2012.
- [38] C. Y. Law and A.-V. Pham, "A high-gain 60 GHz power amplifier with 20 dBm output power in 90 nm CMOS," in *Proc. IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, 2010, pp. 426–427.
- [39] Q. J. Gu, Z. Xu, and M.-C. F. Chang, "Two-way current-combining W-band power amplifier in 65-nm CMOS," *IEEE Trans. Microw. Theory Tech.*, vol. 60, no. 5, pp. 1365–1374, 2012.
- [40] T. Yao, M. Q. Gordon, K. K. W. Tang, K. H. K. Yau, M.-T. Yang, P. Schvan, and S. P. Voinigescu, "Algorithmic design of CMOS LNAs and PAs for 60-GHz radio," *IEEE J. Solid-State Circuits*, vol. 42, no. 5, pp. 1044–1057, 2007.
- [41] J. Chang, K. Kim, S. Lee, and S. Nam, "24 GHz stacked power amplifier with optimum inter-stage matching using 0.13 μm cmos process," in *Proc. 3rd Int. Synthetic Aperture Radar (APSAR) Asia-Pacific Conf.*, 2011, pp. 1–3.
- [42] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*. Cambridge, 1998.
- [43] H.-T. Dabag, J. Kim, L. E. Larson, J. F. Buckwalter, and P. M. Asbeck, "A 45-GHz SiGe HBT amplifier at greater than 25% efficiency and 30 mW output power," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technol. Meeting*, 2011, pp. 25–28.
- [44] C. F. Campbell and S. A. Brown, "A compact, 40 GHz 0.5 W power amplifier MMIC," in *Proc. 21st Annual GaAs IC Symp.*, 1999, pp. 141–147.
- [45] A. Agah, H. Dabag, B. Hanafi, P. Asbeck, L. Larson, and J. Buckwalter, "A 34% PAE, 18.6dBm 42-45GHz stacked power amplifier in 45nm SOI CMOS," in *Proc. IEEE Radio Freq. Integr. Circuits Symp.*, 2012, pp. 57–60.

- [46] J. Kim, H. Dabag, P. Asbeck, and J. F. Buckwalter, “Q -band and W -band power amplifiers in 45-nm CMOS SOI,” *IEEE Trans. Microw. Theory Tech.*, vol. 60, no. 6, pp. 1870–1877, 2012.
- [47] A. Chakrabarti and H. Krishnaswamy, “High power, high efficiency stacked mmWave class-E-like power amplifiers in 45nm SOI CMOS,” in *Proc. IEEE Custom Integr. Circuits Conf.*, 2012, pp. 1–4.
- [48] M. Fujishima, M. Motoyoshi, K. Katayama, K. Takano, N. Ono, and R. Fujimoto, “98 mW 10 Gbps wireless transceiver chipset with D-band CMOS circuits,” *IEEE J. Solid-State Circuits*, vol. 48, no. 10, pp. 2273–2284, 2013.
- [49] S. Kulkarni and P. Reynaert, “14.3 A push-pull mm-wave power amplifier with $<0.8^\circ$ AM-PM distortion in 40nm CMOS,” in *2014 IEEE International Solid-State Circuits Conference (ISSCC)*, 2014, pp. 252–253.
- [50] B. Hanafi, O. D. Gürbüz, H.-T. Dabag, S. Pornpromlikit, R. Rebeiz, and P. Asbeck, “A CMOS 45 GHz power amplifier with output power >600 mW using spatial power combining,” in *IEEE MTT-S Int. Microw. Symp. Dig.*, 2014, (in press).
- [51] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, “A generalized memory polynomial model for digital predistortion of RF power amplifiers,” *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3852–3860, 2006.
- [52] L. Ding, G. T. Zhou, D. R. Morgan, Z. Ma, J. S. Kenney, J. Kim, and C. R. Giardina, “A robust digital baseband predistorter constructed using memory polynomials,” *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 159–165, 2004.
- [53] P. Draxler, J. Deng, D. Kimball, I. Langmore, and P. M. Asbeck, “Memory effect evaluation and predistortion of power amplifiers,” in *IEEE MTT-S Int. Microw. Symp. Dig.*, 2005.
- [54] E. McCune, *Practical Digital Wireless Signals (The Cambridge RF and Microwave Engineering Series)*. Cambridge University Press, 2010.
- [55] *3GPP TS 25.101*, 3GPP Std., Rev. V12.2.0, 12 2013.
- [56] H. Dabag, B. Hanafi, F. Golcuk, A. Agah, J. F. Buckwalter, and P. M. Asbeck, “Analysis and design of stacked-FET millimeter-wave power amplifiers,” *IEEE Trans. Microw. Theory Tech.*, vol. 61, no. 4, pp. 1543–1556, 2013.
- [57] A. Chakrabarti, J. Sharma, and H. Krishnaswamy, “Dual-output stacked class-EE power amplifiers in 45nm SOI CMOS for Q-band applications,” in *Proc. IEEE Compound Semicond. Integr. Circuit Symp.*, 2012, pp. 1–4.

- [58] A. Agah, W. Wang, P. M. Asbeck, L. Larson, and J. F. Buckwalter, "A 42 to 47-GHz, 8-bit I/Q digital-to-RF converter with 21-dBm Psat and 16% PAE in 45-nm SOI CMOS," in *IEEE Radio Freq. Integ. Circuits Symp.*, 2013, pp. 249–252.
- [59] A. Gupta and J. F. Buckwalter, "Linearity considerations for low-EVM, millimeter-wave direct-conversion modulators," *IEEE Trans. Microw. Theory Tech.*, vol. 60, no. 10, pp. 3272–3285, 2012.