UNIVERSITY OF CALIFORNIA, SAN DIEGO

Power-Combining Techniques for Millimeter-wave Silicon Power Amplifiers

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

in

Electrical Engineering (Electronic Circuits and Systems)

by

Jefy Alex Jayamon

Committee in charge:

Professor Peter M. Asbeck, Chair
Professor James F. Buckwalter
Professor Gert Cauwenberghs
Professor Todd P. Coleman
Professor Gabriel Rebeiz

2017

The dissertation of Jefy Alex Jayamon is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2017

DEDICATION

To my parents and sister.

EPIGRAPH

".. where does the power come from, to see the race to its end ? From within .. I believe God made me for a purpose, but he also made me fast. And when I run I feel His pleasure..."

— *Chariots of Fire* (1981)

TABLE OF CONTENTS

Signature Page iii

Dedication iv

Epigraph v

Table of Contents vi

List of Figures viii

List of Tables xii

Acknowledgements xiii

Vita xviii

Abstract of the Dissertation xx

Chapter 1 Introduction 1
1.1 Design Challenges for mm-Wave PAs 2
1.2 Power Combining Schemes 4
1.3 Dissertation Scope and Organization 7

Chapter 2 Spatial Power-Combined W-band Power Amplifier Using Stacked CMOS SOI 10
2.1 Levels of Power Combining 12
2.2 PA-Antenna Array Design 16
2.3 Experimental Results 25
2.4 Conclusion 41

Chapter 3 Multigate-Cell FET Design 43
3.1 Introduction 43
3.2 Multigate-Cell Architecture 45
3.3 Design Considerations for Multigate Stacked FETs . . . 50
3.4 Multigate-Cell PA Implementation 61
3.5 Measurement Results 64
3.6 Conclusion 71
3.7 Appendix : Approximate Thermal Analysis 74

Chapter 4 Millimeter-wave PMOS Power Amplifier 79
4.1 MOSFET Device Physics 81
4.2 32 nm SOI FET 89
4.3 PA Design 101
4.4 Experimental Results 105
4.5 Comparison with State-of-Art 113
4.6 Conclusion 113

Chapter 5 Conclusions and Future Works 118
5.1 Dissertation Summary 118
5.2 Future Work 120

Bibliography 122

LIST OF FIGURES

Figure 2.1: Schematic structure of the power amplifier-antenna array. . . . 11
Figure 2.2: Levels of Power Combining. 12
Figure 2.3: Different PA-Antenna array configuration a) Single PA- Single Antenna, b) High Gain Array and c) Spatial Power Combiner. . 13
Figure 2.4: Block level schematic from chip input to antenna. 16
Figure 2.5: Schematic of 94 GHz amplifier 3-stack final stage. 18
Figure 2.6: Schematics of final five-stage PA (one half of pseudo-differential PA shown). 20
Figure 2.7: Schematics of two stage Line Amplifier. 21
Figure 2.8: Vertical cross section of CMOS + quartz combination. 22
Figure 2.9: Plan of antenna feed on CMOS and antenna on quartz (Designed by Ozan Gurbuz). 24
Figure 2.10: a) Simulated 2 x 4 antenna array gain and directivity; b) Simulated 2 x 4 antenna array radiation pattern in E- and H- plane. 26
Figure 2.11: Measured and simulated S-Parameters of final stage 3-stack PA (PA1). 27
Figure 2.12: Measured and simulated Gain and PAE vs. Output Power at 89 GHz and 94 GHz of final stage 3-stack PA (PA1). 28
Figure 2.13: Measured maximum output power, PAE and DE vs. Frequency for the final stage 3-stack PA (PA1). 28
Figure 2.14: Measured S-Parameters for 4-stage PA (PA2). 29
Figure 2.15: Measured Gain and PAE of 4-stage single ended (PA2) and 5-stage pseudo differential (PA3) PA at 94 GHz. 30
Figure 2.16: Measured drain current of final stage (3-stack), pre-final stage (3-stack) and driver stages (2-stack, three stages vs. output power at 94 GHz for five-stage pseudo differential PA (PA3). . . 30
Figure 2.17: Measured maximum output power and PAE of the five-stage pseudo differential PA (PA3). 32
Figure 2.18: Measured S-Parameters of two-stage Line Amplifier (PA4) (V_{g1} = 0.4 V, V_{DD} = 2.6 V). 32
Figure 2.19: Measured Gain and drain current of two stages of two-stage Line Amplifier (PA4) at 94 GHz. 33
Figure 2.20: CMOS PA chip-quartz antenna wafer assembly wire-bonded to PCB mounted on copper block. 34
Figure 2.21: Radiation measurement setup. 35
Figure 2.22: Measured and simulated E-plane radiation pattern at 94 GHz. . 36
Figure 2.23: (a) Measured maximum EIRP vs. frequency. (b) Measured quasi-optic gain vs. EIRP at 94 GHz. 37
Figure 2.24: Over-the-air modulation measurement setup (Measurements jointly done with Po-Yi Wu). 39

Figure 2.25: Over-the-air radiation measurements of modulated signals for PA-Antenna array with 256 QAM, 375 MS/s single carrier signal (a) AM-AM (b) AM-PM (c) Emission Spectrum (d) Transmitted constellation. (Red is before DPD, blue is after linearization and green is after FIR filtering) (Measurements jointly done with Po-Yi Wu). 40

Figure 3.1: Representative schematic of a stacked FET PA with gate capacitors. 44
Figure 3.2: Schematic, layout and device cross section of a multigate cell. . 47
Figure 3.3: Multigate-cell layout with FET and ground ring. 48
Figure 3.4: Layout of CG2. 49
Figure 3.5: Arranging the unit cells to form an array. 50
Figure 3.6: Representative FET layout showing parasitic resistances and capacitances on gate. 52
Figure 3.7: Schematic of conventional and multigate 4-stack FET with parasitics associated. 54
Figure 3.8: Thermal dissipation pathway with bitie. 57
Figure 3.9: Quality factor of gate capacitors. 59
Figure 3.10: Load-pull simulation of unit multigate-cell. 62
Figure 3.11: 230 μm FET 4-stack PA (PA1). (a) Die microphotograph. (b) Schematic. 63
Figure 3.12: 307 μm FET 4-stack PA (PA2). (a) Die microphotograph. (b) Schematic. 64
Figure 3.13: Measured (solid line) and simulated (dotted line) S-parameters for PA1. 65
Figure 3.14: Measured (solid line) and simulated (dotted line) S-parameters for PA2. 65
Figure 3.15: Measured gain and PAE of PA1 at high and low bias at 29 GHz. 67
Figure 3.16: Measured gain and PAE of PA2 at high and low bias at 29 GHz. 67
Figure 3.17: Measured saturated output power for PA1 and PA2 (Dots - measured points, thin line - best fit curve. 68
Figure 3.18: Measured peak PAE for PA1 and PA2 (Dots - measured points, thin line - best fit curve. 68
Figure 3.19: 28 GHz Modulated Signal measurement setup. 70
Figure 3.20: PA Output constellation with different modulation schemes and bandwidths. 72
Figure 3.21: Schematic cross section of the chip showing thermal pathways and equivalent thermal resistances. 75

Figure 4.1: Current density of NMOS and PMOS transistors of different generations of IBM CMOS FET. 81

Figure 4.2: Cross section of a PFET showing the SiN liner on gate for inducing compressive stress [1]. 82
Figure 4.3: Simulated hole and electron mobility for (100) and (110) silicon substrates as a function of stress [2]. 83
Figure 4.4: Impact ionization rate of electrons and holes in Silicon. 85
Figure 4.5: Simplified CMOS band diagram showing $Si - SiO_2$ energy barrier for electrons in NFET and holes in PFET. 86
Figure 4.6: Simulated time taken in seconds for a 10% drop in ON current due to HCI for NMOS and PMOS vs. stress voltage. 87
Figure 4.7: Reliability data (Mean Time To Failure - MTTF) of 14/16 nm FinFET published by (a) IBM [3], (b) Intel [4] and (c) TSMC [5]. 88
Figure 4.8: 3-D view of portion of FET wired to top level. 90
Figure 4.9: Measured $I_D - V_{DS}$ for W = 28.8 μm FETs : NMOS ($|V_{GS}|$ = 0 - 0.9 V) and PMOS ($|V_{GS}|$ = 0 - 1.1 V) with $\Delta|V_{GS}|$ = 0.1 V (Solid lines for $|V_{DS}|\&|V_{GS}| \leq 0.9$ V and dotted lines for $|V_{DS}|\&|V_{GS}| > 0.9$ V 91
Figure 4.10: Output power and effective efficiency factor (η_{DC}) of an NMOS and PMOS amplifier biased in class-A with constant V_{DD} and varying load or V_{min}. 92
Figure 4.11: Measured DC transconductance (G_m) of NMOS and PMOS FETs (W = 28.8 μm) vs. current density (ID_{den}) for $|V_{DS}|$ = 0 - 0.9 V, $\Delta|V_{DS}|$ = 0.1 V. 93
Figure 4.12: Measured intrinsic gain (G_m/G_{ds}) of NMOS and PMOS FETs (W = 28.8 μm) vs. current density (ID_{den}) for $|V_{DS}|$ = 0.2 - 1.2 V, $\Delta|V_{DS}|$ = 0.1 V. Gain at $|ID_{den}|$ = 0.5 mA / μm vs $|V_{DS}|$ for both NFET and PFET shown in inset. 94
Figure 4.13: Measured and simulated S-parameters for 28.8 μm NFET - a) S_{11} and S_{22} (smith chart), b)S_{21} (polar plot). 96
Figure 4.14: Measured and simulated S-parameters for 28.8 μm PFET - a) S_{11} and S_{22} (smith chart), b)S_{21} (polar plot). 97
Figure 4.15: Equivalent circuit parameters estimated from measurement and simulation for the 28.8 μm NFET - a) g_m and g_{ds}, b) C_{gs} and C_{gd} and c) R_g and C_{ds} (Solid lines are measurement and dotted lines simulation). 98
Figure 4.16: Equivalent circuit parameters estimated from measurement and simulation for the 28.8 μm PFET - a) g_m and g_{ds}, b) C_{gs} and C_{gd} and c) R_g and C_{ds} (Solid lines are measurement and dotted lines simulation). 99
Figure 4.17: Measured and simulated gain for 28.8 μm NFET - a) short circuit current gain ($|h_{21}|$) and b) maximum available gain (MAG). (Solid lines are measurement and dotted lines are simulation. Thin dotted line in (a) is estimated linear fit for f_t calculation.) 100

Figure 4.18: Measured and simulated gain for 28.8 μm PFET - a) short circuit current gain ($|h_{21}|$) and b) maximum available gain (MAG). (Solid lines are measurement and dotted lines are simulation. Thin dotted line in (a) is estimated linear fit for f_t calculation.) 100
Figure 4.19: Layout of portion of double side gate contacted FET showing finger dimensions (only device layers and contacts shown, metal routings not shown). 102
Figure 4.20: Schematic for (a) NMOS PA (PA1), (b) PMOS PA (PA2) and (c) PMOS PA with inter-stack tuning (PA3). 104
Figure 4.21: Die micro-photograph of 3-stack PMOS PA (PA3) with shunt tuning. 105
Figure 4.22: Measured (solid line) and simulated (dotted line) S-parameters of 3-stack PFET PA with inter-stack tuning (PA3) ($|V_{G1}|$ = 0.35 V, $|V_{DD}|$ = 3.6 V. 106
Figure 4.23: Measured (solid line) and simulated (dotted line) S-parameters of 3-stack NFET PA (PA1) and PFET PA (PA2) - a) S_{11}, b) S_{22} and c) S_{21}. 108
Figure 4.24: Measured and simulated Gain and PAE vs. P_{out} at 78 GHz for PA3 with low and high bias (Low bias : $|V_{G1}|$ = 0.35 V, $|V_{DD}|$ = 3.6 V, High bias : $|V_{G1}|$ = 0.4 V, $|V_{DD}|$ = 4.5 V. 109
Figure 4.25: Measured maximum P_{out}, PAE and DE vs. frequency at low bias for PA3. 110
Figure 4.26: Measured maximum P_{out} and PAE vs. V_{DD} at 78 GHz for PA3. 111
Figure 4.27: Measured Gain and PAE at 78 GHz for two samples of 3-stack NFET PA with low and high bias (Low bias : $|V_{G1}|$ = 0.3 V, $|V_{DD}|$ = 3.0 V, High bias : $|V_{G1}|$ = 0.4 V, $|V_{DD}|$ = 3.6 V. 112
Figure 4.28: Measured Gain and PAE at 78 GHz for two samples of 3-stack PFET PA with low and high bias (Low bias : $|V_{G1}|$ = 0.3 V, $|V_{DD}|$ = 3.6 V, High bias : $|V_{G1}|$ = 0.4 V, $|V_{DD}|$ = 4.2 V. 113
Figure 4.29: Output power vs. time at full power with different supply voltage for a) 3-stack NFET PA (PA1) and b) 3-stack PFET PA (PA2). 114
Figure 4.30: P_{sat}, PAE of state-of-art E-band Silicon PA. 116

LIST OF TABLES

Table 3.1: Comparison of Device and Wiring Parasitics 53
Table 3.2: Comparison of Series Resistance of Stacked FET in Conventional Style and Multigate-cell 55
Table 3.3: Comparison with current State-of-the-Art 69
Table 3.4: Modulated Signal Measurement Results 73

Table 4.1: Estimated equivalent circuit parameters for NFET and PFEt (from simulation of extracted 28.8 μm FET) 101
Table 4.2: Comparison to previously reported Silicon high power PA in V- and E-band 115

ACKNOWLEDGEMENTS

I owe my greatest gratitude to Prof. Peter Asbeck for guiding me through my studies at UCSD. Not only that he had been an excellent advisor, through his own perfect example he taught us how a professor and scientist ought to be. I hope I carry forward with me his insatiable desire to keep on learning new things. Prof. James Buckwalter as my co-advisor taught and helped me to pursue challenging technical goals. I am grateful to learn from him a great deal about identifying and engaging in future research problems. I am greatly indebted to Prof. Gabriel Rebeiz for teaching me Antennas and Phased Arrays, co-advising me in most of my projects and giving me access to his lab resources which was very critical for all my research projects at UCSD. I consider myself very lucky to be at the right place at the right time to be able to learn from all these three excellent teachers.

I want to thank Prof. Gert Cawenberghs and Prof. Todd Coleman for reviewing and providing valuable feedback on my research and thesis. They helped me to take a step back and view, analyse and present my research in a larger perspective.

Most of my research at UCSD was part of and funded by ELASTx program from DARPA which we are grateful to, for both the guidance and support.

Most of the work we did at UCSD was heavily collaborative. I am very thankful to Ozan Gurbuz for his help with the W-band Antenna array designs and Po-Yi Wu and Narek Rostomyan for their help with the 94 GHz and 28 GHz modulation measurements.

Prof. Nagendra Krishnapura (IIT Madras), Prof. Steve Long (UC Santa Barbara) and Prof. Mark Rodwell (UC Santa Barbara) taught me Analog, RF and Millimeter-wave IC design respectively. I owe them all my design fundamentals and more specifically the choice of my current research field.

I am greatly thankful to both my previous managers at TriQuint Semiconductors, Tarun Juneja and Steve LeSage for their invaluable support in applying back to grad school.

I am very fortunate to have immensely helpful and selfless senior grad students who taught me the minute details of IC design and lab measurements. I acknowledge my sincere thanks to Munkyo Seo (UCSB), Bassel Hanafi, Amir Agah, Woorim Shim and Ozgur Inac.

I can never thank enough my friends Adarsh Krishnan, Vimal Raj and Vipindas Pala, who were constant source of motivation and support through out my grad school. Much of the time I have spent outside the lab at San Diego was with my dearest friends Jose Maria and Ananda Theertha, with whom I share most of my cherished memories at UCSD. Also I thank John Alexander for being my best friend and source of inspiration from times as far back as I could remember.

I want to thank all my colleagues who helped and worked with me in so many of the tape-outs - Cooper Levy, Gang Liu, Hayg Dabag, Kelvin Fang, Saeid Daneshgar, Tissana Kijsanayotin, Vincent Leung and Young Pyo Hong. Half the courses I did at UCSD was along with Varish Diddi and Kumar Thasari and it was fun doing all those late night homework submissions.

Stay at UCSD was made all the more happy with all my friends from our research group (HSDG) and co-groups (HSIC and TICS). I especially want to thank Don Kimball, Fuji Takanobu, Hamed Gheidi, Hidenori Yamada, Hideyuki Nakamizo, Jie Min, Johana Yan, John Fairbanks, Kangmu Lee, Kenji Mukai, Mustafa Ozen, Paul Theilmann, Shintaro Shinjo, Shuichi Sakata, Toshi Nakatani, Voravit Vorapipat, Youjiang Liu, Yuji Komatsuzaki, Arpit Gupta, Chris Thomas, Joon Li, Najme Ebrahami, Seyed Rabet, Tim Gathman, Fatih Golcuk, Hasan Alrubaye, Samet Zihir and Tumay Kanar.

I would specially like to thank Paul Draxler and Joonhoi Hur at Qualcomm Research for the guidance they offered me during my summer internship there.

I am grateful to the advices received from and consultations had with Prasad Gudem (Qualcomm), Prof. Sorin Voingescu (University of Toronto) and Dennis Wang (Broadcom).

My schooling for nearly last three decades was so much enriched by awesome friends both in and away from school. I wish to thank all my friends who were part of my school life - Aneesh Reddy, TJ Karthik and other friends at San Diego, Edwin Jose and other La Jolla Malayalee friends, Karthik Srinivasan and other friends at TriQuint Semiconductor, Raibin Raphy and Thomas Mathew and other friends at Portland, Sriram Venkateswaran and other friends at UCSB, all the coolmalloos friends and Ganga Hostel wingmates at IIT Madras, especially Jerrin Jacob Jose and Jubin Jose along with whom I did most of my undergraduate courses, Unnikrishnan K.P and all my other school friends. Above all I express my gratitude beyond words to all my teachers to whose blessings I owe my accomplishments.

The material in this dissertation is based on the following papers.

Chapter 2 is mostly based of materials used in the following publications

The material in preparation to be submitted to *IEEE Transactions on Microwave Theory and Techniques,* J. A. Jayamon, O. D. Gurbuz, P.-Y. Wu, J. F. Buckwalter, G. Rebeiz and P. M. Asbeck, "Spatial Power Combined W-band Power Amplifier using Stacked CMOS SOI with 33 dBm EIRP and 3 Gbps with 256 QAM Modulation". The dissertation author was the primary investigator and author of this material, and co-authors have approved the use of the material for this dissertation. The dissertation author was responsible for the design of the integrated CMOS chip, the overall system assembly

and measurements. The design of the antenna array was conducted by Ozan Gurbuz and Prof. Gabriel M. Rebeiz. The modulated data measurements were carried out with assistance from Po-Yi Wu.

The material as it appears in 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," in *2013 IEEE 13th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems (SiRF),* Jan 2013. The dissertation author was the primary investigator and author of this material, and co-authors have approved the use of the material for this dissertation.

The material as it appears in J. Jayamon, O. Gurbuz, B. Hanafi, A. Agah, J. Buckwalter, G. Rebeiz, and P. Asbeck, "Spatially Power-Combined W-band Power Amplifier Using Stacked CMOS," in *2014 IEEE Radio Frequency Integrated Circuits Symposium,* June 2014. The dissertation author was the primary investigator and author of this material, and co-authors have approved the use of the material for this dissertation.

Chapter 3 is mostly based of materials used in the following publications

The material as it appears in J. A. Jayamon, J. F. Buckwalter, and P. M. Asbeck, "Multigate-cell Stacked FET Design for Millimeter-wave CMOS Power Amplifiers," *IEEE Journal of Solid-State Circuits,* Sept 2016. The dissertation author was the primary investigator and author of this material, and co-authors have approved the use of the material for this dissertation.

The material as it appears in J. A. Jayamon, J. F. Buckwalter, and P. M. Asbeck, "28 GHz $>$ 250 mW CMOS Power Amplifier using multigate-cell design," in *2015 IEEE Compound Semiconductor Integrated Circuit Symposium*

(CSICS), Oct 2015. The dissertation author was the primary investigator and author of this material, and co-authors have approved the use of the material for this dissertation.

Chapter 4 is mostly based of of materials used in the following publications

The material in preparation to be submitted to *IEEE Journal of Solid-State Circuits,* J. A. Jayamon, J. F. Buckwalter, and P. M. Asbeck, "Millimeter-wave PMOS Power Amplifier". The dissertation author was the primary investigator and author of this material, and co-authors have approved the use of the material for this dissertation.

The material as it appears in J. A. Jayamon, J. F. Buckwalter, and P. M. Asbeck, "A PMOS mm-wave Power Amplifier at 77 GHz with 90 mw Output Power and 24% Effiency," in *2016 IEEE Radio Frequency Integrated Circuits Symposium (RFIC),* May 2016. The dissertation author was the primary investigator and author of this material, and co-authors have approved the use of the material for this dissertation.

VITA

2006	B. Tech. in Electrical Engineering, Indian Institute of Technology Madras
2007	M. S. in Electrical Engineering (Electronics and Photonics), University of California, Santa Barbara
2008-11	TriQuint Semiconductors, Hillsboro, Oregon
2017	Ph. D. in Electrical Engineering (Electronic Circuits and Systems), University of California, San Diego

PUBLICATIONS

J. A. Jayamon, J. F. Buckwalter and P. M. Asbeck, "Millimeter-wave PMOS Power Amplifier," to be submitted to *IEEE Journal of Solid-State Circuits*.

J. A. Jayamon, O. D. Gurbuz, P.-Y. Wu, J. F. Buckwalter, G. Rebeiz and P. M. Asbeck, "Spatial Power Combined W-band Power Amplifier using Stacked CMOS SOI with 33 dBm EIRP and 3 Gbps with 256 QAM Modulation," to be submitted to *IEEE Transactions on Microwave Theory and Techniques*.

N. Rostomyan, J. A. Jayamon, and P. M. Asbeck, "15 GHz High Power Doherty PA with RF Predistortion Linearizer in CMOS SOI," accepted for publication at *IEEE Transactions on Microwave Theory and Techniques*.

N. Rostomyan, J. A. Jayamon, and P. M. Asbeck, "15 GHz 25 dBm Multigate-cell Stacked CMOS Power Amplifier with 32% PAE and $\geq$ 30 dB gain for 5G applications," in *2016 11th European Microwave Integrated Circuits Conference (EuMIC),* Oct 2016, pp. 265-268

J. A. Jayamon, J. F. Buckwalter, and P. M. Asbeck, "Multigate-cell Stacked FET Design for Millimeter-wave CMOS Power Amplifiers," *IEEE Journal of Solid-State Circuits,* vol. 51, no. 9, pp. 2027-2039, Sept 2016.

J. A. Jayamon, J. F. Buckwalter, and P. M. Asbeck, "A PMOS mm-wave Power Amplifier at 77 GHz with 90 mw Output Power and 24% Effiency," in *2016 IEEE Radio Frequency Integrated Circuits Sym posium (RFIC),* May 2016, pp. 262-265.

J. F. Buckwalter, S. Daneshgar, J. Jayamon, and P. Asbeck, "Series power combining: Enabling techniques for Si/SiGe Millimeter-wave Power Amplifiers," in *2016 IEEE 16th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems (SiRF),* Jan 2016, pp. 116-119.

J. A. Jayamon, J. F. Buckwalter, and P. M. Asbeck, "28 GHz > 250 mW CMOS Power Amplifier using multigate-cell design," in *2015 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS),* Oct 2015, pp. 1-4.

G. Liu, J. Jayamon, J. Buckwalter, and P. Asbeck, "Frequency Doublers with 10.2/5.2 dBm Peak Power at 100/202 GHz in 45nm SOI CMOS," in *2015 IEEE Radio Frequency Integrated Circuits Symposium (RFIC),* May 2015, pp. 271-274.

J. Jayamon, O. Gurbuz, B. Hanafi, A. Agah, J. Buckwalter, G. Rebeiz, and P. Asbeck, "Spatially Power-Combined W-band Power Amplifier Using Stacked CMOS," in *2014 IEEE Radio Frequency Integrated Circuits Symposium,* June 2014, pp. 151-154.

A. Agah, J. A. Jayamon, P. M. Asbeck, L. E. Larson, and J. F. Buckwalter, "Multi-drive Stacked-FET Power Amplifiers at 90 GHz in 45 nm SOI CMOS," *IEEE Journal of Solid-State Circuits,* vol. 49, no. 5, pp. 1148-1157, May 2014.

A. Agah, J. Jayamon, P. Asbeck, J. Buckwalter, and L. Larson, "A 15.8 dBm Two-stage 90 GHz Stacked-FET Power Amplifier in 45-nm SOI CMOS," in *2013 IEEE MTT-S International Microwave Symposium Digest (MTT),* June 2013, pp. 1-3.

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," in *2013 IEEE 13th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems (SiRF),* Jan 2013, pp. 79-81.

ABSTRACT OF THE DISSERTATION

Power-Combining Techniques for Millimeter-wave Silicon Power Amplifiers

by

Jefy Alex Jayamon

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

University of California, San Diego, 2017

Professor Peter M. Asbeck, Chair

Emerging millimeter-wave applications, including high speed wireless communication using 5G standards, favor silicon technologies, both CMOS and SiGe, for transceiver design, due to the high level of integration at reduced cost and availability of high speed transistors. Efficient, linear and reliable high power amplifiers with broad bandwidth are needed at the transmitter front-ends to enable high data rate links at long distances. But the low breakdown voltage of CMOS FETs due to gate length scaling and other transistor non-idealities make the design of high power mm-wave amplifiers in deeply scaled CMOS nodes difficult. Circuit

techniques like FET stacking provide a compact and efficient way of implementing high power mm-wave amplifiers reliably. Other power combining techniques such as on-chip and spatial power combining can be used along with FET stacking to achieve even higher output power levels. This thesis investigates the design of high power mm-wave power amplifiers at frequencies from 28 GHz to 94 GHz, using multiple power combining techniques.

This work extends the use of FET stacking for high power PA design to 94 GHz. A 3-stack PA designed in 45 nm CMOS SOI with 17 dBm output power and 9% efficiency is presented. Using this PA as front-end, a CMOS PA-antenna array is designed, to additionally provide spatial power combining. The CMOS chip has a 2 x 4 array of pseudo-differential power amplifiers along with the signal distribution networks and pre-drivers. A quartz wafer with a 2 x 4 array of differential microstrip antennas deposited on it is placed on top of the CMOS chip, electromagnetically coupled to the PA outputs on the CMOS chip. The spatially power combined PA-antenna array achieved a measured equivalent isotropic radiated power (EIRP) of 33 dBm and an estimated output power of 24 dBm at 94 GHz. Modulated data measurements at 3 Gbps (375 MS/s, 256 QAM) speed using digital pre-distortion are demonstrated with the PA-antenna array.

A novel layout style is introduced for stacked FET design at low mm-wave frequencies. A small multi-finger FET is laid out with fingers connected in series to create the stacked FET. The gate capacitors are realized around the FET with the back-end-of-line metal available in the CMOS process. Multiple multigate cells are interconnected to implement the stacked FET PA. A PA designed in this style in 45 nm CMOS SOI process achieved 24.8 dBm of output power and 29% PAE at 28 GHz with high reliability. This PA is very broadband and linear as shown by the modulated data measurements achieving a data rate of 36 Gbps (6 GS/s,

64 QAM) at 14 dBm with 9.3% PAE, with no digital predistortion.

NFETs and PFETs available in nano-scale CMOS processes are compared and it is shown that in deeply scaled processes, PMOS devices are a viable alternative to NFETs due to their cut-off frequencies similar to those of NFETs, and higher breakdown voltages than NFETs. The first exclusively PMOS mm-wave PA design is reported. This 3-stack PA, made in 32 nm CMOS SOI process, achieved a maximum output power of 19.6 dBm and maximum efficiency of 24% at 78 GHz.

All the designs reported in this thesis achieved either the highest output power or the highest PAE for a CMOS PA at their respective frequencies.

Chapter 1

Introduction

Highly integrated millimeter-wave transceivers with high output power and efficiency are needed for emerging wireless communication (broadcast/point-to-point), imaging (medical/industrial/tactical) and radar (automotive/backscatter) applications. Extreme scaling of MOSFET gate lengths has led to steady increase of operating frequencies for CMOS devices. Along with scaling, other process improvements including material system engineering and structural changes, have helped the Silicon transistors, both SiGe HBTs as well as CMOS FETs, to have cut-off frequencies (f_t/f_{max}) beyond 500 GHz. This has enabled the transistors to have high enough gain to have compact signal generation and amplification capabilities at high mm-wave frequencies. Thus in the past decade, silicon devices have become the favored technology choice for highly integrated mm-wave wireless front ends.

At RF frequencies, since the wireless capacity is primarily limited by interference, the linearity of the front-end is the primary concern. But mm-wave wireless links are primarily limited not by interference but by Signal-to-Noise Ratio due to the high spatial spreading loss and atmospheric attenuation [6]. In this

scenario, as given by Shannon's theorem, the link capacity is directly related to the received SNR [7] and hence the power transmitted. Therefore having capability for high generated and transmitted power levels is very critical to achieve high data rates. The upcoming fifth generation (5G) mobile wireless communication standards envision mm-wave links with effective isotropic radiated power (EIRP) of up to 75 dBm/100 MHz for base stations and 43 dBm for mobile units. These are going to be implemented as antenna array solutions using spatial power combining. Assuming a 2 dB post PA loss, a 29 dBm unit PA is needed for a 256 element array for base station front end and similarly a 21 dBm unit PA is needed for a 16 element array for mobile unit front end to reach the above specified maximum EIRP limits. Since the power amplifiers could constitute nearly half of the total power dissipation of the transmitter, the efficiency of the PA is very important to have viable mobile implementations. To be able to use spectrally efficient higher order QAM signals the PA should have good linearity. Having broadband or frequency-tunable PA helps to reduce the complexity of multi-band multi-standard transceivers. Use of modulation schemes with high peak-to-average-power-ratio (PAPR), such as OFDM, necessitates the need to have power amplifiers with high efficiency at backed-off power levels also.

1.1 Design Challenges for mm-Wave PAs

For mm-wave systems with multiple antennas, the implementation of CMOS-based power amplifiers with output power in the range 50 - 1000 mW has significant interest, particularly using low cost approaches that enable integration with other transceiver components. However, dimensional scaling leads to a reduction of the transistor breakdown voltage [8]. This can be roughly described by the Johnson

Figure-of-Merit (JFoM) which is the product of charge carrier saturation velocity (v_{sat}) and breakdown field (E_{BD}). For a given combination of process (material system) and charge carrier this can be translated to the product of current gain cut-off frequency (f_t) and breakdown voltage (V_{BD}) which tends to be a constant for a given material. The restricted voltage handling of devices limits the attainable output power. The output impedance matching network required to translate to 50 Ω outputs further increases loss and reduces bandwidth. Non-ideal ID-VDS characteristics, like higher knee voltage as a fraction of supply voltage (V_{knee}/V_{DD}) and lower output resistance (r_o), of the short channel FETs reduce the maximum PAE achievable with these devices. Also at mm-wave frequencies the increased impact of parasitics leads to higher loss and non-ideal current-voltage switching. This makes low frequency efficiency enhancement techniques like wave-shaping by harmonic control difficult to be implemented at mm-wave frequencies.

Even though the fineline CMOS processes can have relatively high mobility and hence transconductance (g_m), the thin gate-length and narrow interconnects increase the gate resistance. The fine pitch contacts and interconnects increases the parasitic gate capacitances also. This causes the cut-off frequencies of the large microwave transistors with the interconnects to the top metal layer, as used in mm-wave PA design, to be much lower than that of the intrinsic device. A higher value of gate-to-drain capacitance (C_{gd}) makes the device bilateral, reduces the gain and increases the design difficulty. Also the reliability and thermal modelling of high power transistors are not usually well captured in the foundry device models. The quality factor of the passive components are low at the mm-wave frequencies due to the increase in the insertion loss of the interconnects due to skin effect, surface roughness of the metal wires (at high mm-wave frequencies $>$ 100 GHz) and the presence of dummy fill due to silicon density rules.

1.2 Power Combining Schemes

The output power scales directly proportional to the width of the transistor for a simple common source amplifier. The need for power combining schemes is due to the inefficiencies which arise from scaling the transistor width for higher power. This is limited by two factors; i). due to wiring parasitics the value of maximum oscillation frequency (f_{max}) and hence the gain at design frequency decreases; and ii). if the FET optimum load impedance is very different from the desired output impedance (usually 50 Ω), the required matching network would have higher loss thereby reducing efficiency. Also large impedance-transformation-ratio matching networks cause narrowband frequency response. A device designed to have optimum load impedance of 50 Ω, with a supply voltage of 1 V, assuming ideal class-A configuration, could achieve 10 dBm of output power. One can scale the device width even larger with some penalty of f_{max}, but the output matching losses would increase considerably. Assuming finite values for quality factor ($Q_{passive}$) of matching elements, FET knee voltage (V_{knee}) and FET output conductance (g_{ds}), an upper limit of about 13 - 14 dBm exists for common source amplifiers if an output matching network of Q less than 4 - 5 is needed.

A variety of design strategies have been discussed to overcome this limitation, including on-chip transmission-line [9,10], LC-based power combining [11,12], transformer-based power combining [13, 14], radial power combiner [15], spatial power combining [16, 17] and FET stacking [18–22].

1.2.1 Stacking and On-Chip Power Combining

Among these techniques, FET stacking (Fig. 1.1) leads to the most compact amplifiers, and has advantages for increased efficiency because of the potentially

low additional losses introduced for power combining.

In stacked FET design the impedance seen out of the drain of a FET into the source of the FET stacked above scales linearly with the number of stacked FETs. This is achieved by proper scaling of the gate capacitance of the FET above in the stack [23]. Stacking k FETs allows the use of FETs with k times width as of common source for the same output load impedance. In ideal case this would result in k^2 times output power compared to common source FET. At mm-wave frequencies the impedance seen at the drain has a considerable imaginary part due to the parasitic capacitances at the drain node as well as parasitic series resistance of the gate capacitor. This can cause impedance mismatch at the drain node and misalignment of the drain voltages along the stack, leading to lower output power and efficiency. Also some of the drain current is lost to ground through the shunt parasitics. Various techniques including reactive tuning of the drain nodes [23,24], variable sizing of the FETs in the stack [25] and active driving of top gates [26] have been attempted to improve the efficiencies at mm-wave frequencies. With less loss than on-chip power combining one could stack up to 4 - 5 FETs at low mm-wave frequencies (20 - 50 GHz / Ka, Q band), 3 - 4 FETs are medium mm-wave frequencies (50 - 110 GHz / V, W band) and 2 FETs at higher mm-wave frequencies (110 - 140 GHz / D band). Beyond that, cascode or common source is the optimum choice as the design becomes gain limited. With stacking FETs the power gain grows linearly whereas with cascading same number of FETs the gain grows exponentially. Due to the availability of thick back-end-of-line (BEOL) and (semi)-insulating substrate, on-chip power combining is ideal for III-V processes. Due to the lack of these two in CMOS, FET stacking is more advantageous. Stacked FETs occupy much less area compared to on-chip power combining. Stacking of 2/3/4 FETs is equivalent to 4-/9-/16- way on-chip power combining. Voltages

higher than nominal supply voltage of the process need to be generated for drain and gate biasing of stack FET PA. On-chip power combining leads to thermally better layout as the unit cells are now distributed across a larger area leading to less heat concentration and cross heating. FET stacking leads to thermally worse layout if the FETs are laid out in a small area.

1.2.2 Spatial Power Combining

Spatial/quasi-optic power combining is the most efficient power combing scheme. Ideally, spatial power combining is lossless and can be scaled to any number of elements. It causes minimum coupling/load pulling between the individual PAs. Addition of phase shifters can make the transmitter into a beam-former. For an N element array the output power is increased N fold and EIRP is increased N^2 fold relative to a single element. In SNR-limited links this leads to N fold scaling of distance as suggested by Friis equation. Separate control of antenna array elements would enable techniques like windowing/weighing of unit element power levels for lower side-lobes and selective element choosing for specific radiation properties [27]. Spatial power combining allows graceful power degradation in case of failure of individual elements. In transceivers the antenna array gives receive gain also. The narrow beam width caused by large scale arrays may not be desirable for broadcast applications. At low frequencies spatial power combining is area inefficient due to half wavelength spacing requirement of the antenna elements. Separate and redundant matching network for PA and antenna can lead to suboptimal efficiency if PA and antenna are not co-designed which is rarely the case due to the need to have both characterized separately. Conventionally multi-chip solutions suffer from difficulty of routing LO to different chips with accurate phase. Phase shifters with high enough number of bits or injection locking tech-

niques can be used to address this. Also very large scale arrays to be used in beam former fashion would need true time phase-shifters if large bandwidth (inverse of which is order of magnitude close to the array length) signal modulation is to be used.

1.3 Dissertation Scope and Organization

This thesis studies the realization of high power mm-wave power amplifiers in scaled CMOS nodes at mm-wave frequencies ranging from 25 GHz to 100 GHz. Design techniques to improve key PA performance metrics - output power, efficiency, linearity, bandwidth and reliability, are investigated. The trade-offs involved in choice of device types, the layout styles and the power combining architectures are explored. The study makes heavy use of the stacked FET technique to achieve high output power levels with high efficiency and reliability at mm-wave frequencies. Efforts are made to ensure the designs are suitable for highly integrated MIMO systems. Optimal combination of different power combining schemes together to achieve the required output power and other specifications is studied. Improved layout styles and non-classical device options are investigated for better reliability of the high power mm-wave PAs. Along with desired output power and efficiency, the PA designs are optimized to have the linearity and bandwidth needed for high speed mm-wave wireless communication.

Chapter 2 presents a spatial power-combined power amplifier antenna array at 94 GHz in 45 nm CMOS SOI process. The power amplifiers are implemented as 3- stack FET multi-stage pseudo-differential PAs. The output of each PA is coupled to an antenna on a quartz substrate placed on top of the CMOS chip. The CMOS chip consists of an array of 2 x 4 PA channels, all operating at the

same phase. The quartz wafer has the 2 x 4 differential microstrip antenna array (with half wavelength spacing between the antenna elements) deposited on top of it. Wafer probed output power measurements of the individual PA and the drivers as well as radiation measurements of the whole array are presented. Modulated data measurements of the transmitter array with digitally pre-distorted 256 QAM signals are reported. Limits of different power combining schemes are also studied.

Chapter 3 presents a novel multi-gate cell design for CMOS SOI FET stacking. Instead of conventionally-used large width lumped FETs for stacking, a small multi-finger FET with multiple gate connections is designed. Gate capacitances needed for the stacked FET amplifier are implemented around the FET as metal-over-metal capacitors using the back-end-of-line metals available in the CMOS process. Many of these unit cells are wired together to implement the equivalent large FET width required for the PA. A high power PA was designed using this configuration at 28 GHz. Both output power measurements as well as broadband modulated signal measurements of the PA are presented. Thermal analysis of the heat dissipation pathways in the PA layout is discussed and shown to be a prime contributing factor to the demonstrated high reliability of the PA.

At deeply scaled nodes, below 40 nm, PMOS FETs offer a viable alternative to NMOS FETs and SiGe HBTs as a fast, high voltage device. Chapter 4 explores the device physics behind the high cut-off frequencies and reliability of MOSFETs and provides a comparison of NMOS and PMOS performance. Device measurements of large transistors are done and PMOS FETs are shown to be as fast as and more reliable than NMOS FETs. As a demonstration a 3-stack PMOS PA is designed in 32 nm CMOS SOI process at 78 GHz and the power measurements are presented. A controlled experiment with similar NMOS and PMOS stacked FET PA is also presented at the same frequency.

The thesis concludes with a summary of the research and discussion of further research topics related to the thesis.

Chapter 2

Spatial Power-Combined W-band Power Amplifier Using Stacked CMOS SOI

Spatial power combining of silicon amplifiers has previously been demonstrated in wafer-scale using SiGe at W-band [17] and in CMOS SOI at Q-band [28] using antenna array on PCB. Reference [17] uses SiGe HBT in common emitter configuration for a power amplifier. Due to the lower breakdown voltage of scaled CMOS FETs a similar power amplifier array implementation using CMOS needs to use more than one power combining scheme to achieve the same output power. This work uses FET stacking and differential antenna drive to increase the output power capability (also used in [28]). In this work a 94 GHz integrated power amplifier antenna solution is demonstrated. This implements spatial power combining of CMOS power amplifiers coupled to a 2 x 4 array of differential microstrip antenna on a quartz substrate on top of the silicon chip (Fig. 2.1). Each of the unit power amplifiers in-turn is implemented in pseudo-differential fashion and

uses FET stacking to achieve higher output power. The silicon chip contains the front-end power amplifiers, driver amplifiers, the signal distribution network and the antenna feed. The chip occupies 4.3 mm x 6.0 mm area and achieves peak output power of 24 dBm and Effective Isotropic Radiated Power (EIRP) of 33 dBm at 94 GHz. Over-the-air modulation measurements of the PA-Antenna array using Digital Predistortion (DPD) demonstrates 375 MS/s of 256 QAM (3 Gbps) signals with 2.5% EVM and -32 dBc ACPR. This chapter discusses the trade-offs in different power combining schemes and uses the measurements of sub-blocks of the amplifier array to investigate the discrepancy between measured and simulated power measurements.

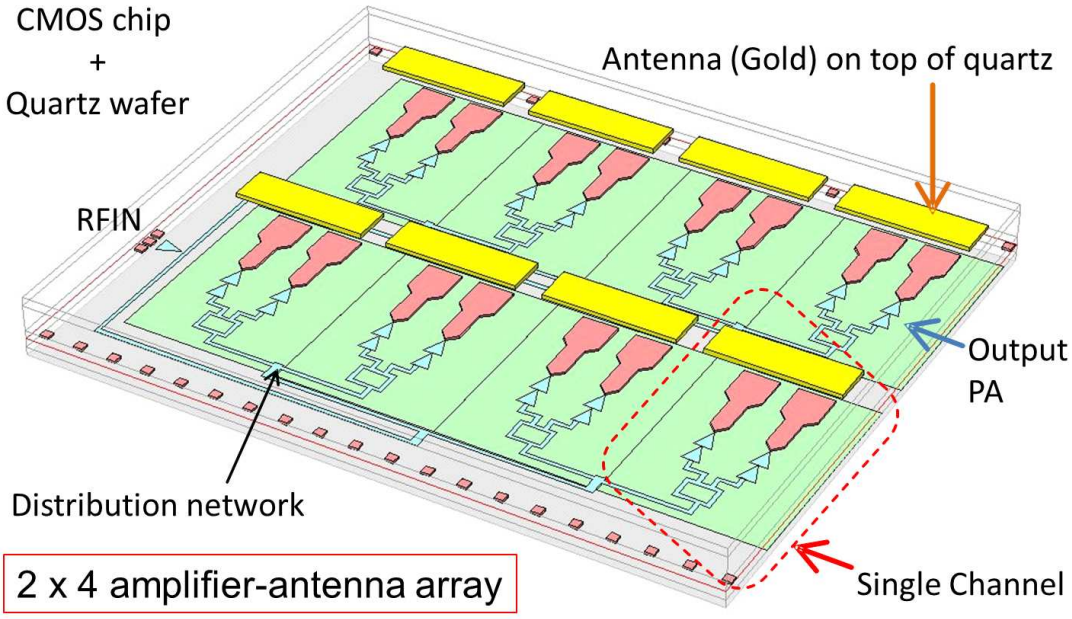

Figure 2.1: Schematic structure of the power amplifier-antenna array.

In the following, Section 2.1 discusses various power combining schemes and their trade-offs. The PA-Antenna Array design is presented in Section 2.2 and measurement results in Section 2.3.

2.1 Levels of Power Combining

The common mm-wave power combining techniques ranging from the device level to system level are broadly - FET scaling, FET stacking, on-chip power combining, multi-driven antenna and spatial power combining (Fig. 2.2).

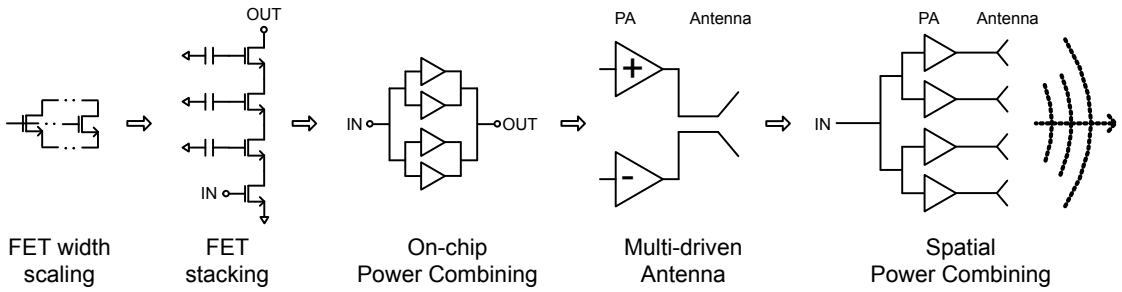

Figure 2.2: Levels of Power Combining.

Scaling the transistor width for higher power is limited by - i) reduction of f_{max} and hence gain - due to wiring parasitics; and ii) increase in the impedance-transformation-ratio of the output matching network (Q_{MN}) and resulting higher loss and lower bandwidth.

Stacking k FETs allows the use of FET with k times width as of common source for the same output load impedance. In ideal case this would result in k^2 times output power compared to common source FET [18]. At mm-wave frequencies FET stacking suffers from impedance mismatch between different FETs on the stack. Also stacking with constant output impedance needs FET scaling which would lead to lower f_{max}. The number FETs which can be stacked without more loss of PAE than on-chip power combining in that frequency, drop from 4-5 at low mm-wave frequencies (30 GHz) to 2-3 at high mm-wave frequencies (140 GHz). At even higher frequencies, cascode or common source is more desirable as the design becomes gain limited. Due to the availability of thick Back-End-Of-Line (BEOL) and (semi)-insulating substrate, on-chip power combining is ideal

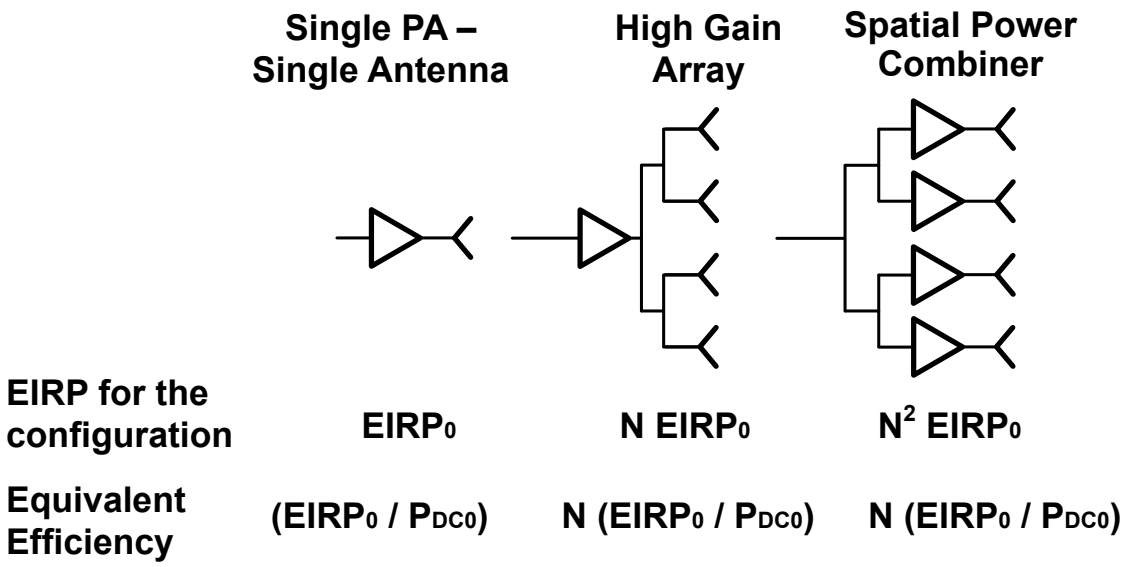

Figure 2.3: Different PA-Antenna array configuration a) Single PA- Single Antenna, b) High Gain Array and c) Spatial Power Combiner.

for III-V processes. Assuming simple binary tree power combining scheme (like Wilkinson Combiner) with a dB insertion loss per 2:1 combining, for an N level combining $(2^N : 1)$ the output power scales as $P_0 + (3 - a)N$ dBm, gain scales down as $G_0 - 2Na$ dB and drain efficiency scales down as $\frac{\eta_0}{10^{(\frac{Na}{10})}}$ % where P_0, G_0 and η_0 are output power in dBm, gain in dB and efficiency in percentage for a unit PA.

On-chip power combining networks can be merged with on-chip antenna to have multi-driven antenna [29,30]. The simplest case would be going from a single ended antenna fed from one PA to a differential antenna fed from two PAs with 180^0 phase difference. This can be extended to N PA, each with $\frac{360}{N}$ degree phase difference feeding into the same antenna at appropriate location determined by phase/input impedance. Also co-designing the PA and antenna can eliminate the need for separate matching networks for each.

Typical cases of PA-Antenna array configurations are - Single PA-Single Antenna (SPSA), High Gain Array (HGA) and Spatial Power Combiner (SPC)

(Fig. 2.3). The EIRP and the equivalent efficiency ($EIRP/P_{DC}$) of each case are shown in the figure .

By using the same unit PA element and unit antenna element, in order to have M times EIRP than SPSA, either one PA and M antenna in HGA configuration, or $\sqrt{M}$ PAs and $\sqrt{M}$ antenna in SPC configuration are needed. The $EIRP/P_{DC}$ of HGA is M times as of SPSA and $\sqrt{M}$ times for SPC for the same EIRP (assuming same PA efficiency). So if there are no constraints on the beamwidth and the output power of the PA being used, in order to the maximize $EIRP/P_{DC}$, it is best to use maximum number of antenna possible and create a high gain array. But this assumes lossless routing at the PA output to the different antenna elements in the array. In a more realistic scenario usually the routing losses limit the maximum number of antenna elements which can be used. Generically for a k-dimensional arrangement of N elements with at unit spacing d, for equi-phase binary split signal distribution from a single source, the path length is $k\sqrt[k]{N}d$. Since the antenna elements are usually placed at $\frac{\lambda}{2}$ spacing for minimum side-lobes, the routing distances for N element linear array is about $\frac{(N\lambda)}{4}$. For an N element square (2D) array it is $2\sqrt{N}\frac{\lambda}{4}$. If we assume routing loss of $\alpha\frac{dB}{\lambda}$, for a one PA - N antenna linear array, the EIRP can be given as

$$EIRP_{dBm,1D} = P_{0,dBm} + G_{ant.ele,dB} + 10\,log(N) - \frac{\alpha\,N}{4} \tag{2.1}$$

For a square array the EIRP would be,

$$EIRP_{dBm,2D} = P_{0,dBm} + G_{ant.ele,dB} + 10\,log(N) - \frac{\alpha\,2\,\sqrt{N}}{4} \tag{2.2}$$

The value of EIRP increase is maximized for a linear array and 2D array respectively at

$$N_{max,lineararray} = \frac{10 \times 4}{2.3\alpha} \tag{2.3}$$

$$N_{max,squarearray} = \left(\frac{10 \times 4}{2.3\alpha}\right)^2 \tag{2.4}$$

This happens around N=16 for $\alpha = 1\ \frac{dB}{\lambda}$ and around $N = 8$ for $\alpha = 2\ \frac{dB}{\lambda}$ for a linear array. For a square array one can have $N = 256$ for $\alpha = 1\ \frac{dB}{\lambda}$.

For SPC there is very little output routing loss as the PA can be placed very close to antenna. Hence if PA has high gain (to compensate for input routing losses), the EIRP always increases with number of PA-Antenna elements (N).

$$EIRP_{dBm,SPC} = P_{0,dBm} + G_{ant.ele,dB} + 20\,log(N) \tag{2.5}$$

The analysis above assumes no constraint on N as well as the output power capability of the PA. But in reality, due to the form-factor limitation of the final implementation, the area available for antenna array and hence the value of N has an upper bound. This leads us to the question of what is the best technology option for highest efficiency operation. We can see from eq 2.1 and 2.5 that for comparing high power PA used in HGA vs. low power PA used in SPC, to achieve same EIRP, we should be comparing efficiencies of high power PA with output power $P_{0,dBm}$ in one PA - N Antenna HGA against low power PA with output power $P_{0,dBm} - 10\,log(N) - \frac{\alpha N}{4}$ in an N PA - N Antenna system (SPC). For example, assuming linear array with $\alpha = 1\ \frac{dB}{\lambda}$, both an eight element HGA using a GaN PA of 30 dBm and an eight element PA-Antenna array of Silicon PA of 19 dBm gives same EIRP. Looking at current state-of-art we can see that [31] reports 30 dBm GaN PA at 94 GHz with 15% PAE and [26] reports 19 dBm CMOS SOI PA at 90 GHz with 15% PAE. This shows that both technologies can yield very similar

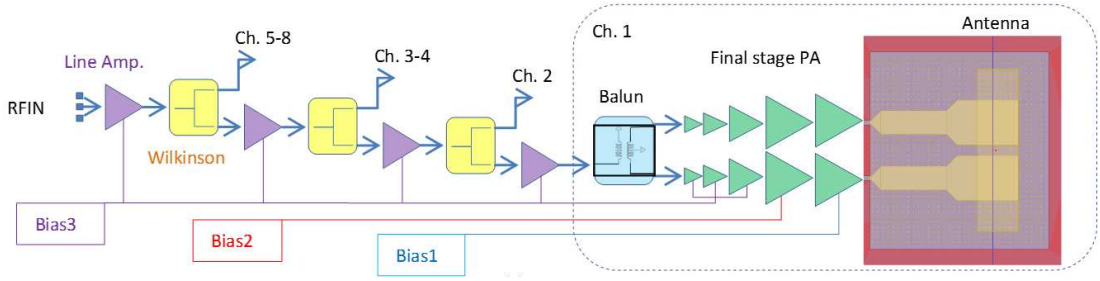

Figure 2.4: Block level schematic from chip input to antenna.

results with the given set of requirements ($N \leq 8$).

2.2 PA-Antenna Array Design

The amplifier design reported in this chapter uses FET scaling, stacking, multi-driven antenna and spatial combining.

2.2.1 System Architecture

The spatial power combining chip is designed as a 2 x 4 array of PA + Antenna. The differential antenna feed, the pseudo-differential five stage power amplifier and the input balun form one channel on-chip. The chip has eight such channels. The chip also contains a signal distribution network which takes the single input signal and divides and feeds it to each of the eight channels. The distribution network is made of three stages of Wilkinson dividers to create a 1:8 split and two-stage cascode line amplifiers in between to compensate for the routing losses. There is a line amplifier at the input of the chip and after every Wilkinson divider. Thus from the chip input to the antenna feed the signal sees thirteen stages of amplifiers (Fig. 2.4). A 100 μm quartz substrate placed on top of the CMOS chip has the antenna array etched in gold on top of it. The antenna feed on CMOS chip capacitively couples with the antenna on top and facilitates

the radiation. The input to the chip is provided by wafer-probing and output is radiated. DC power supply is provided by wirebonds on the sides of the chip.

2.2.2 PA Design

The PA design consists of the 3-stack final stage, the 2-stack/cascode driver stages, the cascode line amplifiers, Wilkinson divider, balun and transmission lines for signal distribution.

Final Stage

IBM 45 nm CMOS SOI process is used for the PA design. Regular V_t, floating body, 1 μm wide finger, single-side gate-contacted, relaxed gate-pitch FETs are used. The measured f_{max} of the FETs (wired up to top metal layer) is 210 GHz [26]. The final two stages of the amplifier chain are designed as 3-stack PA. The PA size is chosen so that the optimum output impedance is close to 50 Ω and hence no output impedance transformation is required. A 4-stack FET ideally should achieve more than 2 dB more power than a 3-stack FET. But due to the lower gain per FET and increased mismatch, the 4-stack PA simulations show only less than 1 dB more output power and almost 4% lower PAE than 3-stack PA. Therefore 3-stack is an optimum choice at this frequency to maximize the output power while still maintaining $>$ 10% PAE.

Intra-stack tuning (shunt inductance implemented as a shorted transmission line) is used at drain of first (common source / bottom) FET to provide impedance match between stacked FETs. A similar tuning is not used at the drain of second FET as the PAE improvement thus would be only less than 1%, but the layout complexity increases. The top common gate FETs of the stack are sized smaller than the bottom common source FET (192 μm vs. 256 μm) [25]. This helps us to

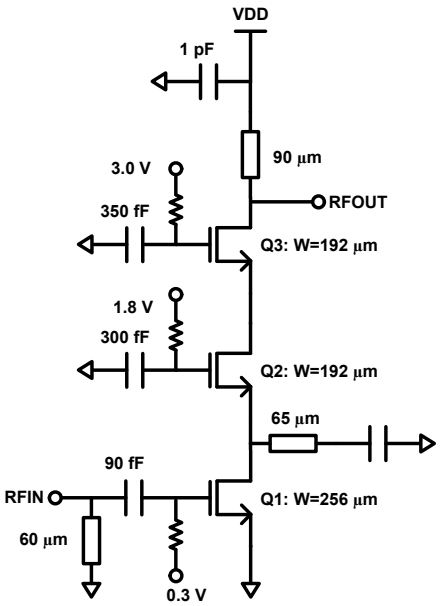

Figure 2.5: Schematic of 94 GHz amplifier 3-stack final stage.

improve the gain by having smaller and hence higher f_{max} FETs as common gate device, while still having the large transconductance (g_m) provided by the larger common source device. Both the skewing of the common gate device width and the shunt tuning help to impedance match the intra-stack node in addition to the capacitive degeneration provided by the finite small value gate capacitor. A shunt transmission line (implemented using CPW) acts as the RF choke for drain supply as well as to tune out the capacitance at the top drain node. This brings the output node impedance to 50 Ω and hence no further impedance transformation is needed. For simulation the FETs are parasitic (RC) extracted along with the lower thin metal layer interconnects (eight layers with thickness < 250 nm, spaced at interlayer dielectric < 250 nm). The higher thick metal layer interconnects (three layers with thickness > 1.2 μm), routing and passive structures like transmission lines are simulated with EM tool (Sonnet) and S-parameter models are used. The 3-stack PA designed (Fig. 2.5) has a simulated output power of 18 dBm and PAE of 12%.

An identical 3-stack amplifier is designed as a pre-final stage to drive the final stage PA. This is much less efficient than using a smaller size 3-stack or a 2-stack. But the 3-stack PA can be operated under a wider range of supply voltages yielding a > 1 dB range of maximum output power. Thus overdesigning the pre-final stage makes sure that the final stage can be saturated even under non-ideal conditions at the cost of PAE dropping to 8% from 12%.

Driver Stages

A three stage 2-stack/cascode driver chain is designed to drive the final two stages (Fig. 2.6). For ideal cascode amplifier, the gate capacitor (CG_2) of the top transistor (common gate FET), should be large enough to provide a very low impedance at the gate. Hence $CG_2 \gg C_{gs2}$ for a cascode amplifier. For ideal 2-stack amplifier, the value of CG_2 is nearly equal to C_{gs2} for reliable operation [23, 25]. The final stage of the amplifier chain is designed to be in deep class-AB mode and is expected to operate into saturation for higher output power and efficiency. This causes high voltage and current swings at the FET nodes and hence the amplifier has to be designed for reliable operation following ideal FET stacking methodology. The initial stages of the amplifier chain are designed in class-A mode for linearity and high gain and hence designed more like cascode stages. In this 45 nm CMOS SOI process, the value of C_{gs} is about 0.8 fF/μm. The designed value of CG_2 varies from 1.5 fF/μm for final stage to 7 fF/μm for first stage of the amplifier chain and 9.3 fF/μm for the first stage of the line amplifier. Thus the design-class varies from 2-stack at the front end (with large FETs, higher voltage swings and lower gain per stage) to cascode at the back end (with small FETs, lower voltage swings and higher gain per stage) of the amplifier chain. The 2-stack amplifiers also have shunt-inductor tuning at the first drain node for impedance

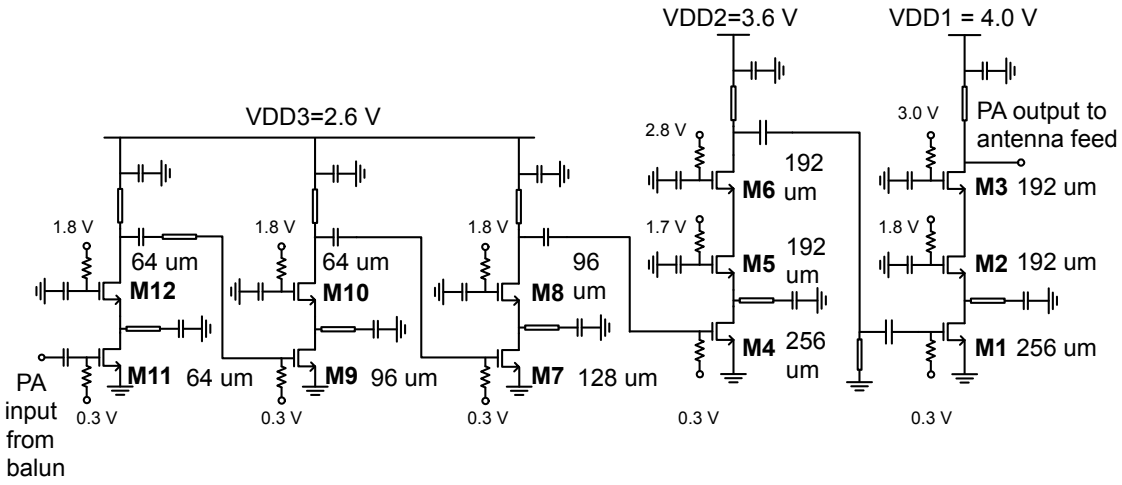

Figure 2.6: Schematics of final five-stage PA (one half of pseudo-differential PA shown).

matching.

The five-stage amplifier chain is used in pseudo-differential fashion to implement one channel of the array. The input to the channel is single ended and a simple balun is used to convert it to differential. The balun has a simulated insertion loss of 1.5 dB and output gain imbalance of less than 0.25 dB and output phase imbalance of 5^0. The first driver stage of the two differential sides are laid out very close to each other and are fed the differential output from the balun. The top FET gates are provided both capacitors to ground as well as capacitor between the two differential sides utilizing virtual ground. The output of this stage is taken as a pseudo-differential signal and passed through the rest four stages. This helps to reduce the imbalance between the differential sides. The rest four stages of amplifier are laid out at a differential pitch which matches the antenna feed pitch so that final stage is close to the antenna feed.

The five stage pseudo-differential amplifier has a simulated linear gain of 32 dB. Typical values of on-chip isolation at W-band are roughly about 40 dB depending on the substrate type and the design of electromagnetic structures like

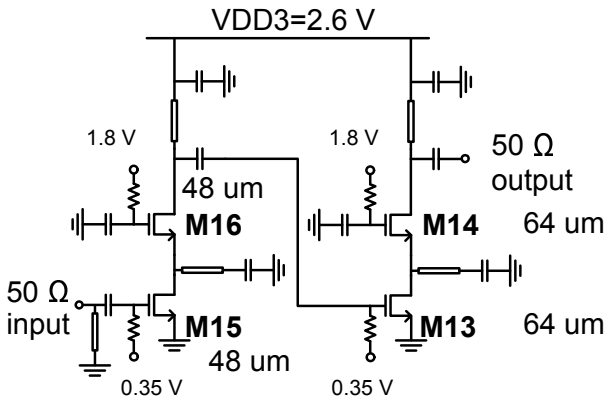

Figure 2.7: Schematics of two stage Line Amplifier.

inductors, antenna, ground plane etc [32]. Thus the five-stage PA provides high enough gain without having the risk of oscillation due to unintentional coupling.

Line Amplifier

The line amplifiers are used to compensate the routing and power division losses in the input signal distribution chain. It is designed as a two stage cascode amplifier with simulated gain of more than 10 dB (Fig. 2.7). For comparison with the main stage amplifiers, the top gate capacitor (CG_2) values are 9.3 fF/μm and 7 fF/μm for the first and second stages of the line amplifier which makes it almost cascode-like. The line amplifier is matched to 50 Ω at input and output and hence can be used as an easily reusable element across the signal distribution chain which is designed in a 50 Ω environment. The final five stage amplifier chain needs only less than 5 dBm input power for achieving saturated output power. But the line amplifier is designed to have a 1-dB output compression power of much more than 5 dB to have an error margin for variations. This helps to have the line amplifiers always operate in deep back-off and hence be very linear.

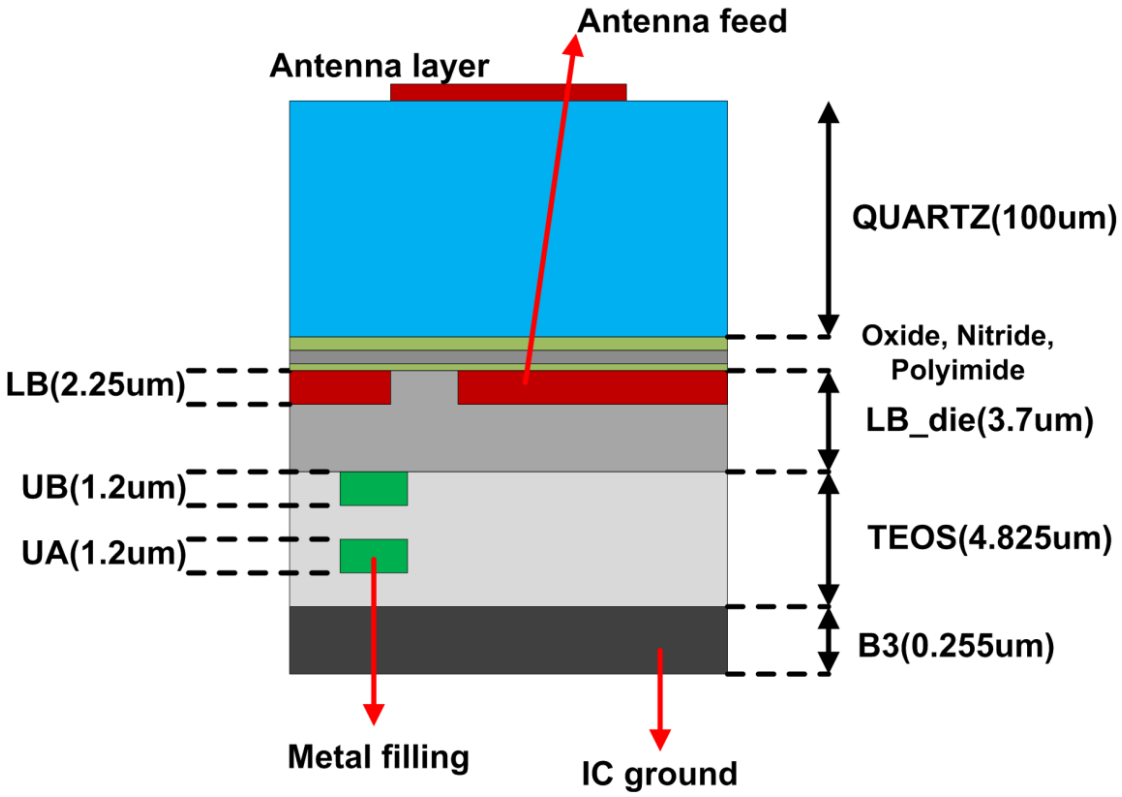

Figure 2.8: Vertical cross section of CMOS + quartz combination.

Signal Distribution Network

A three stage cascade of 2-way Wilkinson power divider is used to achieve the 1:8 split of the input signal on-chip to each of the eight channels. The signal distribution loss includes routing loss, ideal Wilkinson power split and Wilkinson divider loss. 50 Ω grounded coplanar waveguide (GCPW) transmission lines are used for the signal routing. The measured insertion loss of the GCPW is 1.2 dB/mm at 94 GHz. The simulated insertion loss of Wilkinson divider is 1.0 dB at 94 GHz.

2.2.3 Antenna Design

Wafer scale antenna array on a quartz substrate initially demonstrated in [17, 33] is used in this work. A 2 x 4 array of differential microstrip antenna is deposited on top of a quartz substrate and placed on top of the CMOS chip. The antennas on quartz are aligned to the differential antenna feed lines on the CMOS chip at the top most metal layer (LB). The output from the pseudo-differential power amplifier to the antenna feed gets EM coupled to the antenna through the quartz substrate (figs. 2.8 and 2.9). Antenna on quartz rather than antenna on top CMOS metal layer is preferred at W-band frequencies. Using the top metal layer as the antenna layer would cause a very small antenna to ground spacing ($\approx$ 6.3 μm) which does not allow efficient radiation. Also with the external quartz antenna the resonator structure can be changed easily in the case of a shift of peak gain frequency of the transmitter chip. Three 250 nm thick metal layers (B1, B2 and B3) are stitched together to form the antenna ground plane beneath the antenna feed. Since these antennas have very low impedances at the antenna edge the feed line is connected to the amplifier output (50 Ω) by wide quarter-wave impedance transformers.

To satisfy metal density rules dummy metal fill structures are needed to be placed in the layers between the ground and antenna feed. The CMOS process used for this work needs a minimum density requirement of 9% metal fill for the layers between antenna feed and ground plane beneath. The dummy fill reduces the efficiency of the antenna from more than 50% down to about 40%. The placement and configuration of the dummy fill has been optimized for maximum efficiency. Larger dummies increase the eddy current loss and smaller dummies cause more effective capacitive shielding between antenna and ground, thereby reducing the antenna impedance even further. Also dummy elements should be placed carefully

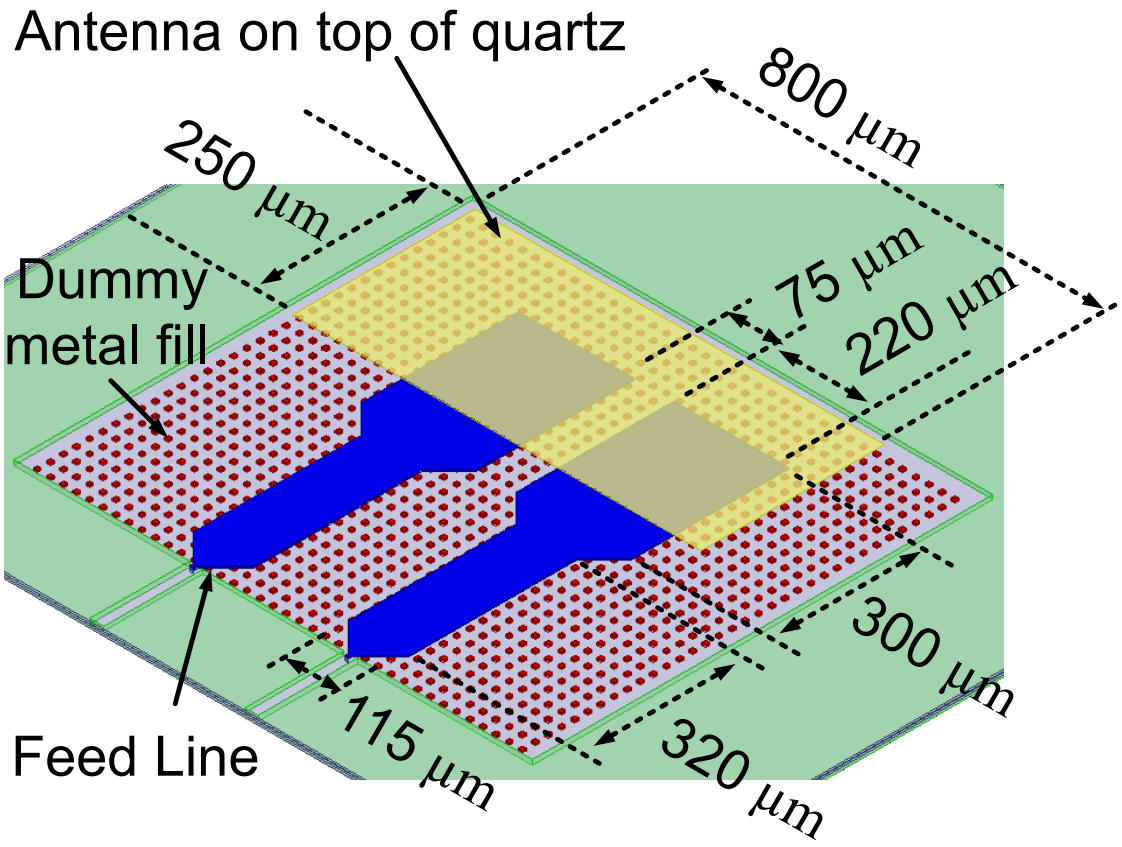

Figure 2.9: Plan of antenna feed on CMOS and antenna on quartz (Designed by Ozan Gurbuz).

away from beneath high current density location on antenna feed like bent edges. In this design, the dummy fills are laid out as floating square pieces of 7.5 μm x 7.5 μm at 25 μm pitch, not tied to each other or the ground beneath.

The antenna is a thin gold layer of dimensions 800 μm x 250 μm and is fabricated on a 100 μm thick quartz wafer. The antenna is designed to resonate at 94 GHz and has a peak gain of 1.0 dB and peak efficiency of 41% at 94 GHz. The array elements are placed at nearly $\frac{\lambda}{2}$ in air spacing for minimum grating lobes. The final implementation is at 0.42 λ spacing in E plane (4 columns) and 0.59 λ spacing in H plane (2 rows). The antenna array simulations show a directivity of 13 dB and gain of 9 dB (Fig. 2.10). The antenna has been designed by Ozan D. Gurbuz.

2.3 Experimental Results

The CMOS PA - Antenna array chip as wells as each sub-block breakouts are fabricated and measured.

2.3.1 Breakout Measurements

Separate breakout test structures are made for the final stage three-stack PA, final four-stage PA, one channel (five-stage pseudo differential PA), line amplifier (two-stage cascode), balun and Wilkinson divider.

Final Stage

The final stage PA (PA1) has a measured maximum small signal gain (S_{21}) of 8 dB at 90 GHz (Fig. 2.11). The 3-dB small signal gain bandwidth of the PA is 18 GHz (81 - 99 GHz). It has a measured maximum saturated output power of

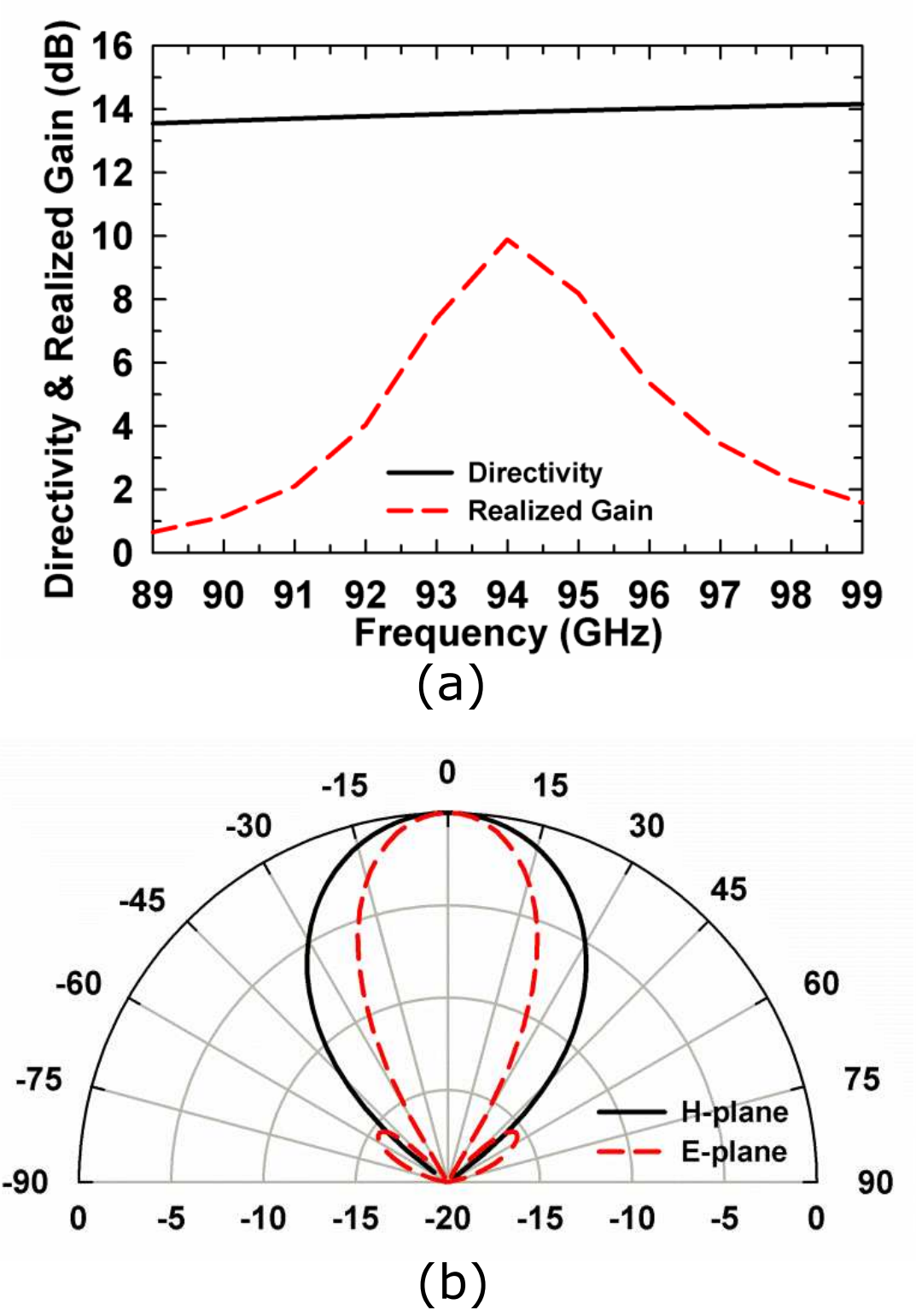

Figure 2.10: a) Simulated 2 x 4 antenna array gain and directivity; b) Simulated 2 x 4 antenna array radiation pattern in E- and H- plane.

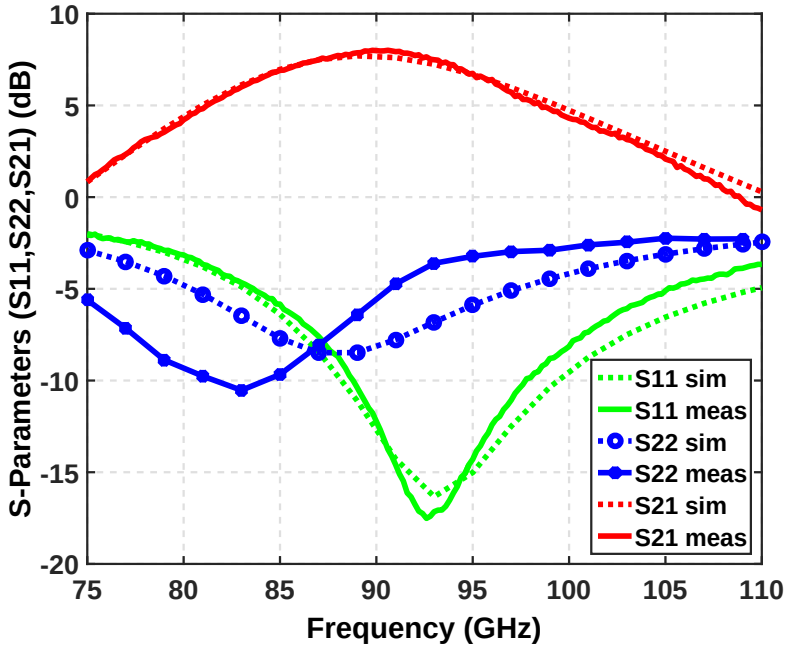

Figure 2.11: Measured and simulated S-Parameters of final stage 3-stack PA (PA1).

17 dBm, maximum drain efficiency of 14% and PAE of 9% at 89 GHz (Fig. 2.12). At 94 GHz the PA achieves S_{21} of 6.5 dB, P_{sat} of 15 dBm, DE of 10% and PAE of 5% (Fig. 2.13). The PA is biased in class-AB mode with 0.15 mA/μm of current density (I_{Dq} = 40 mA, V_{G1} = 0.3 V, V_{DD} = 4.2 V).

The measured output power at 94 GHz is 1.2 dB lower than simulation. The nominal maximum supply voltage for the CMOS process used is 1.1 V. The PA have been designed and measured with 1.4 V per stacked FET. The devices still operate reliably since the stress time per cycle is very low due to the very high frequency of operation. The PA is designed such that at all points of operation the drain-to-gate or drain-to-source instantaneous RF voltage swings of all FETs are less than 2.5 V.

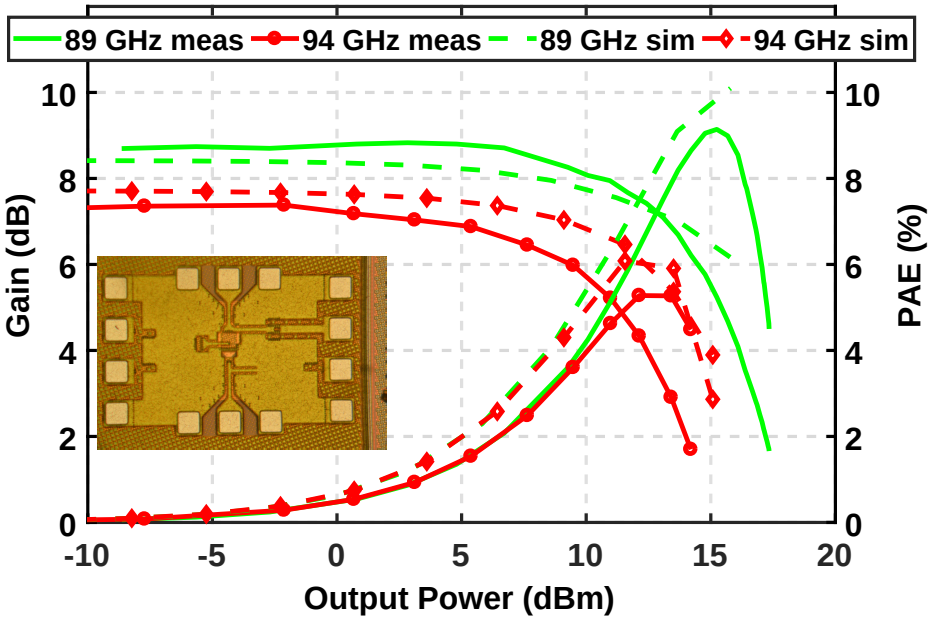

Figure 2.12: Measured and simulated Gain and PAE vs. Output Power at 89 GHz and 94 GHz of final stage 3-stack PA (PA1).

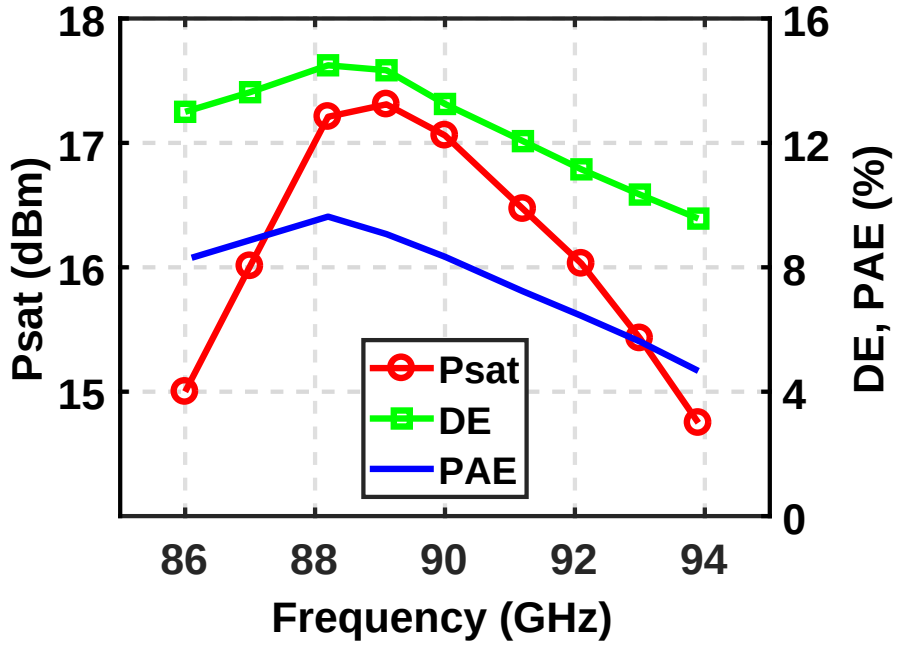

Figure 2.13: Measured maximum output power, PAE and DE vs. Frequency for the final stage 3-stack PA (PA1).

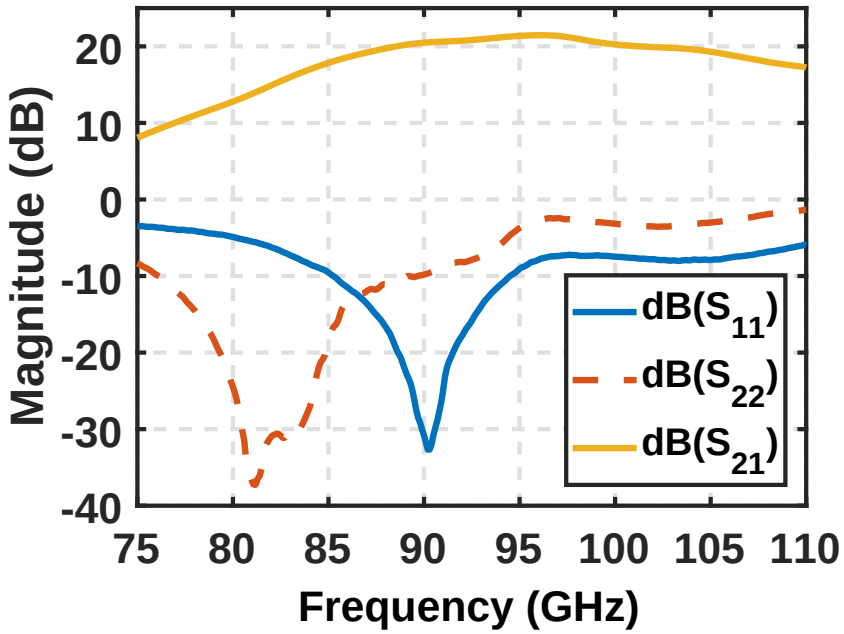

Figure 2.14: Measured S-Parameters for 4-stage PA (PA2).

One Channel (Final + driver stages)

Two breakouts of the multistage PA are measured a single ended four-stage (PA2) and pseudo differential five-stage with balun (PA3).

The four-stage PA (PA2) has a measured maximum S_{21} of 22 dB at 96 GHz (Fig. 2.14). At 94 GHz the maximum measured output power is 15 dBm and PAE is 5% (same as of PA1) (Fig. 2.15). The final stage 3-stack PA is biased at 4.2 V and the pre-final 3-stack uses only 3.9 V supply. The 2-stack driver stages use 2.6 V supply. The measured small signal gain is 2 dB lower than simulation. The drivers are biased in class-A mode for linearity.

The pseudo differential five-stage PA (PA3) has single ended input to the balun and differential output from the final stage of amplifier chain. For measurements a WR-10 GSGGSG waveguide probe is used with one differential side terminated with a 50 Ω waveguide termination and measurements made on the

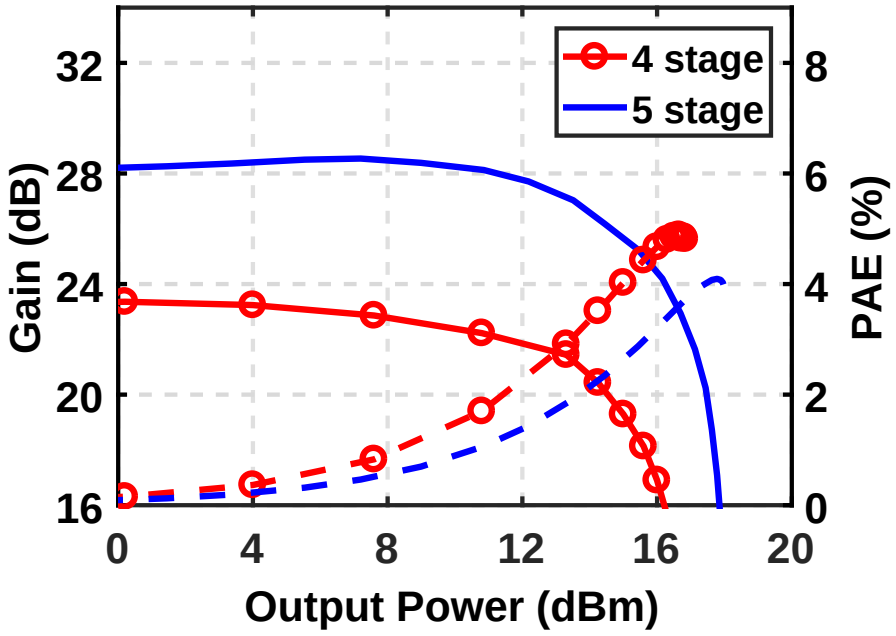

Figure 2.15: Measured Gain and PAE of 4-stage single ended (PA2) and 5-stage pseudo differential (PA3) PA at 94 GHz.

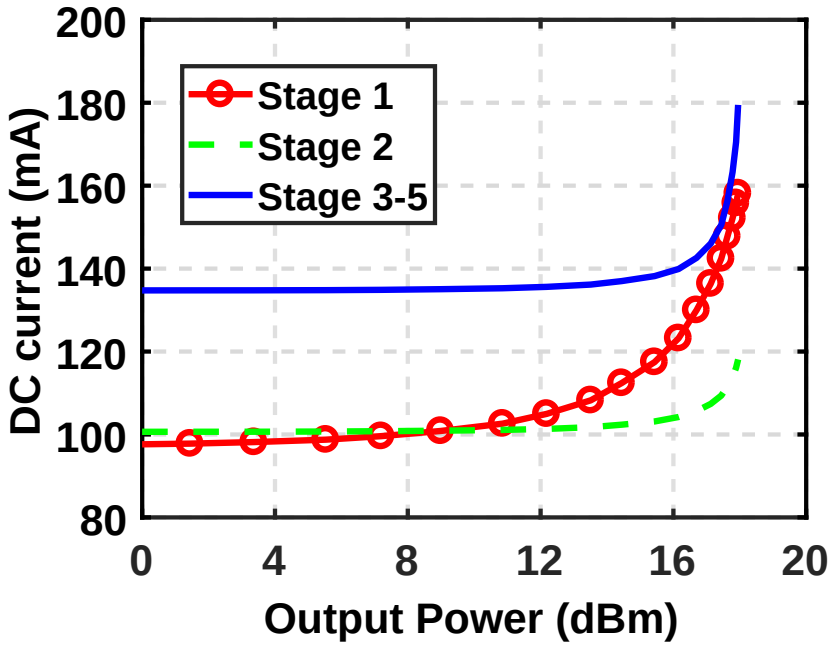

Figure 2.16: Measured drain current of final stage (3-stack), pre-final stage (3-stack) and driver stages (2-stack, three stages vs. output power at 94 GHz for five-stage pseudo differential PA (PA3).

other differential side. At 94 GHz PA3 has a measured small signal gain of 28 dB, maximum output power of 17.5 dBm (14.5 dBm in single ended measurement) and a maximum PAE of 4%. PA3 measured output power is lower than that of PA1 by 0.5 dB and PAE is lower by 1%. Fig. 2.16 shows the measured drain current of final 3-stack, pre-final 3-stack and three 2-stack stages (of two differential sides together). As seen from the plot the pre-final stage current increases from 50 mA only to 60 mA per PA whereas the driver stage currents keep on increasing. Hence it can be inferred that the lower saturated output power compared to four-stage is due to the first stage driver (following the balun) saturating prematurely. PA3 has a 1-dB P_{sat} bandwidth of 14 GHz (86 GHz - 100 GHz) (Fig. 2.17). The measured output power difference between the two differential sides was below measurement accuracy. Since the final stage amplifier maximum gain is centered around 90 GHz the driving stages were designed to have maximum gain at 96 GHz so that together the PA chain has broadband gain centered at 94 GHz. But the saturated output power is still maximum at 90 GHz as it depends only on the final stage.

Line Amplifier

The line amplifier (two-stage cascode) (PA4) has a measured small signal gain of 11 dB and P_{sat} of 11 dBm at 94 GHz. The S-parameters shows very good input and output matching, better than -15 dB at 94 GHz (Fig. 2.18). PA4 has a 3-dB small signal gain bandwidth of 22 GHz (88 GHz - 110 GHz). The PA is linear till 6 dBm output power as seen from the Gain and drain current measurements (Fig. 2.19). This is much higher than the input power needed to saturate the five-stage amplifier chain (0 dBm). PA4 is biased with 2.7 V drain supply.

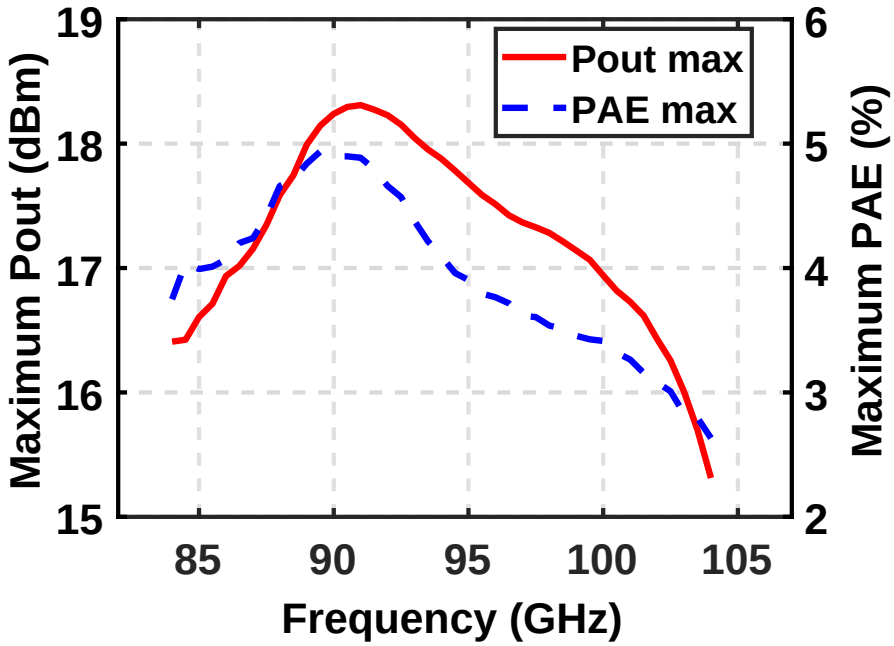

Figure 2.17: Measured maximum output power and PAE of the five-stage pseudo differential PA (PA3).

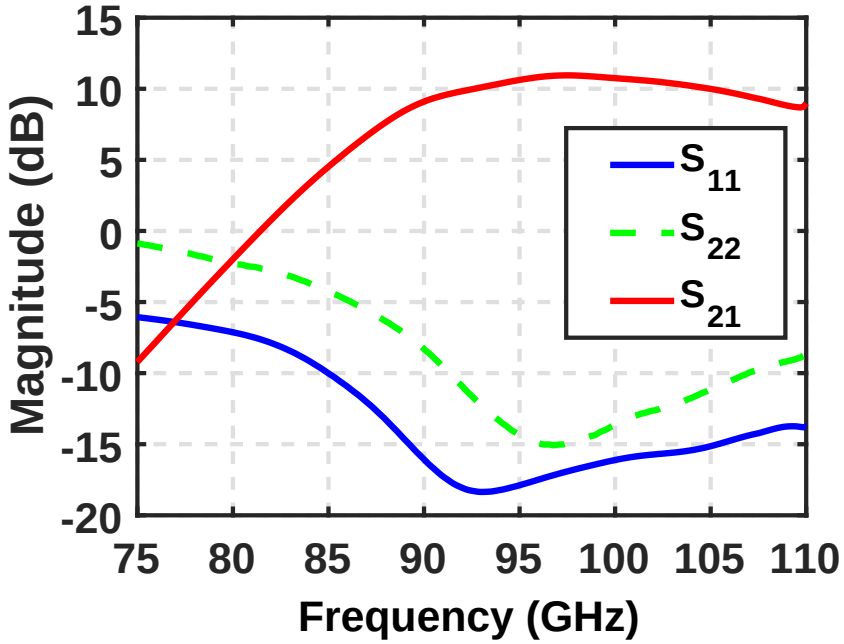

Figure 2.18: Measured S-Parameters of two-stage Line Amplifier (PA4) (V_{g1} = 0.4 V, V_{DD} = 2.6 V).

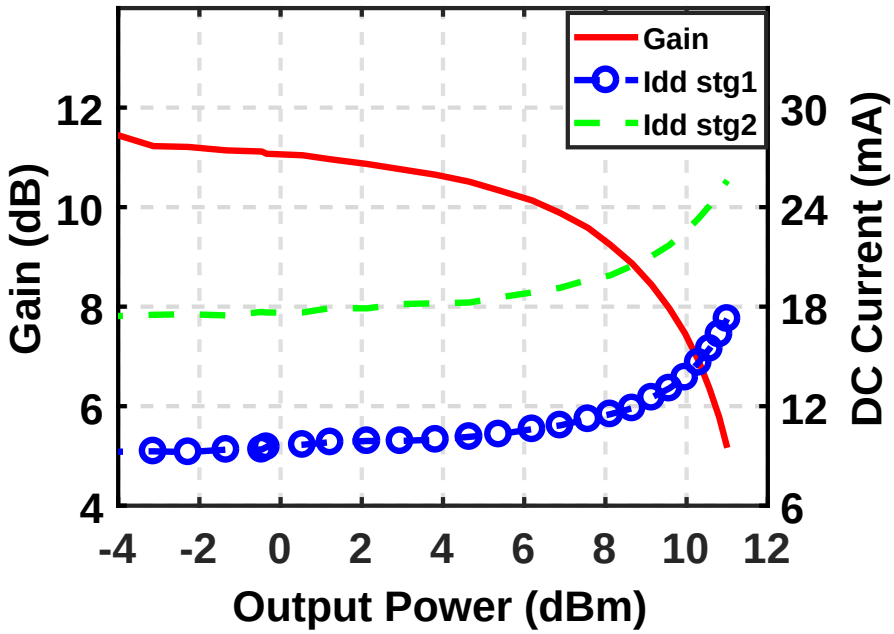

Figure 2.19: Measured Gain and drain current of two stages of two-stage Line Amplifier (PA4) at 94 GHz.

Wilkinson Divider

A test structure of back to back Wilkinson divider is measured and has 1.5 dB insertion loss per Wilkinson divider at 94 GHz. This is 0.5 dB higher than simulation.

2.3.2 Chip Assembly

The PA-Antenna array fabricated in 45 nm CMOS SOI process occupies 4.3 mm x 6.0 mm area on-chip. The chip is mounted on a one inch thick copper block using silver thermal epoxy for heat sinking. The chip has wirebonds on two sides to a PCB placed around the chip on the copper block (Fig. 2.20). The DC supply for the chip is provided from the PCB through these wirebonds. Redundant pads and bonds are used to reduce the routing resistance of the supply lines. Different value

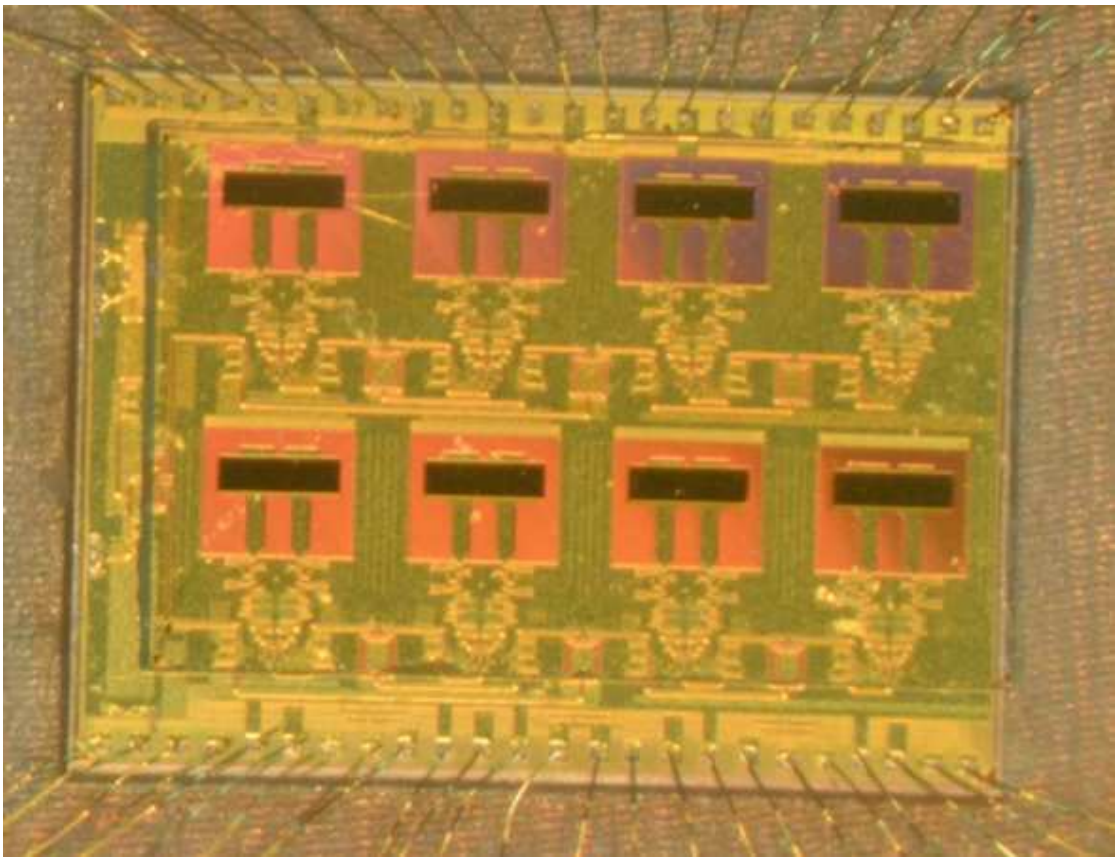

Figure 2.20: CMOS PA chip-quartz antenna wafer assembly wire-bonded to PCB mounted on copper block.

bypass capacitors are placed on the PCB for providing supply bypass at different frequencies. The quartz wafer is placed on the top of the chip and hand-aligned to the antenna feed on the chip. A 10-13 GHz signal from Agilent E8257D signal generator is multiplied and amplified using a x8 VDI multiplier chain to create the W-band signal. The W-band signal is provided to the chip by a WR10 waveguide probe to GSG pad on the chip.

2.3.3 Radiation Measurements

The radiation measurements are done with a WR10 receiver horn antenna placed 25 cm above the chip (Fig. 2.21). This distance is chosen to be much

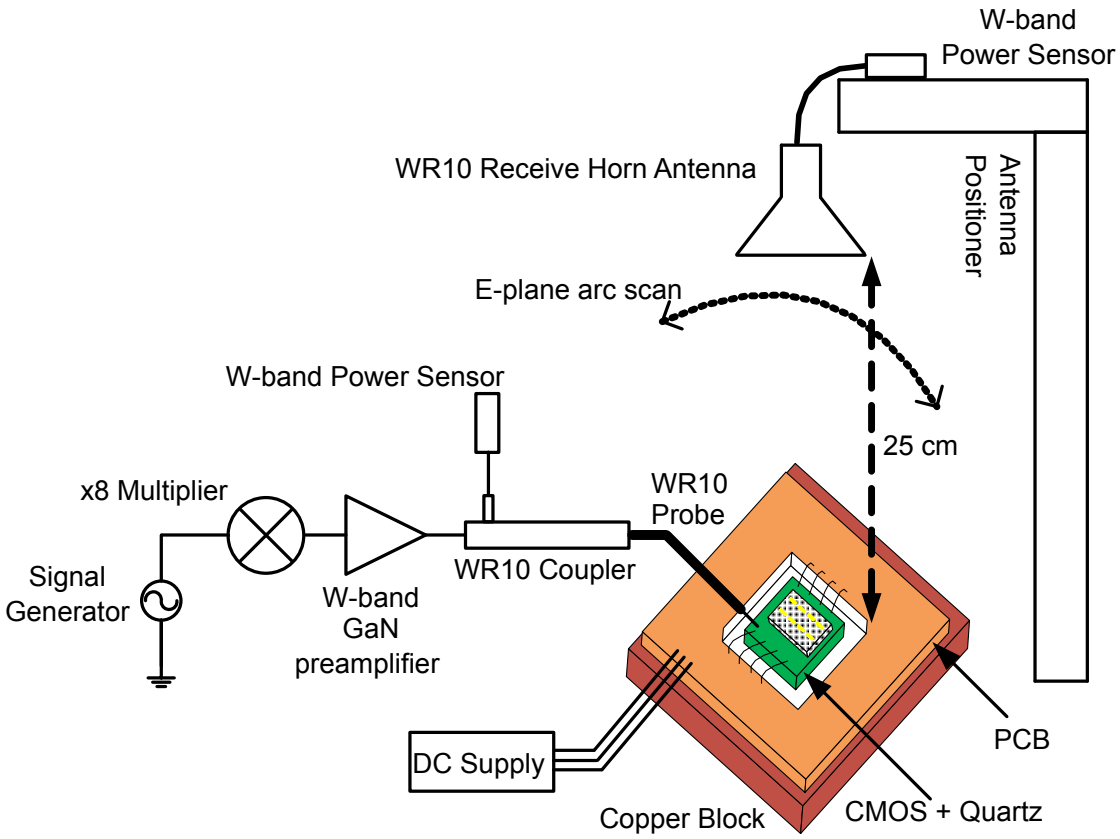

Figure 2.21: Radiation measurement setup.

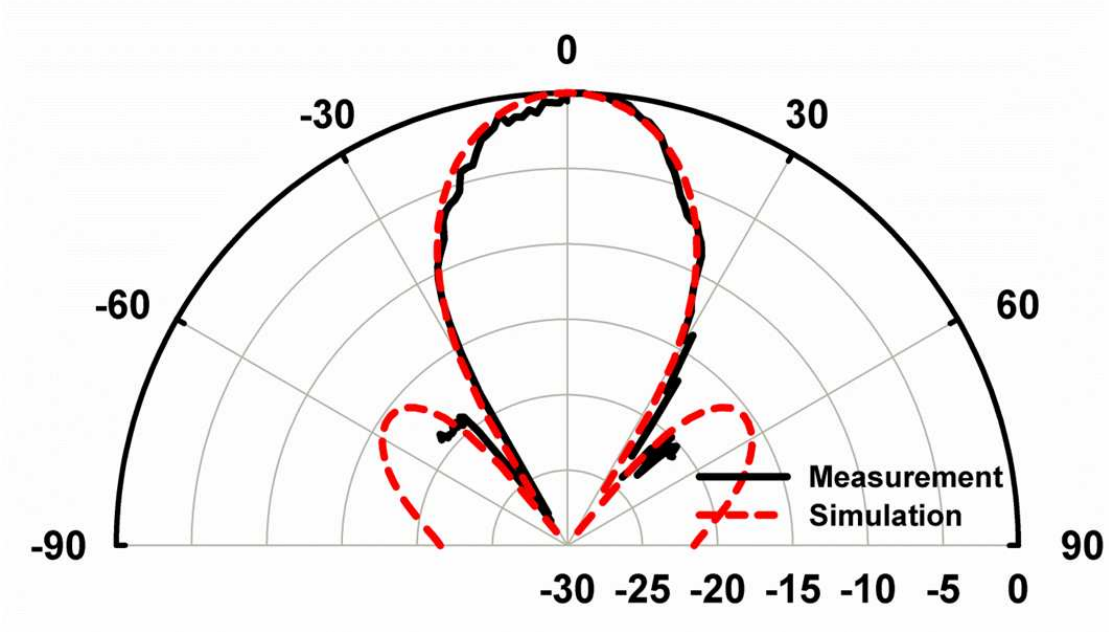

Figure 2.22: Measured and simulated E-plane radiation pattern at 94 GHz.

greater than the far-field distance for the array ($\approx$ 2 cm) and still have enough dynamic range at receiver W-band sensor. The receive antenna is moved in an arc above the chip to measure the radiation pattern in the E-plane. The measured radiation pattern matches excellently with the simulated pattern (Fig. 2.22). The half-beam-width is 27^0 and the side-lobes are at +/- 38^0 in E-plane at 94 GHz. Since the measured radiation pattern is centered and matching with the simulated pattern we can infer that all the eight channels are active and there is no significant phase or magnitude imbalance between them.

The chip-quartz assembly achieved a maximum measured EIRP of 33 dBm at 94 GHz (Fig. 2.23). The antenna gain (G_{TX}) is estimated to be 9.0 dB at 94 GHz from EM simulation using HFSS. Therefore the output power (P_T) from the power amplifier array into the antenna grid can be estimated to be 24 dBm (250 mW). This implies each channel is outputting 15 dBm, which is 2 dB lower than PA3 measurement. More than 9 W of DC power is dissipated leading to

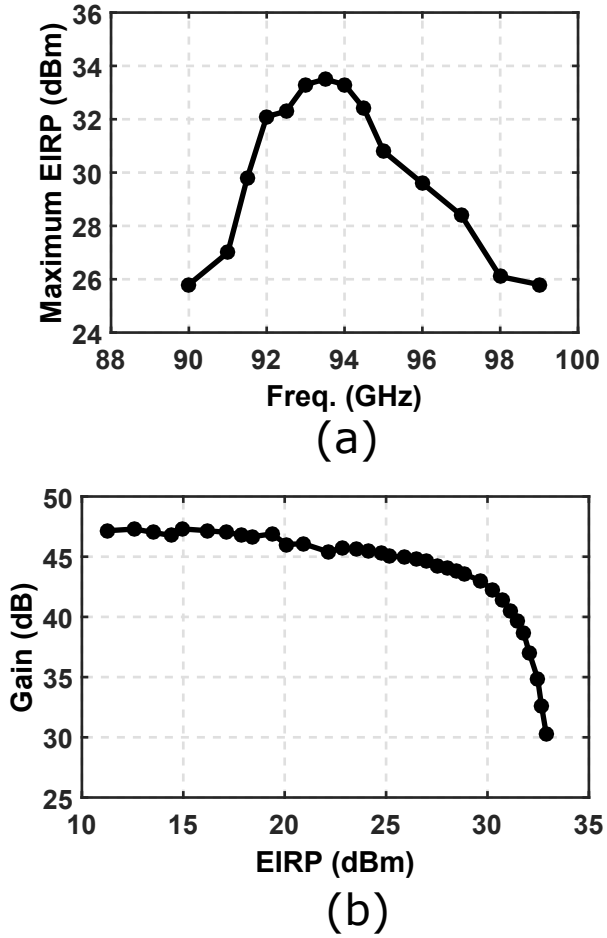

Figure 2.23: (a) Measured maximum EIRP vs. frequency. (b) Measured quasi-optic gain vs. EIRP at 94 GHz.

a 2% total system efficiency considering the on-chip PA power output and 40% antenna efficiency. The efficiency is 1% lower than the array estimate from PA3 and PA4 individual measurements. The performance degradation could be due to compounded effect of increased cross heating between PA, change in properties of CPW between breakouts with air on top and PA-Antenna array chip with quartz on top, finite air gap between chip and quartz causing impedance mismatch at final stage output, increased drain supply line resistance due to longer routing using thin metal layers etc.

The chip was continuously operated at peak power (EIRP of 33 dBm) for

many hours and little irreversible power drop was noticed. The bandwidth of the total system is primarily determined by the antenna. The PA-antenna array achieved an EIRP of $>$ 30 dBm from 91 to 95 GHz (Fig 2.23). Since the power amplifier chain has broad bandwidth, the center frequency of the PA-antenna array can be shifted by using a different quartz antenna designed for the appropriate frequency. This allows for post-fabrication tuning compared to entirely on-chip antenna solutions. The quasi-optic power gain is defined as the ratio of EIRP to the input signal power to the chip. The measurement shows more than 47 dB of linear gain at 94 GHz (Fig 2.23).

2.3.4 PA-Antenna Array Modulation Measurements

Modulated signal measurements are conducted using the spatial power combining chip as shown in Fig 2.24. The baseband signal is generated using Arbitrary Waveform Generator (AWG) Keysight M8190A. The IF signal at 2 GHz from AWG with 6 GS/s is up converted in two steps using 21 GHz and 75 GHz LO to 94 GHz and fed to the PA-Antenna array. The amplified-radiated signal from the chip is captured with a WR10 horn antenna. This is down converted to 2 GHz and sampled using a 20 GS/s Oscilloscope.

Computationally efficient Digital Predistortion algorithm is used to linearize the transmitter. The transmitter non-ideality is decomposed into non-linearity and memory. DPD linearization is applied first (Memory Polynomial DPD with harmonic order of 5 and memory depth of 7) and an FIR filter (memory depth of 100) is applied later to correct the memory effect due to the wide bandwidth. This approach decouples the memory and non-linearity and reduces the number of DPD coefficients from 500 to 135.

256 QAM single carrier signal with PAPR of 7.5 dB is used. The PA is

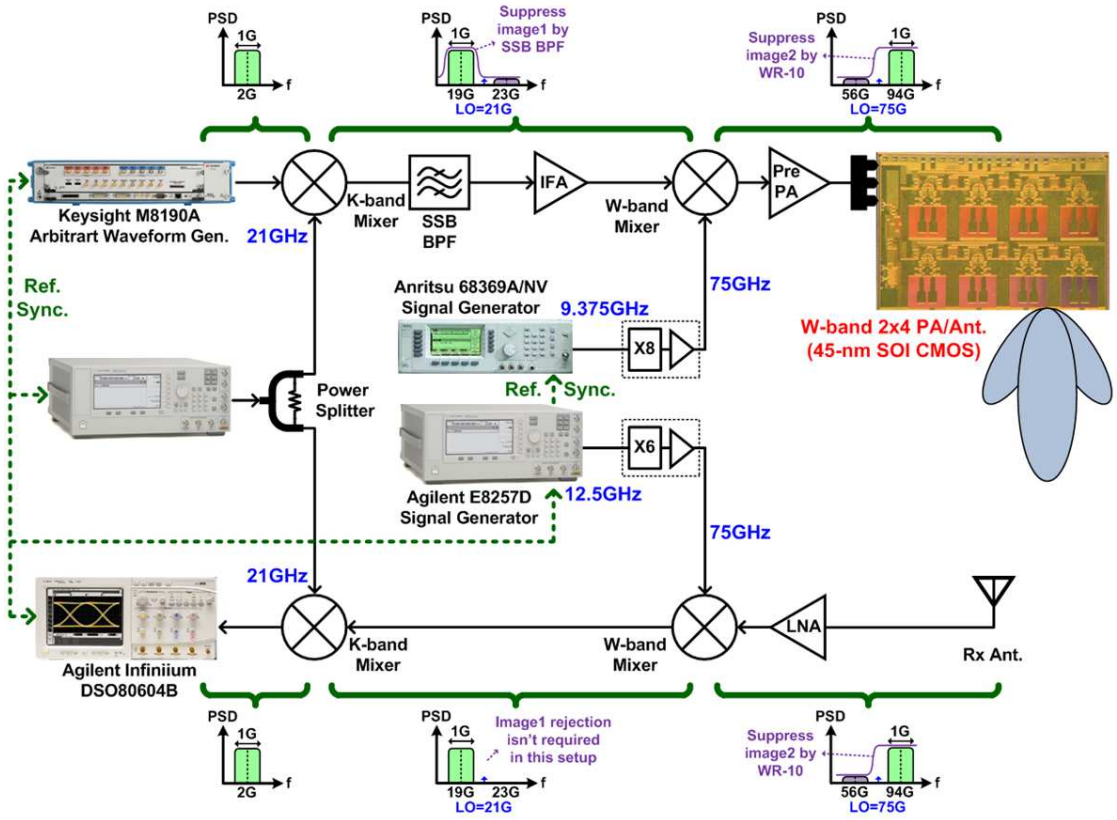

Figure 2.24: Over-the-air modulation measurement setup (Measurements jointly done with Po-Yi Wu).

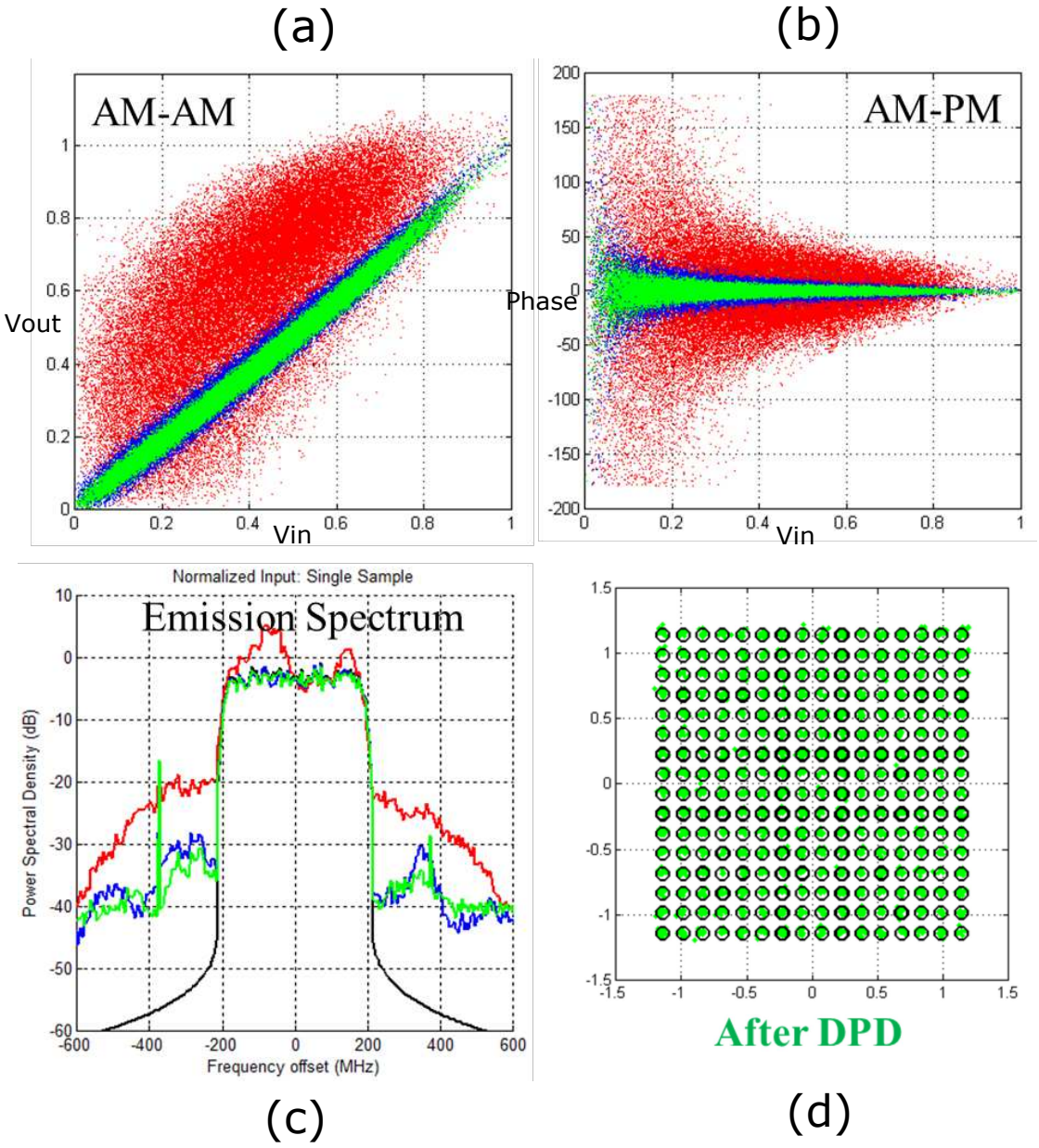

Figure 2.25: Over-the-air radiation measurements of modulated signals for PA-Antenna array with 256 QAM, 375 MS/s single carrier signal (a) AM-AM (b) AM-PM (c) Emission Spectrum (d) Transmitted constellation. (Red is before DPD, blue is after linearization and green is after FIR filtering) (Measurements jointly done with Po-Yi Wu).

biased at low gain to avoid heating. This reduces the maximum EIRP to 26 dBm. Due to the soft roll-off of CMOS PA an additional 5 dB back-off is applied on top of the maximum EIRP PAPR. The PA is operated with an average EIRP of 13 dBm. With 375 MS/s (3 Gbps) the PA-Antenna array after DPD achieves 2.5% EVM and -32 dBc ACPR (Fig. 2.25). The maximum modulation bandwidth is limited by the band pass filter (1 GHz at 19 GHz) in the up convert chain. The modulation measurements were done jointly with Po-Yi Wu and Youjiang Liu.

2.4 Conclusion

A 3-stack CMOS SOI power amplifier is designed to have an output power of 17 dBm at W-band. Using this amplifier as front end, a spatially power-combined, a stacked-FET, CMOS-SOI power amplifier array with integrated quartz microstrip antenna has been designed and tested. The system achieves a maximum EIRP of 33 dBm at 94 GHz. An estimated 24 dBm (250 mW) of power is delivered by the power amplifier grid to the antenna. This is the highest output power from a CMOS chip at 94 GHz. A 3 Gbps over the air modulated signal measurement has been demonstrated with this system using 256 QAM signals with DPD.

Acknowledgment

Chapter 2 is mostly based of materials used in the following publications

The material in preparation to be submitted to *IEEE Transactions on Microwave Theory and Techniques,* J. A. Jayamon, O. D. Gurbuz, P.-Y. Wu, J. F. Buckwalter, G. Rebeiz and P. M. Asbeck, "Spatial Power Combined W-band Power Amplifier using Stacked CMOS SOI with 33 dBm EIRP and

3 Gbps with 256 QAM Modulation". The dissertation author was the primary investigator and author of this material, and co-authors have approved the use of the material for this dissertation. The dissertation author was responsible for the design of the integrated CMOS chip, the overall system assembly and measurements. The author is grateful to Ozan Gurbuz for the design of the antenna and Po-Yi Wu for the modulation measurements.

The material as it appears in 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," in *2013 IEEE 13th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems (SiRF),* Jan 2013. The dissertation author was the primary investigator and author of this material, and co-authors have approved the use of the material for this dissertation.

The material as it appears in J. Jayamon, O. Gurbuz, B. Hanafi, A. Agah, J. Buckwalter, G. Rebeiz, and P. Asbeck, "Spatially Power-Combined W-band Power Amplifier Using Stacked CMOS," in *2014 IEEE Radio Frequency Integrated Circuits Symposium,* June 2014. The dissertation author was the primary investigator and author of this material, and co-authors have approved the use of the material for this dissertation.

Chapter 3

Multigate-Cell FET Design

3.1 Introduction

In mm-wave ICs to date, the transistor stacking has been implemented with discrete, relatively large devices (e.g. gate width 200 μm and above) [12,23–25,34], and large external capacitors are placed at the gates (Fig. 3.1). Connections between the distinct transistors, as well as between the transistors and capacitors, must be carefully implemented in order to avoid excess inductance which can disturb matching and create instability [28].

The maximum single ended output power demonstrated at mm-wave frequencies in silicon using FET stacking alone as the power-combining technique is limited as a result of (a) degenerated device performance due to the layout parasitics; (b) thermal issues at high current density high power operation; and (c) modeling inaccuracies. Representative saturated power levels correspond to 21.6 dBm for CMOS [23] at 41 GHz and 23.4 dBm for SiGe [24] at 41 GHz.

In this work, a modified design is presented for stacked FET mm-wave amplifiers based on the use of transistors with multiple gates. This technique

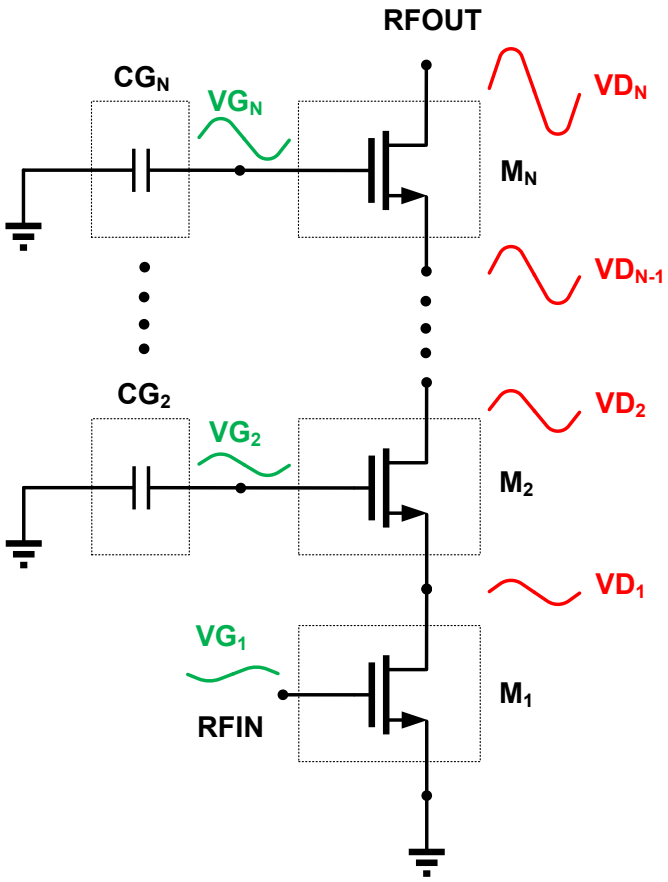

Figure 3.1: Representative schematic of a stacked FET PA with gate capacitors.

eliminates a significant fraction of the parasitic resistances and capacitances of the conventional stacked FET design. The gate connections in the stack can be terminated in capacitors conveniently implemented in the metallization levels available in CMOS technology. The structure leads to a compact unit cell that can be arrayed to provide a wide choice of aggregate device sizes and output currents. Experimental FET stacks and mm-wave amplifiers are demonstrated in CMOS SOI technology. The use of SOI significantly reduces capacitances and leakage, provides isolation between FETs and eliminates FET body-effects. The passive components have higher quality factors due to the insulating buried oxide layer. This makes SOI desirable for mm-wave design, particularly with stacked FETs. SOI, however, can also lead to increased thermal resistance of the FETs to the

heat-sink below the substrate. The multigate design can potentially improve the thermal connection between the FET stack and the silicon substrate relative to what is obtained in the conventional stack.

Amplifiers operating over the 25-35 GHz band (Ka-band) were implemented using the multigate approach. The measured output power is the highest reported to date for CMOS power amplifiers that do not use power combining in Ka-band.

In the following, a description of the multigate-cell design is presented in Section II. Design considerations for stacked FET mm-wave PA, and comparisons of characteristics of conventional FET stacked PA with those of the multigate approach are given in Section III. Section IV presents the design of Ka-band PA using the multigate cell technique, and Section V presents their measured performances. Conclusions are given in Section VI.

3.2 Multigate-Cell Architecture

The multigate-cell merges the source and drain diffusion regions of multiple transistors, leading to a single effective FET with multiple gate regions. This approach reduces parasitic resistances, capacitances and inductances in the amplifiers. There are numerous precedents for the use of multigate FETs in microwave circuits. It is a common practice in pHEMT switch design to have multiple fingers laid out together to be wired as series connected transistors instead of wiring them in parallel [35]. SOI RF switches have also been demonstrated in this layout style [36–38]. A similar layout was used in a previously reported 2-GHz RF CMOS SOI stacked FET design [22]. For high frequency PA, this layout is not favored due to the difficulty of implementing appropriately small gate capacitors for each stacked cell. In typical current practice in a mm-wave FET layout with

N fingers of width W_g each in the same diffusion, all the gates, sources and drains are connected among themselves to make a single FET of width NW_g. Multiple such FETs are used in series with external capacitors at the gate nodes (Fig. 3.1).

In this work, the multigate device includes four gates prepared on the same diffusion or active FET area as shown in Fig. 3.2. The use of SOI technology electrically isolates the bodies of the distinct FETs. Contacts to intermediate source/drain nodes are eliminated. The external gate capacitors needed for the stack are implemented locally using custom designed metal-oxide-metal (MOM) structures around the device. This provides a compact layout for the unit cell containing FET and associated capacitors.

The width of the gate fingers used in the cell is determined in a way that minimizes gate resistance without introducing excess interconnect capacitance. Gate finger widths in the range of 1 to 1.5 μm are considered in this work for Ka-band amplifiers.

The source of the multigate FET is connected to a ground plane placed around the FET on the lowest metal layer (M1). This ring provides a low inductance ground connection as well as a good pathway for thermal dissipation (Fig. 3.3). Since all the FET fingers sit on the same diffusion all of them are now thermally sunk (connected to the ground ring). To further enhance the thermal conductivity, bities (BI Tie-downs) are placed on the ground ring as close to the device as the process limits allow. The bities are polysilicon filled holes punctured through the buried-oxide (BOX) layer to the substrate. This provides a low thermal resistance path to the substrate.

On top of the M1 ground ring a U-shape strip of second interconnect metal layer (M2) is drawn around the FET and connected to the second stack gate. This strip is covered with third interconnect metal layer (M3) which is shorted to the

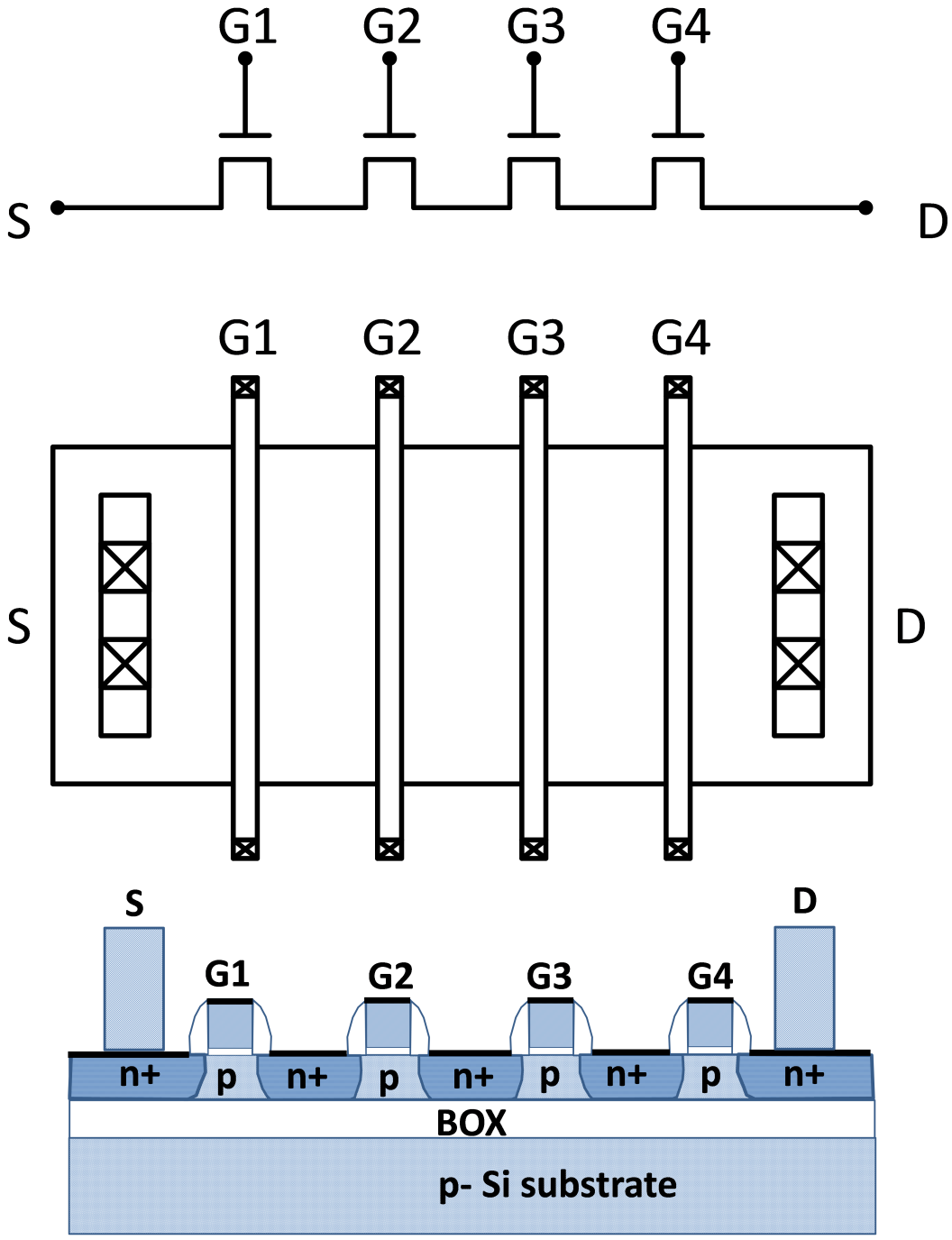

Figure 3.2: Schematic, layout and device cross section of a multigate cell.

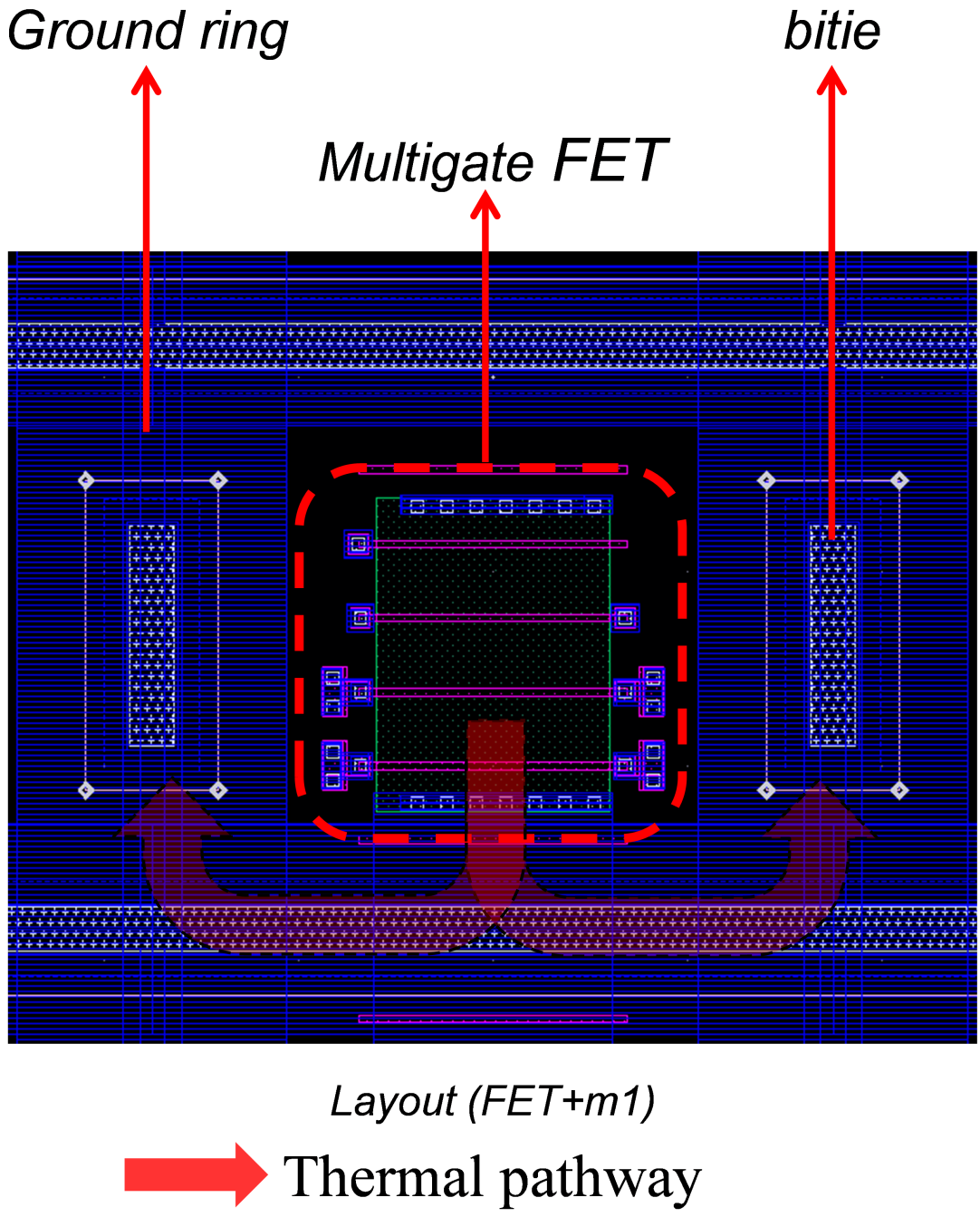

Figure 3.3: Multigate-cell layout with FET and ground ring.

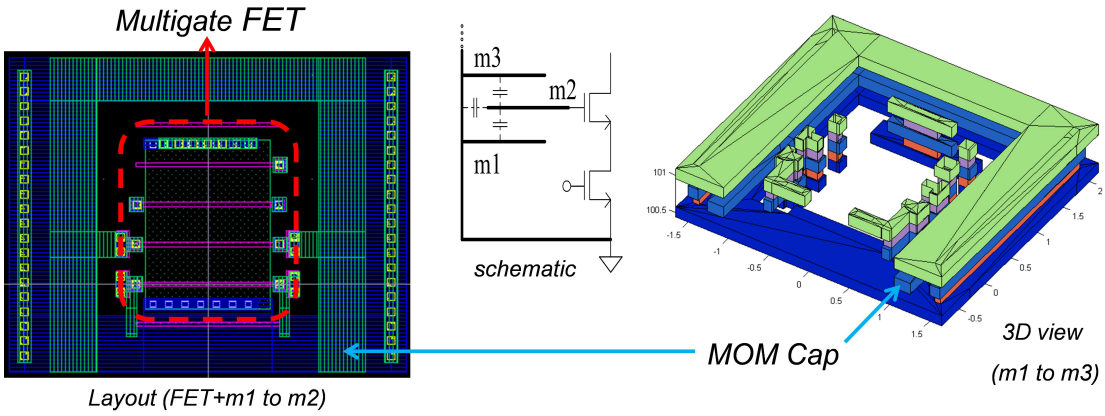

Figure 3.4: Layout of CG2.

ground ring on M1. This forms a Metal-Oxide-Metal (MOM) capacitor between second gate and ground (Fig. 3.4). For higher levels in the interconnect stack, every alternate metal layer is used as capacitor top plate or ground in similar fashion. All the ground layers are tied together with a row of via stacks. All the top-plate layers are connected to the FET gates. The bottom-most eight closely spaced metal layers are used to make the capacitors. This sets the total area density of the capacitors. Since the lower gate capacitors are higher in value the more closely spaced lower metal layers are used for the generation of lower gate capacitors. The multigate-cell capacitance requirements are thus well matched to the capabilities of Silicon back-end-of-line (BEOL) processing.

The maximum current required for the power amplifier, and thus the overall gate width of the aggregate FET assembly determines the number of multigate stacked-cell units needed, which can be arranged in an array (Fig. 3.5).

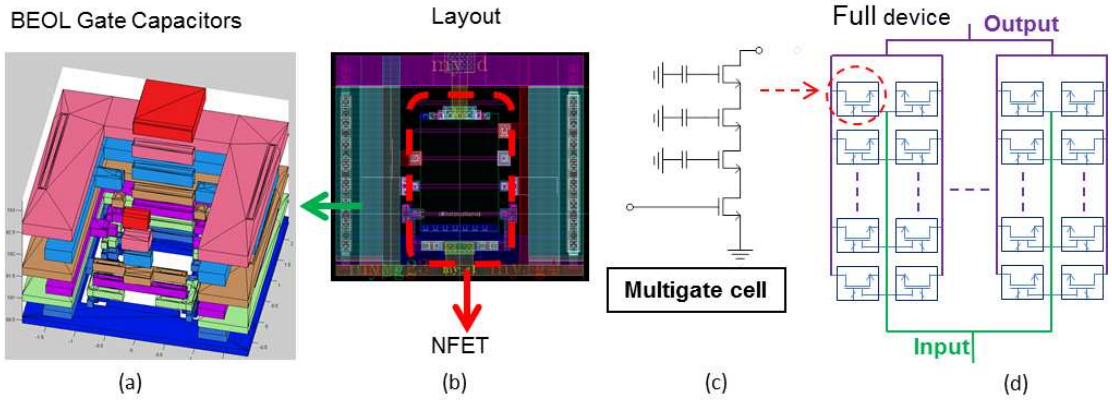

Figure 3.5: Arranging the unit cells to form an array.

3.3 Design Considerations for Multigate Stacked FETs

In the following, the characteristics of the multigate-cell architecture are contrasted with those of the conventional stacking approach in relation to the how they address the significant design challenges for mm-wave PA design.

3.3.1 Wiring Parasitics

Gate length scaling has enabled CMOS devices to have maximum frequency of operation (f_t) as high as 0.5 THz [39]. F_{max} of 430 GHz has been reported [40] for IBM 45-nm CMOS SOI NMOS devices. This result, however, is for a small device (width 8 μm) with port references at the lowest metal layer. As one includes the wiring parasitics for larger devices and considers reference ports at upper layers of metal, the measured value of f_{max} is lower. For a 30 μm device, 285 GHz is reported [41]. For high power mm-wave PA design usually much larger devices (width more than 200 μm) are required so that load lines are closer to 50 Ω. This results in further lowering of f_{max} values to around 200 GHz [26]. The

scaling causes a disproportionate increase in gate resistance (R_g) and capacitance (C_{gs}, C_{gd}) relative to the increase in transconductance (g_m), leading to reduced figures of merit (3.1).

$$f_{max} \approx \frac{1}{4\pi}\sqrt{\frac{g_m}{R_g\, C_{gd}\,(C_{gs}+C_{gd})}} \tag{3.1}$$

A tradeoff exists in the layout of mm-wave FETs between the values of R_g and C_{gg} ($\approx C_{gs}+C_{gd}$). Input gate resistance R_g can be roughly expressed as the sum of three portions (as shown in Fig. 3.6): (a) $R_{g,routing}$, resistance from the global routing on metal layers till the contact via to the gate finger; (b) $R_{g,ext}$, resistance from the gate extension outside the active device; and (c) $R_{g,int}$, intrinsic gate resistance from the active device. $R_{g,routing}$ can be reduced by using multiple metal layers tied together for gate signal routing; $R_{g,ext}$ can be reduced by bringing the contact to gate-poly as close to the active device as possible; $R_{g,int}$ can be reduced by using double side gate contacts [42]. All of these result in increase of parasitic gate capacitances (C_{gsx}, C_{gdx}). In this conventional style layout it becomes difficult to improve f_{max} beyond a value determined by the best achieved $R_g\, C_{gg}$.

For the multigate-cell in this work, a gate-pitch (380 nm) wider than the usual pitch (190 nm minimum allowed by technology) is used. This improves the f_t due to (a) enhanced stress response of the device leading to higher transconductance (g_m); and (b) lower gate-to-contact capacitance due to wider spacing [39]. The contacts to intermediate source/drain nodes are eliminated. This nearly eliminates the extrinsic parasitic capacitance due to wiring on the gate (C_{gsx}, C_{gdx}) for the intermediate FETs in the stack. Since the $R_g - C_{gg}$ trade-off is reduced it is possible to reduce R_g further than in conventional layouts. A double-sided contact gate finger layout is used and the contact via to the gate poly is made at

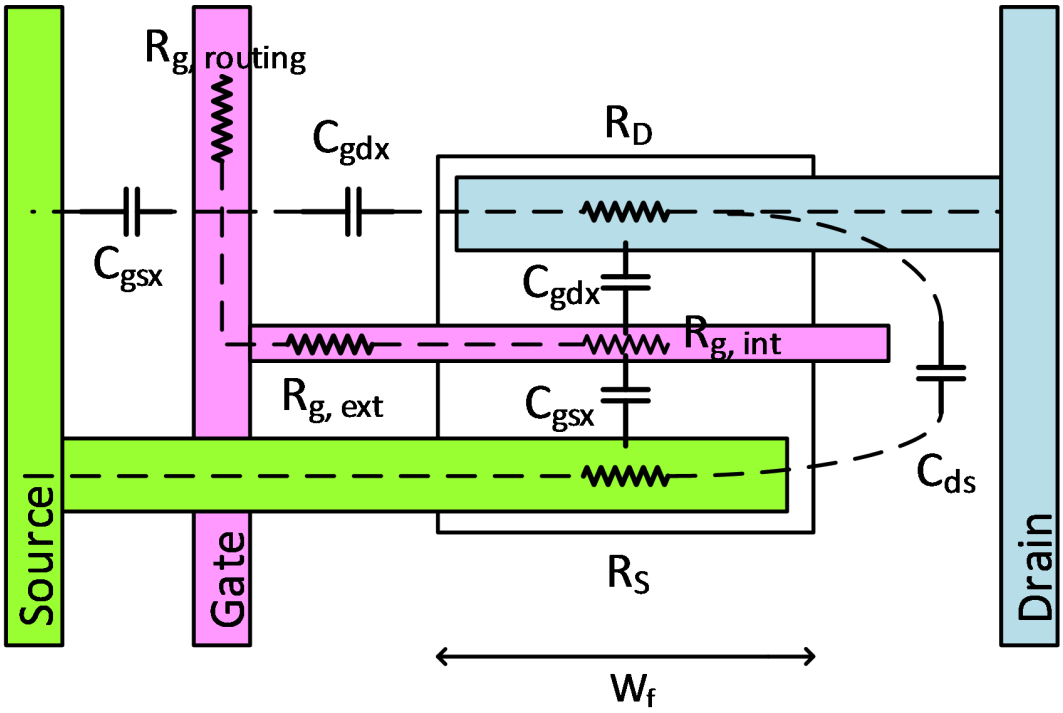

Figure 3.6: Representative FET layout showing parasitic resistances and capacitances on gate.

the minimum allowed distance from the active device.

To compare parasitic capacitances, a conventional FET layout was studied along with a comparable multigate-cell. The 32 μm wide FET (1 μm x 32 fingers) was RC extracted up to the global routing layer (UA). Simulations were done using (a) schematic for intrinsic device; (b) RC extracted FET with wiring and (c) RC extraction of wiring alone without FET. The estimated values for C_{gs} and C_{gd} from the simulation are shown in the Table I. The simulations suggest that by eliminating intermediate source/drain contacts and wiring, the C_{gs} and C_{gd} values can be reduced by 17% and 24% respectively.

Table 3.1: Comparison of Device and Wiring Parasitics

Parameter	**Value from FET schematic simulation**	**Value from RC extracted FET simulation**	**Value from RC extracted wiring simulation**	**Estimated percentage improvement with multigate layout (%)**
C_{gs} (fF / μm)	0.60	0.72	0.12	17%
C_{gd} (fF / μm)	0.25	0.33	0.09	24%
C_{ds} (fF / μm)	0.23	0.29	0.10	21%

3.3.2 Vertical Contact Parasitics

For large FETs used in PA, the source and drain contact resistances (R_D, R_S) are a considerable fraction of R_{ON}. Even with wiring optimized for reducing resistance, the equivalent parasitic series resistance on source/drain for a FET wired to the global routing layer (UA in this example) is approximately 24 $\Omega.\mu m$ as calculated from via contact resistance and routing metal sheet resistance. From this total approximately 12 $\Omega.\mu m$ is from the device to M1 contact and the rest is from the via and metal routing to the top layer. Layouts which attempt to decrease the series resistance result in increase of C_{ds}. The tradeoff is similar to the well-known $R_{ON} - C_{OFF}$ trade-off for switches. The vertical routing also creates series inductance [43]. Particularly for stacked FET designs the added parasitics present a problem as these parasitics are multiplied by the number of FETs in the stack. For a 4-stack FET, we therefore have $8R_v$ and $8L_v$ in series with the stacked-FET as shown in Fig. 3.7. The series resistance reduces the PAE and the series inductance narrows the bandwidth.

Since the current in the stack flows laterally from one finger to another inside the 4-stack multigate FET, six out of the eight vertical interconnects (and their associated resistance and inductance) are eliminated. For the bottom-most

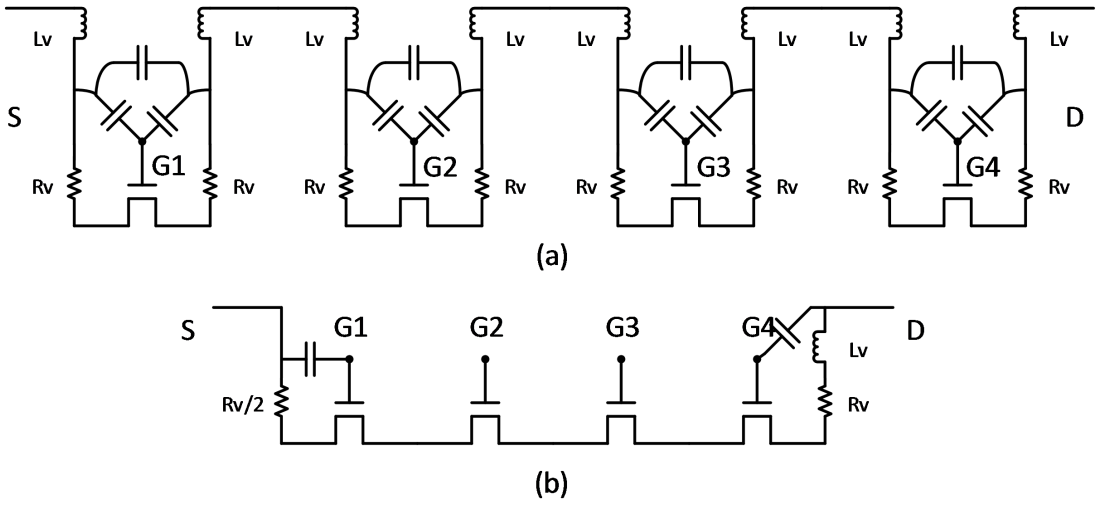

Figure 3.7: Schematic of conventional and multigate 4-stack FET with parasitics associated.

finger, only the resistance of the contact from device to M1 is present since the source is connected to a very wide ground ring. This provides an estimated 82% reduction of the parasitic wiring series resistance of the FET. At the same time nearly all of the C_{ds} arising from the local wiring parasitics is eliminated. It is not possible, however, to completely eliminate parasitic capacitance at the drain of the highest FET on the stack since the multigate-cell array requires global drain wiring to connect unit cells.

A comparative study of series resistance was done between the 4-stack multigate-cell and a 4-stack 32 μm gate width conventional-style FET. The estimated value of R_{ON} is shown in Table 3.2 for different scenarios. The multigate-cell layout has 11% lower R_{ON} compared to the conventional style layout. Assuming that the difference in R_{ON} between the RC extracted FET stack simulation and intrinsic schematic FET stack simulation corresponds to the parasitic series resistance, the multigate-cell is able to reduce the resistance by 82%, which matches the series resistance estimation described above based on contact via resistance

Table 3.2: Comparison of Series Resistance of Stacked FET in Conventional Style and Multigate-cell

Parameter	R_{ON} ($\Omega.\mu m$)	**Estimated percentage reduction of parasitic series resistance with multigate layout (%)**	**Estimated percentage reduction on R_{ON} (%)**
Value from FET schematic simulation	1128	-	-
Value from conventional layout FET 4-stack : RC extracted simulation	1304	-	-
Value from multigate-cell 4-stack FET : RC extracted simulation	1161	82%	11%
Measured value for multigate cell IC	1148	89%	12%

(Measured value of R_{ON} is averaged across multiple size devices)

and routing metal sheet resistance. The difference in C_{ds} between conventional style layout and multigate-cell is shown in Table 3.1. The parasitic C_{ds} due to the wiring of conventional style FET (and hence the savings by multigate-cell) is estimated to be 0.06 $fF/\mu m$, as indicated by the difference between RC extracted and schematic. This is less than the 0.10 $fF/\mu m$ as given by RC extracted simulation of wiring alone; the difference can be attributed to the fact that simulation of wiring alone ignores the reduction of C_{ds} due to shielding by the FET and underlying silicon substrate.

3.3.3 Thermal Resistance and Self Heating

For high power PA design, SOI suffers from increased self-heating. The thermal conductivity of SiO_2 ($\approx 1W/mK$) is two orders of magnitudes lower than Si ($\approx 130W/mK$). The active device is surrounded by shallow-trench-isolation (STI) on all sides, by dielectric layers on the top and by buried-oxide on the bottom. This leads to relatively thermally insulated FETs and can lead to elevated junction temperatures while operating at high current densities [28, 44]. This lowers both performance and reliability of the high power amplifiers. Thinning of the substrate cannot alleviate the situation since the majority of the thermal resistance is due to the buried oxide.

Thermal interaction between transistors is an additional concern. In conventional stacked FET designs, the FETs in the middle of the stack do not have a good thermal dissipation pathway. They are also cross-heated by FETs from both sides.

Multigate-cell design reduces - but does not eliminate - these concerns. Since all the (four) FET fingers sit on the same diffusion, they are thermally connected by the continuous layer of Silicon. Heat can subsequently flow through

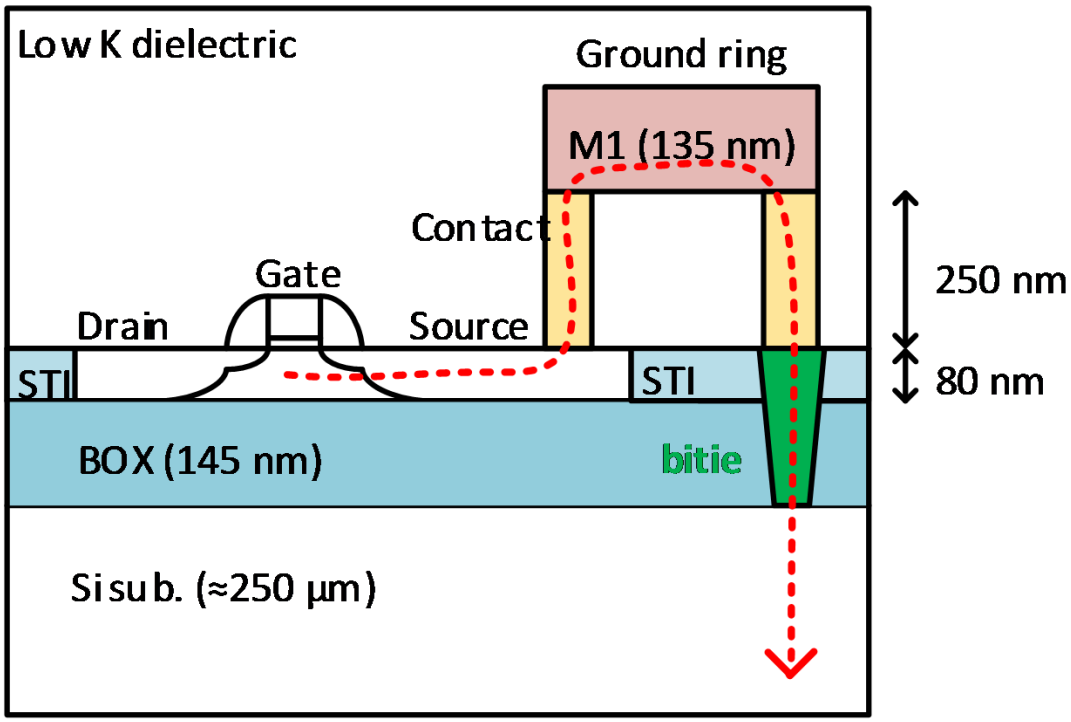

Figure 3.8: Thermal dissipation pathway with bitie.

the metal connections to the source. The source connection to the M1 ground ring thus is instrumental in dissipating the heat across the chip. This is aided by the fact that every alternate interconnect metal layer is connected to the ground plane to create the multigate-BEOL capacitor. A large density of polysilicon filled through-buried-oxide-vias (bities) are placed on the source/ground ring (Fig. 3.8) to improve the thermal conductivity to the substrate. The distributed nature of the multigate-cell FET array allows the designer to put bities very close to and all around each unit cell.

3.3.4 Unequal and Low Quality-Factor Gate Capacitor

Exact sizing of the external gate capacitance is critical in stacked FET design to maintain the impedance levels at internal stack nodes and maintain

equal voltage swing among the different FETs in the stack. Due to the large size of the FET in conventional designs, the series inductance of the connection to the external gate capacitor can be considerably different between gate fingers located at different locations of the FET. At mm-wave frequencies, this leads to having different impedance levels at the gate for different device fingers of the same FET, and hence different voltage swings. The long lead inductances to the capacitors and their ground return path can also make the design narrowband and potentially unstable.

Typically large capacitors have lower quality factor (Q) at mm-wave frequencies caused by wiring parasitic resistance due to spreading resistance and skin depth. This reduces the gain and efficiency of the stacked-FET PA. Active-drive of the stacked-FET gates [41] can reduce this problem, but would be area expensive and can result in very sensitive design.

In multigate-cell, the capacitors are very small and the current return path from gate to source through the capacitor is local, lowering the parasitic resistances and inductances associated with the gate capacitors. The capacitance is equal for all the unit cells. The external gate capacitor CG2 for the multigate-cell is about 3 fF. For a conventional stack design with size equivalent to 192 unit multigate-cells, two gate capacitors of value 384 fF are needed. Electromagnetic simulation of the multigate-cell BEOL capacitor indicates a Q of more than 375 for the 3 fF capacitor at 28 GHz compared to around 100 for the 384 fF vncap (metal finger capacitor) (Fig. 3.9).

3.3.5 Intra-Stack Tuning

It has been shown that to improve the efficiency of mm-wave stacked FET power amplifiers it is desirable to carry out impedance matching at intermediate

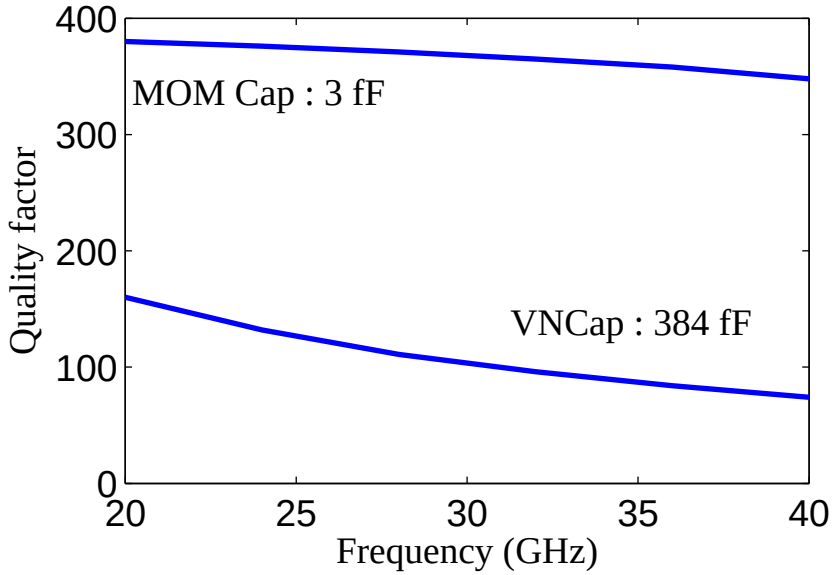

Figure 3.9: Quality factor of gate capacitors.

nodes in the stack in order to compensate for capacitances to ground present on these intermediate nodes (from device parasitics and from intrinsic C_{gs}). This intra-stack tuning is of increasing importance as the frequency of operation is increased. Different types of intra-stack tuning including series-inductor [25], shunt-inductor [45] and C_{ds} shunt-feedback [24] have been used to improve the efficiency of mm-wave stacked power amplifiers by providing impedance match at intra-stack nodes. Due to the compact unit cell and distributed array architecture, these efficiency improvement techniques are difficult to implement for the multigate-cell design. Although the parasitic capacitances are reduced for the multigate-cell design, it is likely that at very high mm-wave frequencies the conventional approach will have superior performance by using intra-stack tuning. Reducing the width of FETs along the stack [25] as well as driving multiple gates on the stack [26] are other techniques reported to adjust for the current leakage to ground due to parasitic capacitances as one moves up in the stack. These are also more difficult to

implement for the multigate-cell design than for the conventional approach. This diminishes the usability of the multigate-cell design presented here for PAs with large FET widths at frequencies higher than 75 GHz.

3.3.6 Array Architecture

The full PA device is laid out as 8 (or 6) x 32 array of unit cells. The 32 columns are fed in a binary corporate tree fashion using thick copper microstrip transmission lines that connect to the gates. A similar binary corporate combining tree is used to on the drain side of the unit cell array. Inside each column the signals to the 8 (or 6) rows of unit cells are routed in series fashion (Fig. 5). Alternate columns share the gate and drain routing, thus reducing the corporate combining tree to be 16-way instead of 32-way. Inside each column the drain routing and gate routing are done on different metal layers to avoid coupling between input and output. As a result the input gate routing transmission line and output drain routing transmission line have different impedances and different velocities of wave propagation.

Let d be the unit cell width, ω be the frequency of operation and v_d and v_g be the respective loaded transmission line propagation velocities on output drain and input gate lines. Then with n unit cells in a row, the maximum phase difference between unit cells at extreme ends can be expressed as

$$\Delta\Phi = n\,\omega\,d\left(\frac{1}{v_d} - \frac{1}{v_g}\right) \tag{3.2}$$

This shows that for a given choice of transmission line structure, the number of unit cells one can have on a row decreases inversely as the frequency increases for a given maximum allowed phase difference between cells. Phase difference affects

gain and impedance and also appears as a dispersion which causes the harmonic frequency contents to be misaligned. Another effect would be unintentional load pulling between the unit cells. One way of achieving phase coherency would be by minimizing the velocity difference between the gate and drain routing lines. In this work, with $n = 8$ and $d = 3.5\mu m$, the overall propagation distance along the row is $28\mu m$, and the degree of dephasing between the lines is estimated to be negligible (less than 1^0).

Unlike bipolar devices (HBT), CMOS FET devices have negative temperature coefficient for I_{Dsat} ($\approx$ -0.6 ($\mu A/\mu m/^0C$) for the 45 nm NFET used for this design). This prevents current hogging and thermal runaway if a temperature difference should arise between different unit cells in the array.

3.4 Multigate-Cell PA Implementation

A four finger/stack multigate cell with finger width $1.2\mu m$ was designed. The FET model was derived using parasitic RC extraction (PEX). An EM modeling tool (EMX) was used to simulate the BEOL capacitor around the FET. The unit cell with the FET and BEOL capacitor occupies an area of $3\mu m \times 3.5\mu m$. To maximize the area density of the BEOL capacitors and thereby minimize the size of the stacked-cell maximum amount of metal allowed by design rules is used. This has the advantage that since the metal densities are maximized, no additional dummy fill [46] is needed. It results in a more accurate simulation of the capacitor structure.

The 4-stack multigate unit cell simulation shows an effective f_{max} of about 240 GHz. A load-pull simulation at 28 GHz with the unit cell yields a saturated output power of 1.5 dBm and 43% PAE. The optimum impedance for maximum

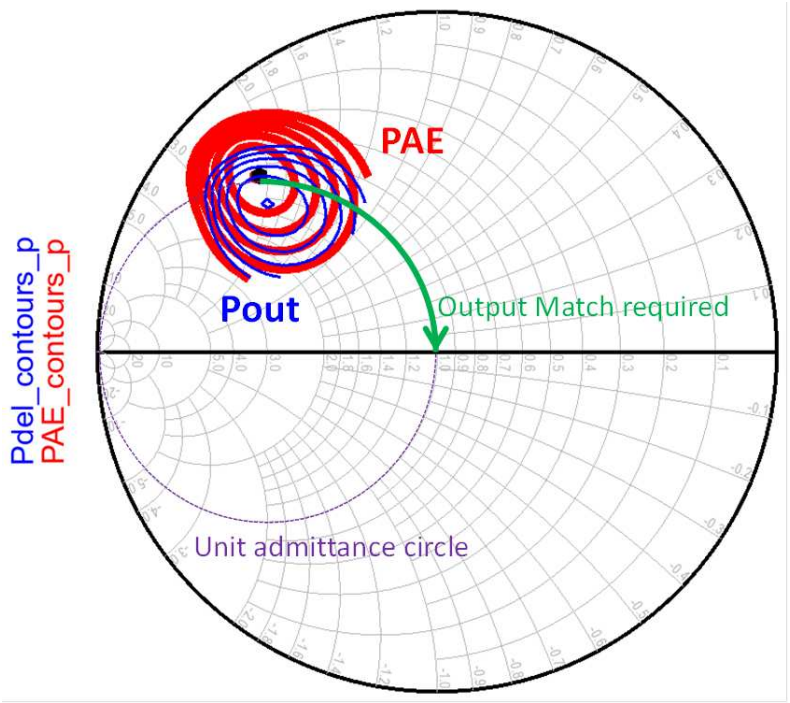

Figure 3.10: Load-pull simulation of unit multigate-cell.

power and efficiency (Z_{opt}) roughly lies in the unit admittance circle when the reference impedance is set to be $192 \times 50\Omega$ (Fig. 3.10 (Reference Impedance = 50 $\Omega\times$ 192; Normalized Y_{opt} = 1 + j 2 S; Max P_{out} = 1.5 dBm; Max PAE = 43%). Thus, an array of 192 unit cells would have Z_{opt} close to 50Ω. This avoids need to have any impedance transformation at the output thereby improving the efficiency and making the design broadband.

A 4-stack FET of aggregate width $230\mu m$ was made using an array of 192 (6 x 32) unit cells. A PA (PA1) was designed at 28 GHz using this device (Fig. 3.11). Simple matching networks with grounded Coplanar Waveguide (CPW) transmission lines and capacitors were used to match input and output. The PA achieves a simulated saturated output power of 24 dBm.

Another PA (PA2) of slightly larger periphery 256 unit cells / 307 μm width, was designed for higher output power (Fig. 3.12). Since the device width

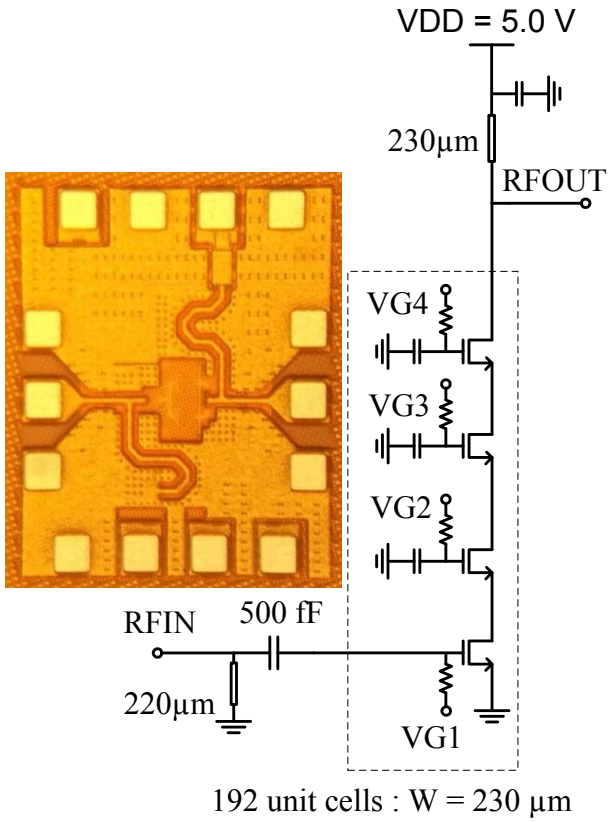

Figure 3.11: 230 μm FET 4-stack PA (PA1). (a) Die microphotograph. (b) Schematic.

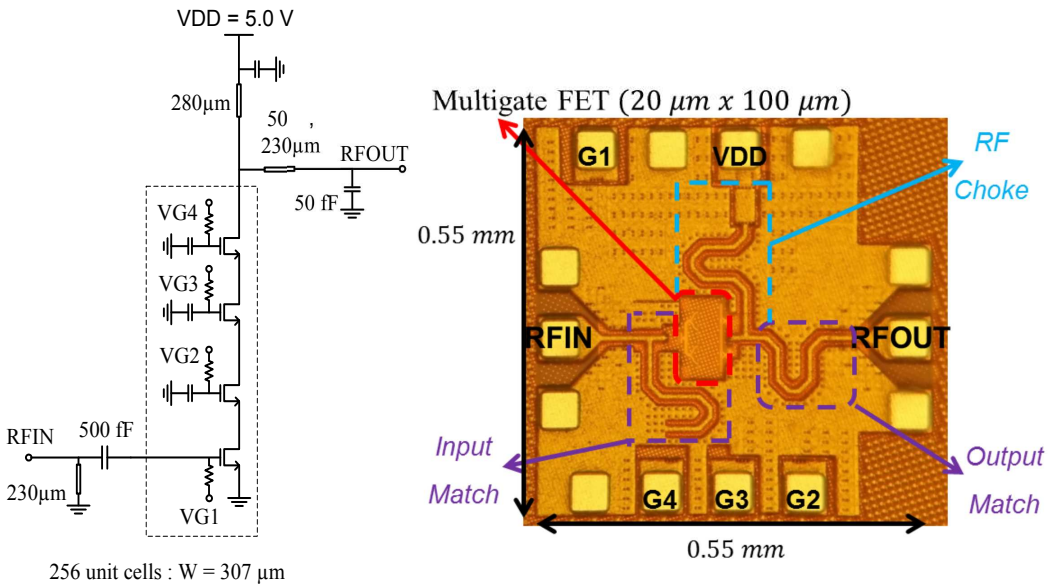

Figure 3.12: 307 μm FET 4-stack PA (PA2). (a) Die microphotograph. (b) Schematic.

increased the optimum load impedance is reduced to 40 Ω. An additional series L shunt C matching network was added to form the impedance transformation from 40 Ω to 50 Ω. The series L is implemented as a transmission line. This matching network also provides a short at the second harmonic at the drain which helps to increase the efficiency. The PA achieves a simulated output power of 25 dBm.

3.5 Measurement Results

The multigate amplifiers (PA1 and PA2) were fabricated in a 45-nm CMOS SOI process (Fig. 3.11, 3.12). These amplifiers occupy an area of 540 μm × 450 μm (0.24 mm^2) for PA1 and 540 μm × 550 μm (0.30 mm^2) for PA2 including the pads. Excluding the pads the PA occupy areas of 340 μm× 190 μm (0.064 mm^2) and 340 μm× 280 μm (0.095 mm^2) respectively.

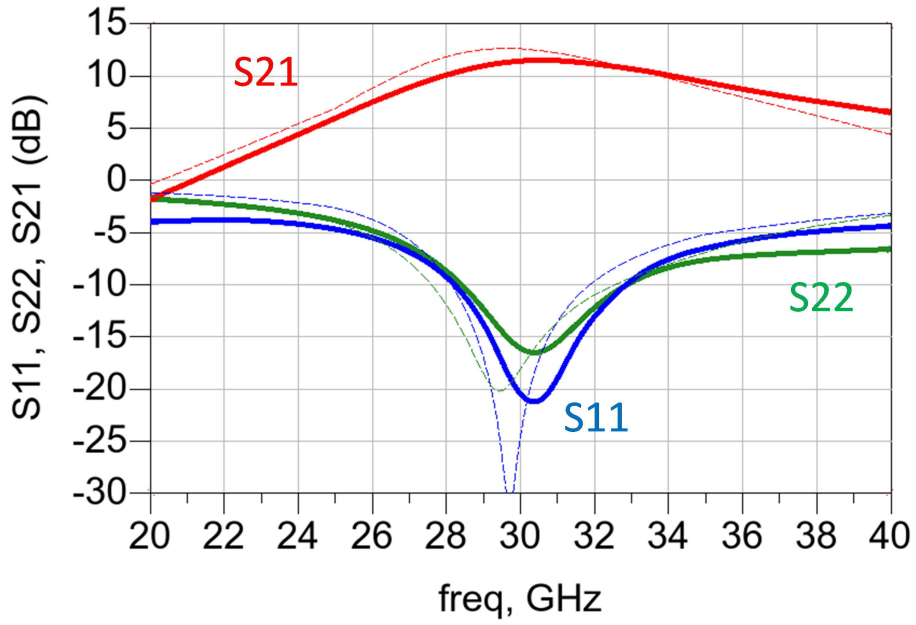

Figure 3.13: Measured (solid line) and simulated (dotted line) S-parameters for PA1.

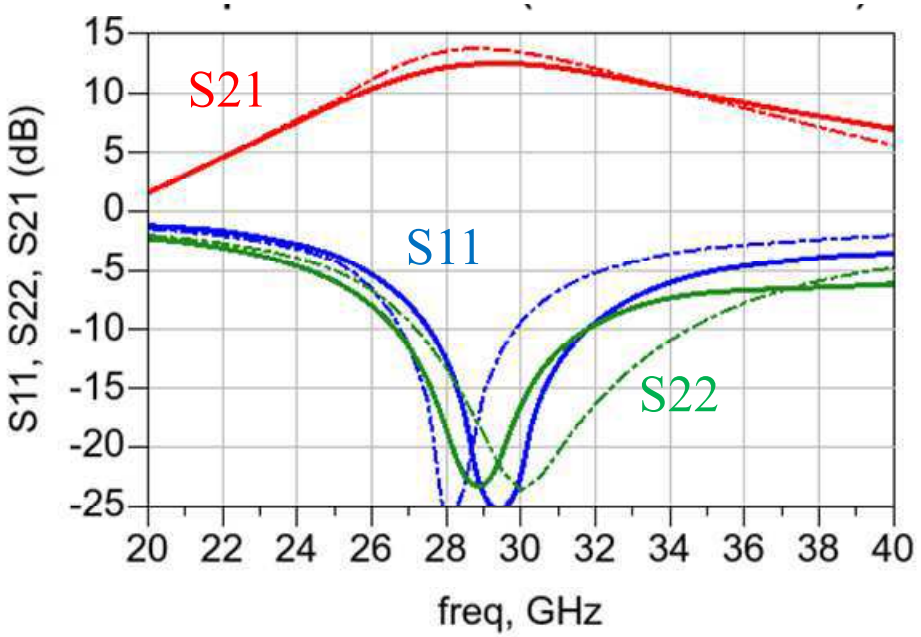

Figure 3.14: Measured (solid line) and simulated (dotted line) S-parameters for PA2.

3.5.1 CW Measurements

S-parameter measurements (Fig. 3.13, 3.14) show that both the amplifiers have input and output well matched at 28 GHz. PA1 demonstrates a small signal gain (S_{21}) of 13 dB at 30 GHz. The 3-dB gain-bandwidth of the PA is about 11 GHz (24 35 GHz), corresponding to a fractional bandwidth of 38%. PA2 demonstrates S_{21} of 13 dB at 29 GHz. The 3-dB gain-bandwidth of the PA is about 10 GHz (25 35 GHz), corresponding to a fractional bandwidth of 33%.

Large signal measurements were done at two bias conditions, high bias (V_{G1} = 0.4 V, V_{DD} = 5.2 V) and low bias (V_{G1} = 0.3 V, V_{DD} = 5 V) bias. Both the bias conditions correspond to class AB bias, at different depths. The low bias condition is very close to class B as seen from the class B like efficiency back-off characteristics. The high bias condition is significantly closer to Class A.

PA1 (Fig. 3.13) demonstrates a saturated output power of 23.7 dBm (230 mW) and 29% PAE at 29 GHz with P_{1dB} of 20 dBm, with the high bias. Under low bias PA1 achieves a peak PAE of 30% and output power of more than 23.5 dBm. PA2, (Fig. 3.14) demonstrates a P_{sat} of 24.8 dBm (300 mW) and 26% PAE at 29 GHz, with P_{1dB} of 21 dBm, with high bias. Under low bias PA2 achieves a peak PAE of more than 29% and output power of more than 24.3 dBm. A frequency sweep indicates a 1-dB P_{sat} bandwidth of 10 GHz (24 34 GHz) and $>$ 25% PAE bandwidth of 5 GHz for both the PA (Fig. 3.17, 3.18). The scatter in the data is due to the frequency ripple in the measurement setup involving long cables.

To check the reliability, the PA was operated for more than 48 hours continuously at peak output power and the output power was measured to be stable within +/- 0.1 dB.

Table 3.3 provides a comparison of the multigate PA results with some of the state-of-art results for Silicon-based and III-V mm-wave power amplifiers in the

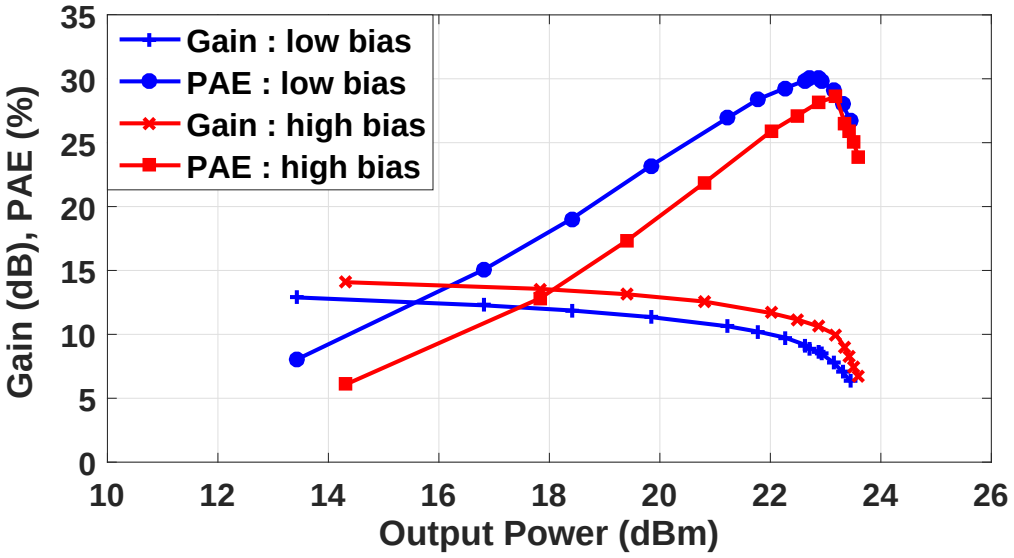

Figure 3.15: Measured gain and PAE of PA1 at high and low bias at 29 GHz.

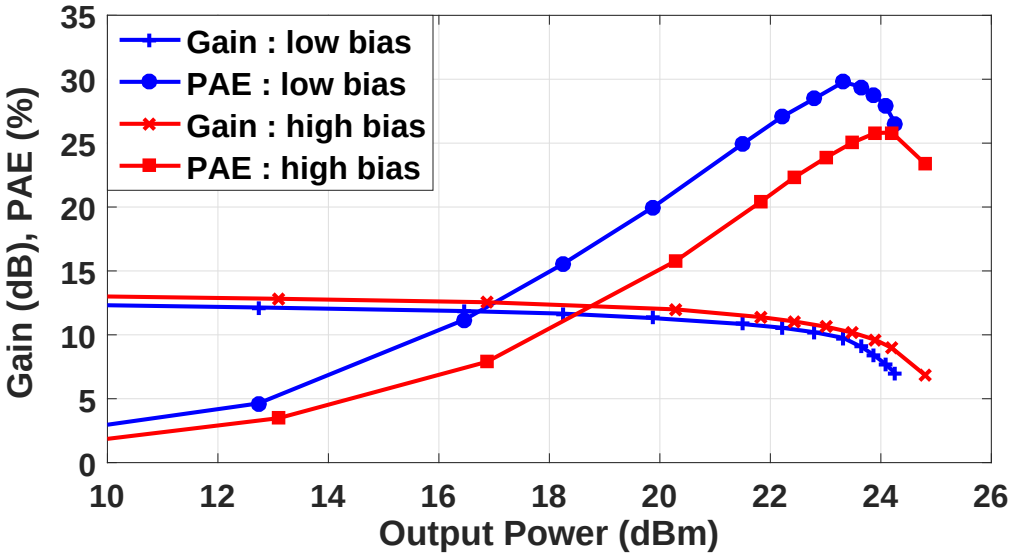

Figure 3.16: Measured gain and PAE of PA2 at high and low bias at 29 GHz.

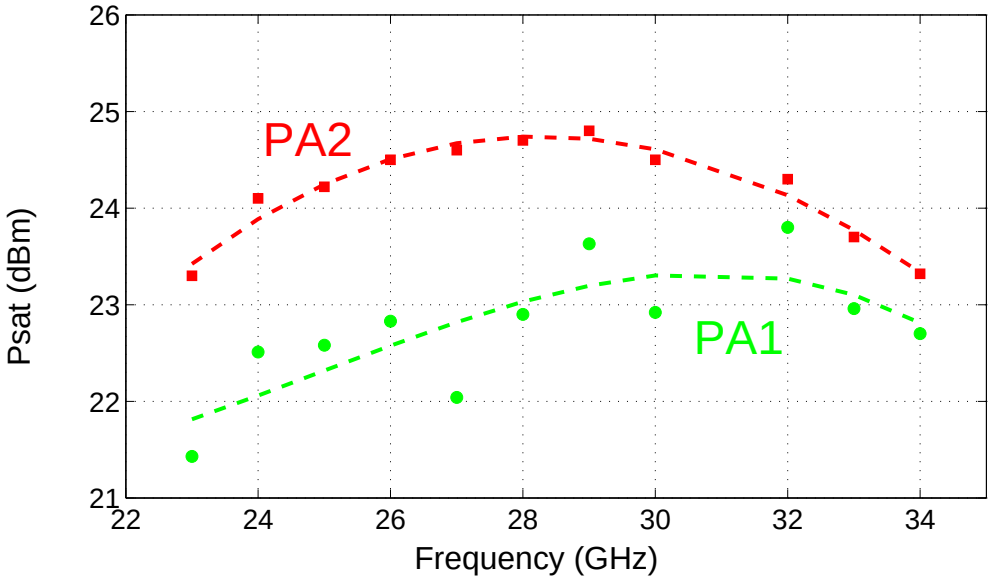

Figure 3.17: Measured saturated output power for PA1 and PA2 (Dots - measured points, thin line - best fit curve.

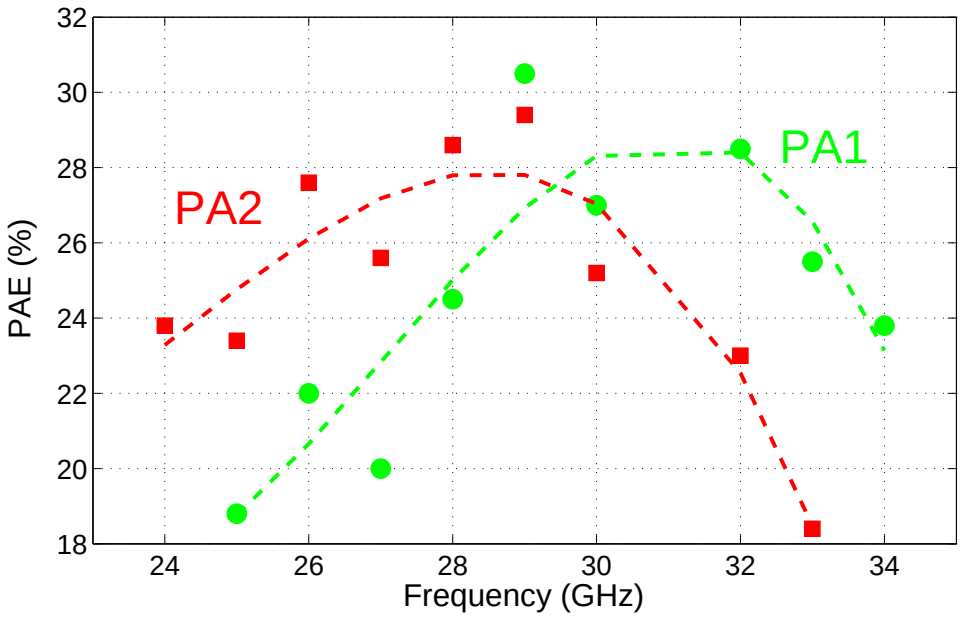

Figure 3.18: Measured peak PAE for PA1 and PA2 (Dots - measured points, thin line - best fit curve.

Table 3.3: Comparison with current State-of-the-Art

Ref.	**Tech.**	**Design**	**Freq. (GHz)**	P_{sat} **(dBm)**	**Peak PAE (%)**	**Gain (dB)**	**Chip area** (mm^2)
This work	45 nm SOI CMOS	Multigate-cell	29	24.8	29	13	0.3
[23]	45 nm SOI CMOS	4-stack	41	21.6	25.1	8.9	0.3
[25]	45 nm SOI CMOS	4-stack, Diff. off-chip load	45	24.3	14.6	> 18	0.8
[47]	45 nm SOI CMOS	6-stack	18	26.1	11	5	0.5
[24]	130 nm SiGe	2-stack	41	23.4	34.9	12.5	1.0
[48]	130 nm SiGe	common source	28	17.1	42	21.2	0.5
[49]	120 nm SiGe	cascode	28	18.6	35.3	15.8	0.4
[50]	150 nm GaAs pHEMT	Doherty	26.4	25.3	38	10.3	25
[51]	150 nm GaN HEMT	8-way power combined	28	39.4	26	24	9.7

20 50 GHz frequency range. The present work has one of the best combinations of P_{sat} and efficiency (Fig. 20). The SiGe HBT 2-stack results [24] are closest in performance to the CMOS result here.

The multigate-cell PA delivers a saturated output power of 300 mW from less than 0.1 mm^2 of chip area (excluding the pads). This corresponds to 3 W/mm^2 of output power density. Various demonstrations of efficient multi-way (up to 16-way) on-chip power combining have been reported at mm-wave frequencies recently [11, 52, 53]. The multigate-cell PA demonstrated here could in principle be used

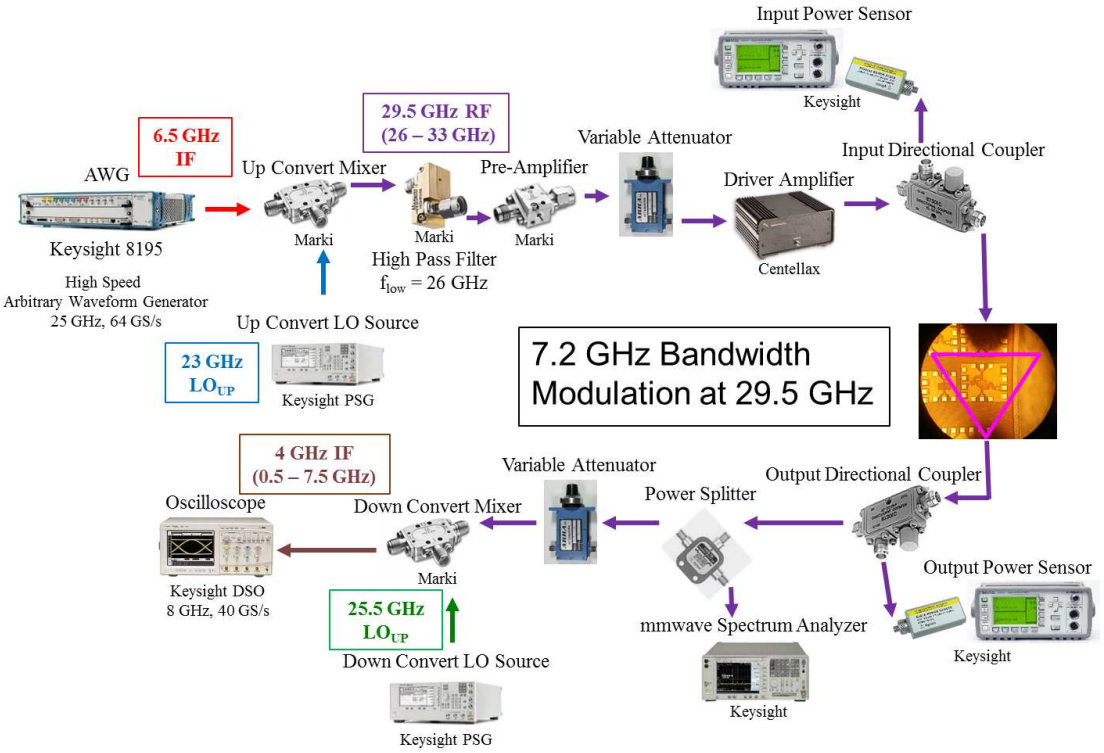

Figure 3.19: 28 GHz Modulated Signal measurement setup.

as a building block to implement a 4- or 8-way power combined amplifier. Such a design would have more than one Watt of output power delivered from a chip of less than 1 mm^2 area and could be useful for 5G access point transceiver design. This would require, however, careful study of heat removal methods since such a chip would be dissipating about 5 W/mm^2 of DC power. The segmented nature of the multigate-cell array also enables it to be readily made into a mm-wave power DAC [54]. However with the distributed layout one should be careful to avoid non-linearities that could arise from differences in output from different unit cells, due to on-chip process/temperature variations.

3.5.2 Modulated Signal Measurements

Modulated Signal measurements (Fig. 3.19) demonstrates the capability of the multigate-cell PA as a broad-band, high-power, high-efficiency, linear PA. The measurements are done without any digital pre-distortion. The input signal to the PA is frequency equalized due the high dispersion of the measurement set up. The PAE at backed off power levels and hence the average PAE of modulated signal shows that the PAE - Pout relationship is very close to a class-B PA while still having linear gain (Fig. 3.20). The broadest bandwidth measured (7.2 GHz) is limited by the measurement set up (capture bandwidth of the Receiver Oscilloscope). The stacked FET PA shows very good ($< 3^0$) AMPM response. The maximum average output power achieved with EVM meeting transmit specs is nearly Psat - PAPR. This shows that no additional power back-off is needed for linearity. Broadband modulated signal measurements of the multigate-cell PA showed 36 Gbps of data rate at 28 GHz (7.2 GS/s of single carrier 32 QAM signals with average output power of 17 dBm and 14% PAE and 6 GS/s of single carrier 64 QAM signals with average output power of 14 dBm and 9.3% PAE)(Table 3.4). This is the maximum data rate to be demonstrated at any frequency below W-band. Also the average power level achieved is nearly an order of magnitude higher than previously reported Silicon PA at same frequency.

3.6 Conclusion

A multigate-cell device layout has been studied for the design of stacked-FET CMOS mm-wave power amplifiers. Design considerations have been presented including advantages from reduction of layout parasitics and topology which facilitates heat-sinking in an SOI process. The difficulty of impedance matching at

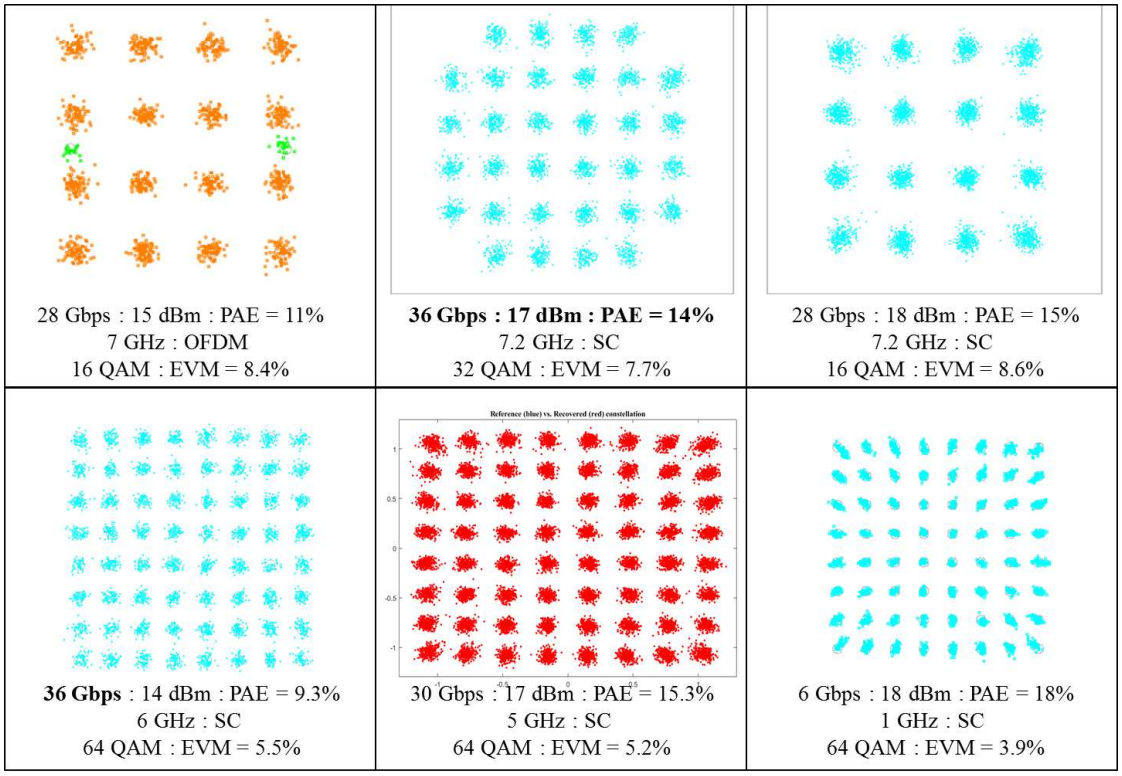

Figure 3.20: PA Output constellation with different modulation schemes and bandwidths.

Table 3.4: Modulated Signal Measurement Results

BW (GHz)	**Carrier (SC/OFDM)**	**QAM**	**Data Rate (Gbps)**	P_{out} **(dBm)**	**PAE (%)**	**EVM (%)**	**SNR MER (dB)**
7	OFDM	16 QAM	28	15	11	8.4	-
7.2	SC	32 QAM	36	17	14	7.7	22.3
7.2	SC	16 QAM	28.8	18	15	8.6	21.3
6	SC	64 QAM	36	14	9.3	5.5	25.1
5	SC	64 QAM	30	17	15.3	5.2	25.6
5	SC	128 QAM	35	16	13	5.1	25.8
1	SC	128 QAM	7	17	16	3.7	28.6
1	SC	64 QAM	6	18	18	3.9	28.1

intermediate nodes in the stack was also highlighted. Several Ka-band mm-wave power amplifiers have been demonstrated using the multigate-cell architecture. A saturated output power of 300 mW and peak efficiency of 30% was achieved at 29 GHz from the amplifiers. The absence of tuning elements favors broadband operation for the multigate-cell approach. In keeping with this expectation, the measured amplifiers had 3-dB small signal and 1-dB large signal bandwidth of 10 GHz (35%) of centered around 29 GHz. Broadband modulated signal measurements of the multigate-cell PA showed 36 Gbps of data rate at 28 GHz without using any DPD. The multigate-cell offers a highly compact, scalable and reliable building block for high power mm-wave PA design.

3.7 Appendix : Approximate Thermal Analysis

Approximate temperature rise at the FET can be estimated from the following thermal resistance calculations. The heat is generated in the FET channel and spreads readily across the silicon device diffusion area (as defined by STI oxide at the multigate transistor edges). The main heat flow pathways from the device are (a) vertical conduction from the silicon device layer to the heat-sink at the silicon substrate backside through the buried-oxide (BOX); and (b) lateral spreading through the silicon device layer to reach the source/drain contacts from where there are various pathways to the heat-sink. Here we consider wire-bond packaging which has a heat-sink on substrate backside; different pathways apply for flip-chip bonding.

For critical heat flow paths in the structure used in this work, a one-dimensional approximation can be used to estimate thermal resistance given by

$$R_{th} = \frac{1}{\kappa_{th}} \frac{l}{A} \tag{3.3}$$

where κ_{th} is the thermal conductivity of the material, l is the length of conductor and A is the area of the conducting plane.

For heat flow from rectangular regions through a substrate of constant thermal conductivity, the heat spreads in a prism-like fashion with an angle of approximately 45^o. The thermal resistance [55] in this case can be approximated as

$$R_{th} = \frac{1}{2\kappa_{th}} \frac{1}{(L-W)} ln \left(\frac{W+2h}{L+2h} \frac{L}{W} \right) \tag{3.4}$$

where L and W are the length and width of the rectangular sheet heat source, h is the height of the heat conducting column (substrate thickness). When L and W are similar in value this reduces to

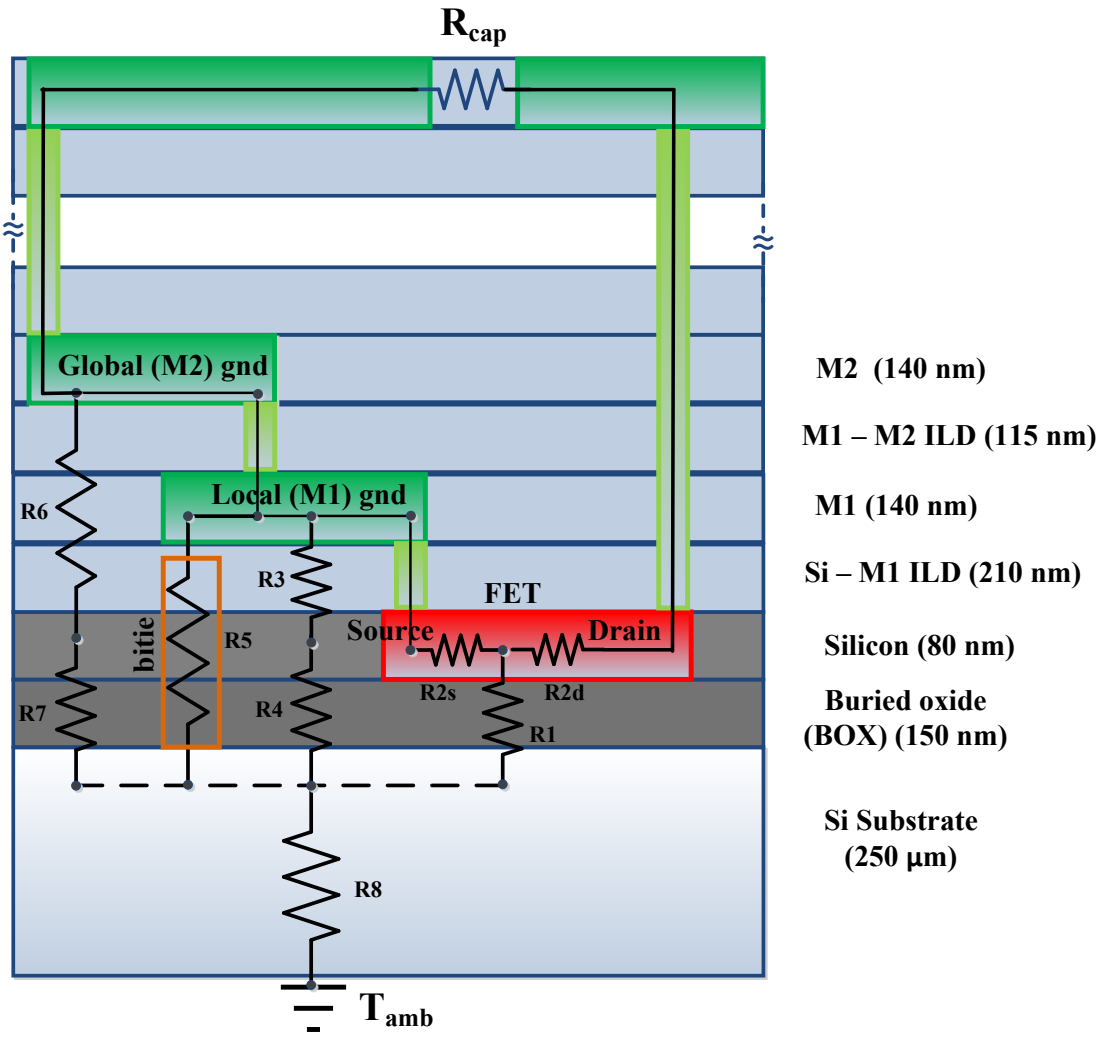

Figure 3.21: Schematic cross section of the chip showing thermal pathways and equivalent thermal resistances.

$$R_{th} = \frac{1}{\kappa_{th}} \frac{h}{L\,(L + 2h)} \tag{3.5}$$

Fig. 3.21 shows the various thermal pathways from the active device to the heat sink at the back of substrate. Fig. 3 shows the layout of a unit cell with 1.2 $\mu m\times$ 1.5 μm device area and 6.3 μm^2 of M1 ground ring area within a unit cell of 3.5 $\mu m\times$ 3.0 μm. For thin film BOX with κ_{th} = 0.8 W / m - K, the thermal resistance from the device layer to the top of the substrate through the BOX, per unit cell, can be estimated as

$$R_1 = \frac{1}{\kappa_{BOX}} \frac{h_{BOX}}{Area_{device}} = 104K/mW \tag{3.6}$$

The lateral conduction path inside the silicon device layer can be modeled as doubly contacted from both source and drain side for the multigate cell. Assuming values of $\kappa_{Si_{thinfilm}}$ = 70 W/m - K, the lateral spreading thermal resistance is

$$R_2 = \frac{1}{12\kappa_{Si_{thinfilm}}} \frac{length_{diffusion}}{W_{diffusion}\, h_{diffusion}} = 18K/mW \tag{3.7}$$

The thermal connection between the source and heat-sink is accomplished via the ground metal as described below. The thermal connection between the drain and the heat-sink is accomplished via the drain supply transmission line, which is bypassed using a large 50 pF capacitor. It can be shown that the thermal resistance of this capacitor (R_{cap}) is negligible compared to other contributions of the thermal pathway. The thermal resistance of the M1 ground plane to the substrate, is calculated as

$$R_{34} = R_3 + R_4 = 42 + 45 = 87K/mW \tag{3.8}$$

If bitie are used, assuming a conservative thermal conductivity of 40 W/m-K for the bitie fill and area of 1.1 μm^2, R_5 = 10 K/mW comes in parallel with R_{34}.

For the 256 unit cells array, a ground plane (180 $\mu m\times$ 180 μm) using multiple metal layers (lowest layer being M2) is laid around the FET array. Due to finite conductivity of metal the heat can be estimated to be spread to a distance

$$L_{eff} = \sqrt{\frac{\kappa_{metal}\, t_{metal}\, t_{dielectric}}{\kappa_{dielectric}}} \tag{3.9}$$

The value of L_{eff} can be estimated to be about 25 μm. This means we can assume an effective ground plane of width 25 μm around the unit cell array. The effective thermal resistance per unit multigate cell of this ground plane can be estimated as

$$R_{67} = 256(R_6 + R_7) = 30K/mW \tag{3.10}$$

This calculation considers the fact that the ground planes are cheesed as per CMOS design rules and hence has a fill factor of only 50%.

All the above calculations assume one-dimensional conduction obeying (3). Since the substrate thickness is larger than the area dimensions of the FET array + ground plane the thermal conduction in substrate becomes three-dimensional, obeying (4). Assuming $\kappa_{Si_{bulk}}$ = 130 W/m - K, the effective thermal resistance per unit cell of the substrate can be estimated as $R_8 \approx$ 4 K/mW.

Now the total thermal resistance of the unit cell can be estimated as

$$R_{eff} = R_8 + (\, R_1 \,||\, [R_2 + (R_{34}||R_5||R_{67})\,]\,) \approx 24K/mW \tag{3.11}$$

The maximum output power each of the unit cells deliver is close to 1 mW

at about 25% drain efficiency. This would lead to peak power dissipation of 3 mW per unit cell. Therefore the maximum temperature rise of the device can be estimated to be 72^0 C.

Acknowledgment

Chapter 3 is mostly based of materials used in the following publications

The material as it appears in J. A. Jayamon, J. F. Buckwalter, and P. M. Asbeck, "Multigate-cell Stacked FET Design for Millimeter-wave CMOS Power Amplifiers," *IEEE Journal of Solid-State Circuits,* Sept 2016. The dissertation author was the primary investigator and author of this material, and co-authors have approved the use of the material for this dissertation.

The material as it appears in J. A. Jayamon, J. F. Buckwalter, and P. M. Asbeck, "28 GHz $>$ 250 mW CMOS Power Amplifier using multigate-cell design," in *2015 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS),* Oct 2015. The dissertation author was the primary investigator and author of this material, and co-authors have approved the use of the material for this dissertation.

Chapter 4

Millimeter-wave PMOS Power Amplifier

E-band frequencies (60 GHz - 90 GHz) are used for both radar applications (77 GHz automotive radar) as well as point-to-point communication for wireless back-haul (71 GHz - 76 GHz and 81 GHz - 86 GHz). Recently 73 GHz band is being investigated for being used for fifth generation (5G) wireless communication systems [56]. 77 GHz radar requires only less than 10 mW of output power and typically SiGe and more recently CMOS chipsets are able to have single chip solutions [57]. But E-band back-haul for point-to-point mm-wave communication requires more than 100 mW - 1 W of transmitted power to be able to sustain very high data rates. Both InP and GaN transistors have demonstrated capabilities to have 100 mW of power from single un-power-combined amplifier at E-band [31,58]. With efficient power combining schemes the output power can be further increased. But the integration needs for emerging complex systems favor silicon designs. Several power combining schemes including FET stacking [34, 59], on-chip transmission line power combining [9,10], transformer power combining [13,14] and spatial

power combining [17, 60] have been used to design reliable high power SiGe and CMOS amplifiers in these frequency ranges. The PAE of most of them is less than 15% however.

This chapter presents an investigation of the use of PMOS devices in a standard CMOS process as an alternative to NMOS and SiGe devices for the design of high power mm-wave power amplifiers. Due to process improvements associated with dimension scaling, material system engineering and structural changes, the PFET devices have nearly similar performance as of NFET devices in deeply scaled CMOS processes. At the same time PMOS breakdown voltages are higher than that of NMOS. This allows a mm-wave power amplifier made exclusively with PMOS to have similar or higher power levels as of an NMOS amplifier depending on the ratio of the respective FET parameters. Previous research efforts have used PMOS FETs in a push-pull [61,62] or inverter-like [63] configuration. In this work a PMOS-alone PA at E-band using 32 nm CMOS SOI process utilizing stacked-FET configuration is presented. This PA achieved a measured maximum output power of 19.6 dBm and a peak efficiency of 24% at 78 GHz. This represents the highest efficiency and the highest output power without elaborate power combining schemes reported for any Silicon power amplifier in this frequency range.

Section II of this chapter describes the device physics associated with the scaling of the CMOS FETs. Section III gives a comparison of measured performance of equivalent NMOS and PMOS devices. The design of the E-band PA is described in section IV and the measurement results are given in Section V.

4.1 MOSFET Device Physics

Dimension scaling of CMOS devices has been primarily motivated by the reduction in the area and hence the cost of Silicon CMOS chipsets as predicted by Moores law. As the gate length (L_g) is scaled down the current and frequency capabilities of the CMOS devices have been consistently increasing, while special techniques have to be implemented to prevent the increase of leakage current. Process and structural innovations like SiGe strained Silicon ($L_g < 90$ nm), High-K Metal Gate (HKMG) ($L_g < 40$ nm) and FinFET and Ultra-Thin Body Buried Oxide Fully Depleted Silicon on Insulator (UTBB-FDSOI) ($L_g < 20$ nm) have helped performance improvement. At the highly scaled nodes ($L_g < 40$ nm) the comparison of mobility and reliability of N-channel and P-channel devices shows interesting merits for PFETs.

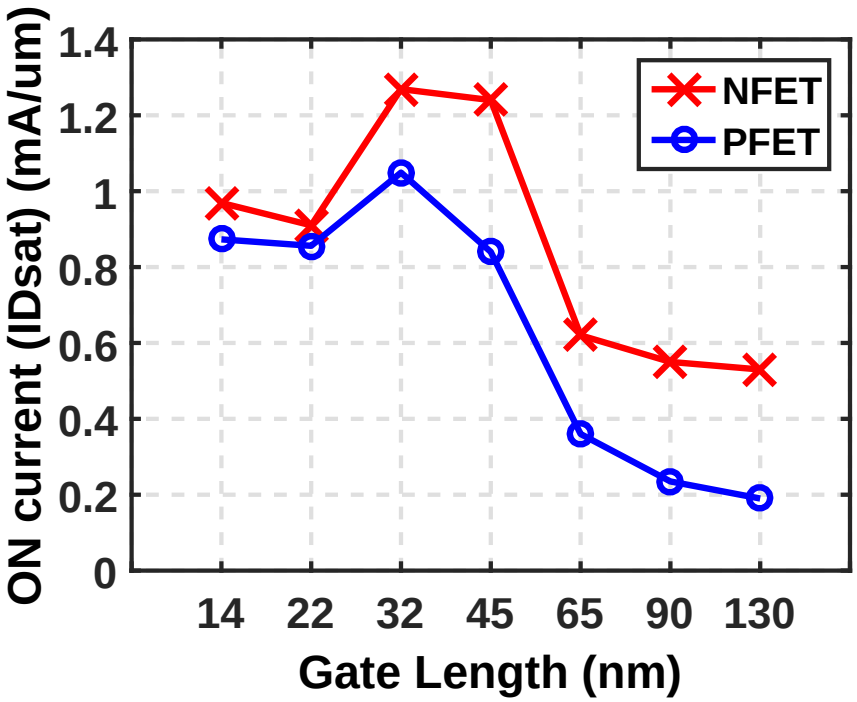

Figure 4.1: Current density of NMOS and PMOS transistors of different generations of IBM CMOS FET.

4.1.1 Mobility

The rate of increase of ON current (I_{ON}) of NMOS FETs with advancement in the CMOS technolgy node has decreased considerably for nodes below 28 nm [64]. PMOS FETs followed a similar trend but the slowing has been soft due to the improvements in strain engineering. Thus initially for long channel devices the PMOS I_{ON} was about half of that of NMOS whereas currently for many extremely scaled processes it is as close as 94% of NMOS I_{ON} [65,66]. The result is apparent in the fact that in CMOS inverter design the W_P/W_N ratio has changed from above 2 to as low as 1.1 in deeply scaled CMOS processes. Fig. 4.1 shows the I_{ON} for different generation of N and P FETs available in commercially available IBM CMOS processes.

This relative improvement of PMOS compared to NMOS has been result of: a. strain engineering; b. change in device orientation from [100] to [110]; c. increasingly ballistic transport and d. wiring parasitics.

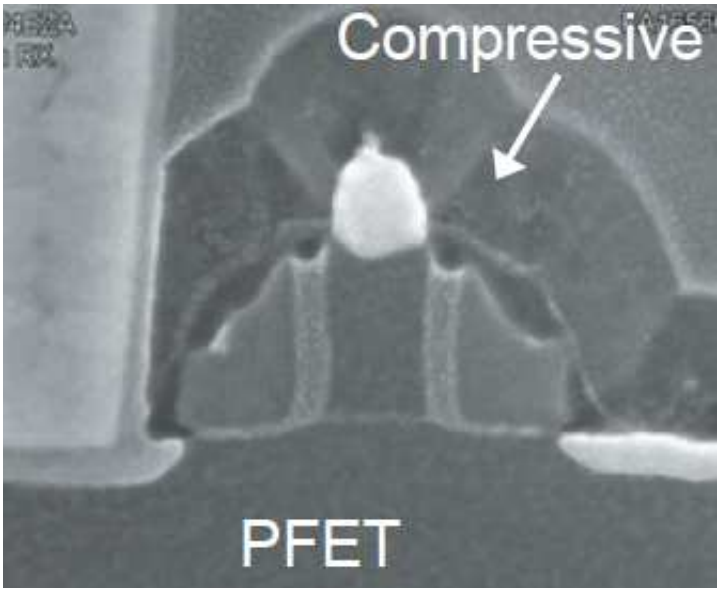

Figure 4.2: Cross section of a PFET showing the SiN liner on gate for inducing compressive stress [1].

Hole mobility increases with compressive stress and electron mobility increases with tensile stress. Addition of Ge to the Silicon channel increases the uniaxial compressive strain and hence embedded SiGe (eSiGe) is used to increase the hole mobility in PFET. Also strain can be induced by the use of nitride (SiN) liners on the gate (Fig. 4.2) [1]. This can be done for NMOS and PMOS by using Dual Stress Liners (DSL) which creates tensile strain for NMOS and compressive strain for PMOS in the same substrate. Other techniques including stress memorization and strain induced by regrown source/drain contacts are also used to increase the mobility of charge carriers. The mobility improvement for holes has been much higher than that of the electrons in most of the above mentioned techniques. This resulted in PMOS current and transconductance improving by a higher fraction than NMOS.

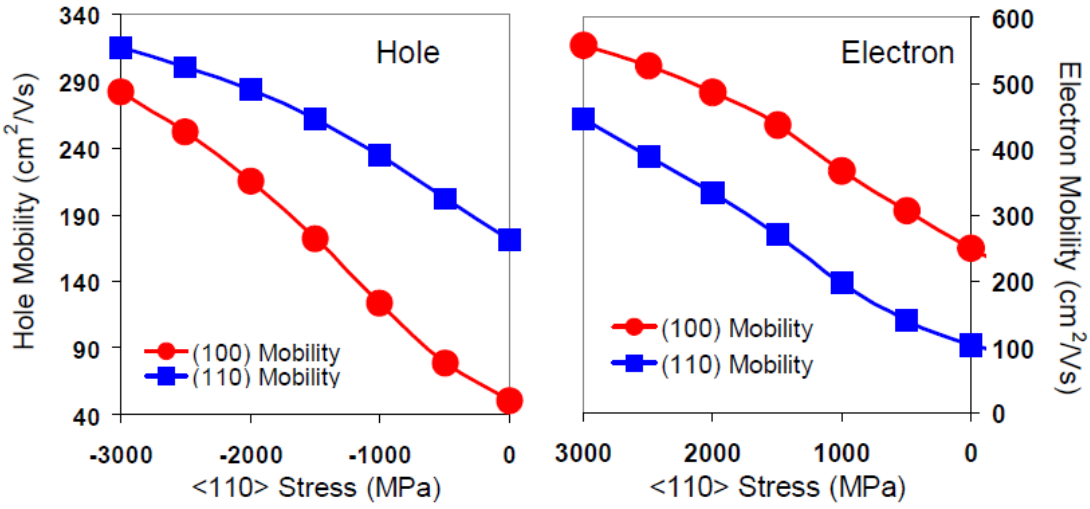

Figure 4.3: Simulated hole and electron mobility for (100) and (110) silicon substrates as a function of stress [2].

In Silicon electron mobility is higher for [100] direction than [110], while hole mobility in [110] is higher than [100] direction (Fig. 4.3) [2]. Traditionally CMOS devices used to be made on [100] orientation. But most of the deeply scaled devices

especially FinFETs are made in [110] orientation. For NMOS the loss in mobility of electrons with change in orientation is partially mitigated by application of stress. But for PMOS the change in orientation and stress considerably increases the hole mobility and brings it closer to the electron mobility for NMOS.

Both the above mentioned results enhances the hole mobility compared to the electron mobility. But at the same time increasingly ballistic nature of the carrier transport in extremely scaled devices makes the current less dependent on channel mobility and more on the source density of states and emission properties. This also makes PMOS and NMOS similar.

For mm-wave power amplifier design f_{max} serves as the most important figure of merit. As the device is scaled, due to the smaller gate length the gate resistance (R_g) increases. Smaller lithographic dimensions bring the FET contacts closer thereby increasing the capacitive wiring parasitics($C_{gd,ext}$, $C_{gs,ext}$) and thinner interconnects increases resistive wiring parasitics. This increases the portion of $C_{g,ext}$ compared to $C_{g,int}$ in the final C_{gs} and C_{gd} values. This is same for both NMOS and PMOS. This leads to similar performance by NMOS and PMOS as the similar values of parasitics overshadow the dissimilar intrinsic FET values.

4.1.2 Reliability

The primary reliability concerns in short channel MOSFETs are due to Hot Carrier Injection (HCI), Time Dependent Dielectric breakdown (TDDB), Bias Temperature Instability (BTI) (Positive BTI (PBTI) for NMOS and Negative BTI (NBTI) for PMOS) and electro-migration. The breakdown voltages of PFETs are typically higher than those for NFETs. This is partially due to the difference in impact ionization rates of holes and electrons and tunneling behavior of the two device structures. The impact ionization rate of holes in Silicon is much lower

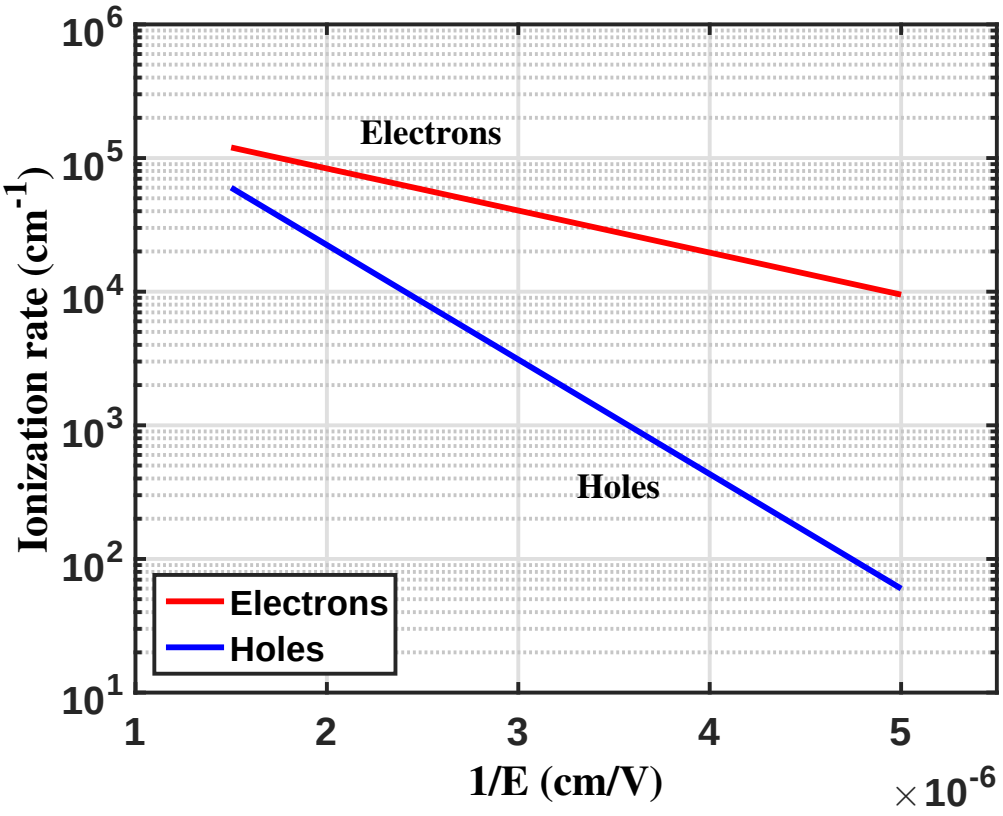

Figure 4.4: Impact ionization rate of electrons and holes in Silicon.

than that of electrons (Fig. 4.4) [67]. This leads to lower probability of avalanche breakdown and creation of hot carriers (hot holes) in p-channel MOSFETs at a given electric field. Also the n+ source and drain junctions are more abrupt than p+ junctions leading to lower breakdown for NMOS [68]. In general, injection from Si into SiO_2 (or related High-K dielectrics) is much more likely for hot electrons than for hot holes because (a) electrons can gain energy from the electric field more readily than holes due to their smaller effective mass; and (b) the Si-dielectric interface energy barrier is larger for hole ($\approx$ 4.8 eV for SiO_2) than for electrons ($\approx$ 3.1 eV for SiO_2) as shown in Fig. 4.6. The difference in energy barrier between holes and electrons is expected to increase further with use of High-K dielectrics instead of SiO_2 .

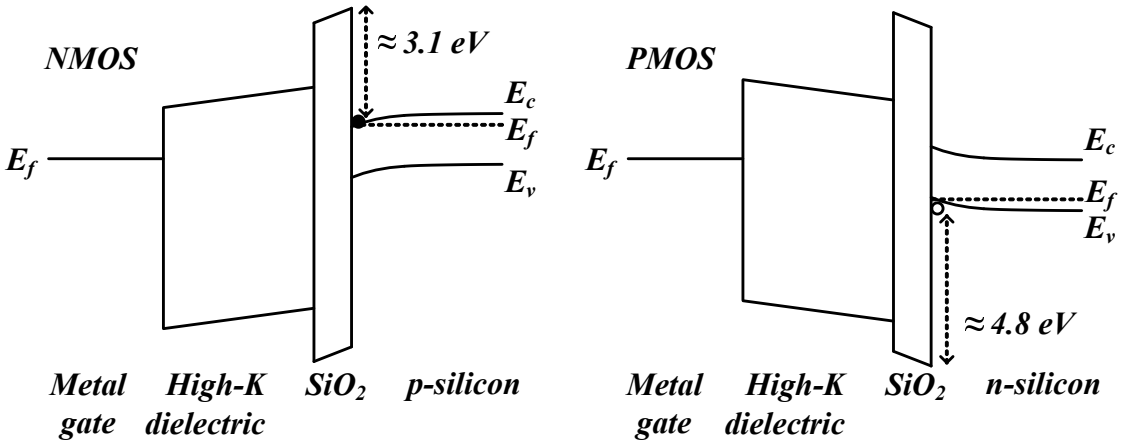

Figure 4.5: Simplified CMOS band diagram showing $Si - SiO_2$ energy barrier for electrons in NFET and holes in PFET.

Fig. 4.5 shows a simple conceptual band diagram for NMOS and PMOS. Commonly the band narrows at the channel for PFET due to the presence of eSiGe. This is followed by a interfacial layer of SiO_2 and then lower band-gap but wider thickness High-K dielectric stack (mostly HfO_2) followed by metal gate (TiN) and then poly-Si. Hot carrier effect in FETs can be caused due to either Conducting Hot Carriers ($V_{gs} > V_t$) or Non-Conducting Hot Carriers ($V_{gs} < V_t$). In amplifiers, especially biased amplifiers, only conducting hot carriers are relevant. They can be two types by origin - channel hot carriers or drain-avalanche hot carriers. These hot carriers once generated can either cross the channel-dielectric band barrier and get trapped in the dielectric or generate interface states. The trapped charges can change the threshold voltage. The interface states can reduce the drain current, degrade subthreshold slope and cause higher leakage. Interface states affect both NFETs and PFETs, whereas charge trapping is more a problem in PFET than NFET [68].

The dependence of current degradation due to HCI on device parameters and operating conditions can be roughly expressed as

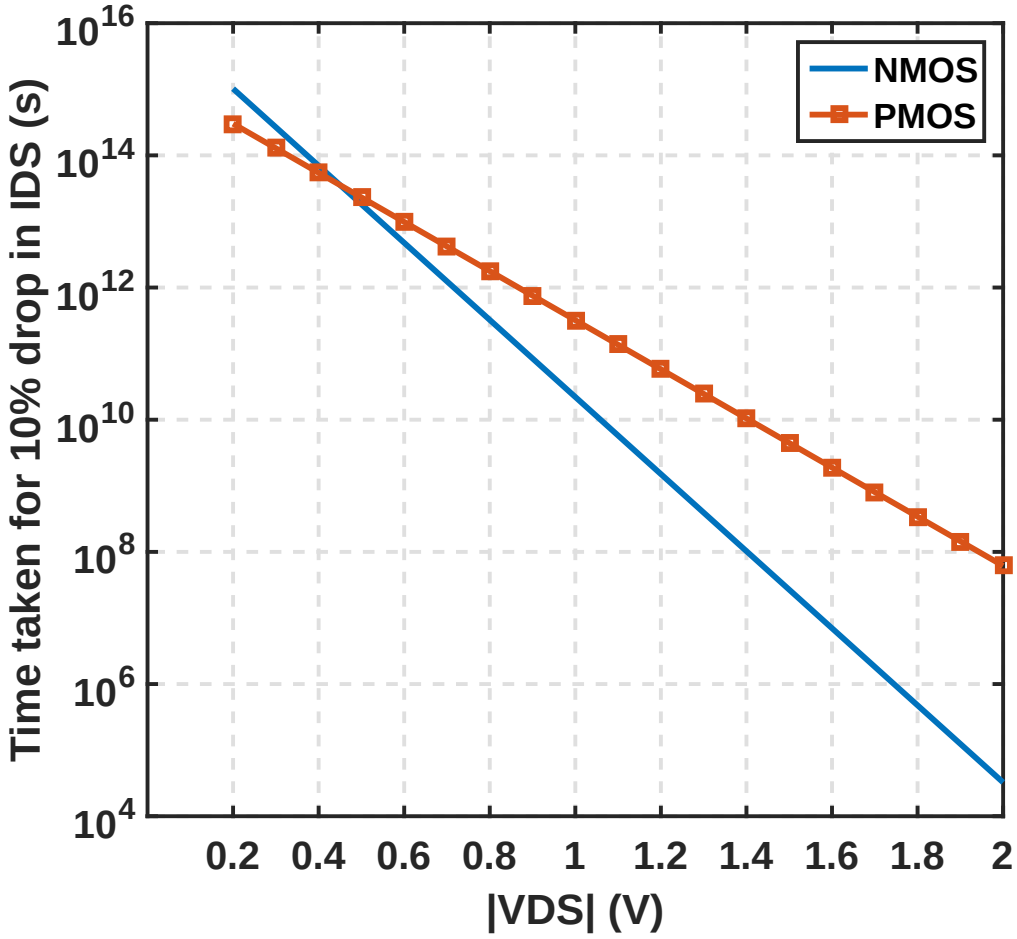

Figure 4.6: Simulated time taken in seconds for a 10% drop in ON current due to HCI for NMOS and PMOS vs. stress voltage.

$$\Delta I_D \propto \frac{W^a}{L^b} \frac{e^{(c|V_{DS}|+d|V_{GS}|}}{e^{\frac{m}{T_{junc}}}} t^n \tag{4.1}$$

Where a, b, c, d, m, n are process and charge carrier dependent aging variables and t is the duration of stress [69]. For 32 nm SOI process the time taken for 10% drop in I_D is calculated using the aging parameters provided by IBM. This calculation assumes a 1 μm wide FET, V_{gs} = 0.5 V and T_{junc} =100 degree Celsius. The results are plotted (Fig. 4.6) for different drain supply voltage for both NFET and PFET. As seen from the figure for a suggested 10 year lifetime

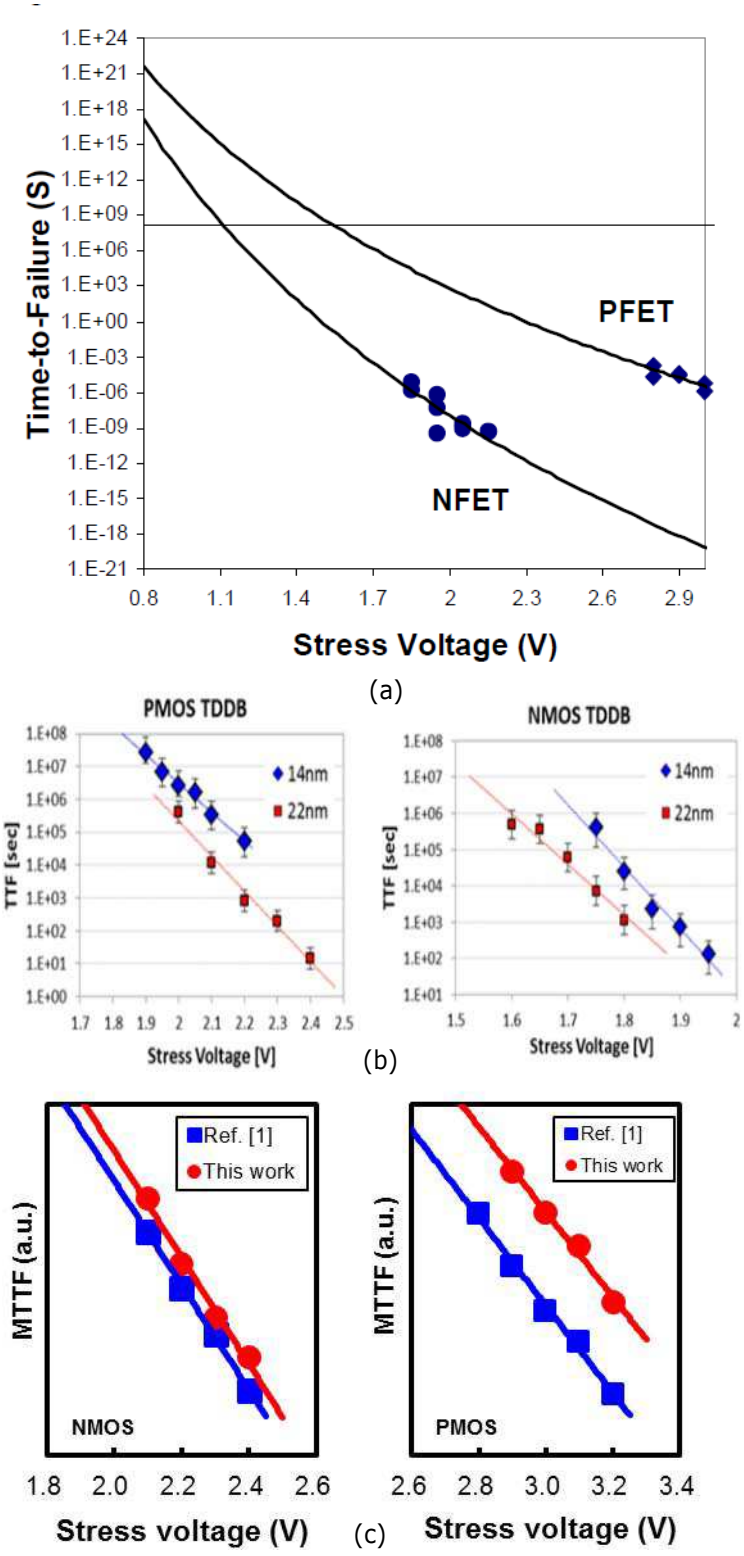

Figure 4.7: Reliability data (Mean Time To Failure - MTTF) of 14/16 nm FinFET published by (a) IBM [3], (b) Intel [4] and (c) TSMC [5].

(3×10^8 seconds) the NFET VDD has to be less than 1.3 V whereas PFET can operate up to 1.8 V supply for the same lifetime.

One should also note since there are multiple degradation mechanisms the leading cause for failure is extremely dependent on the device configuration like carrier type (N/P), nature of the dielectric (SiO_2 vs. HK), nature of interface (surface channel vs. buried channel), the gate length etc. For example at 32 nm HKMG SOI devices HCI is the leading cause for failure for NFETs whereas NBTI is a leading cause failure for PFETs.

These factors suggest that PMOS devices can be more robust than NMOS devices [70]. Recent data published by multiple foundries show that these trends carry forward to 14 nm FinFET devices [3–5] in relation to TDDB degradation mechanism. It has been shown that for the same Mean-Time-To-Failure (MTTF) due to TDDB the PMOS device can sustain 50% higher stress voltage than NMOS devices (Fig. 4.7).

4.2 32 nm SOI FET

IBM 32 nm CMOS SOI process is used in this study for performance comparison of very short channel NMOS and PMOS devices. This process uses High-K Metal Gate (HKMG) for the FETs and has an effective channel length of 32 nm ($L_{g,drawn}$ = 40 nm). ($L_{g,drawn}$ is the lithographic gate length and L_g is the effective gate length which is lower than $L_{g,drawn}$ due to the diffusion of dopants from source/drain into the area beneath gate.) The NFET / PFET has an equivalent gate oxide thickness T_{ox} of 1.4 nm / 1.55 nm and suggested process digital supply operating voltage of 0.9 V (nominal) / 1.0 V (maximum). Unlike bulk CMOS, SOI FETs do not need different well structure for N and P devices due to the isolation

provided by the buried oxide (BOX). Therefore the layouts of NFETs and PFETs in SOI are identical. Partially Depleted SOI (PDSOI), Floating Body, regular V_t devices are used for both NMOS and PMOS. Similar NFET and PFET of same size (28.8 μm), having identical layout (double side gate contacted, 0.8 $\mu m\times$ 36 fingers), were fabricated and measured (Fig. 4.8).

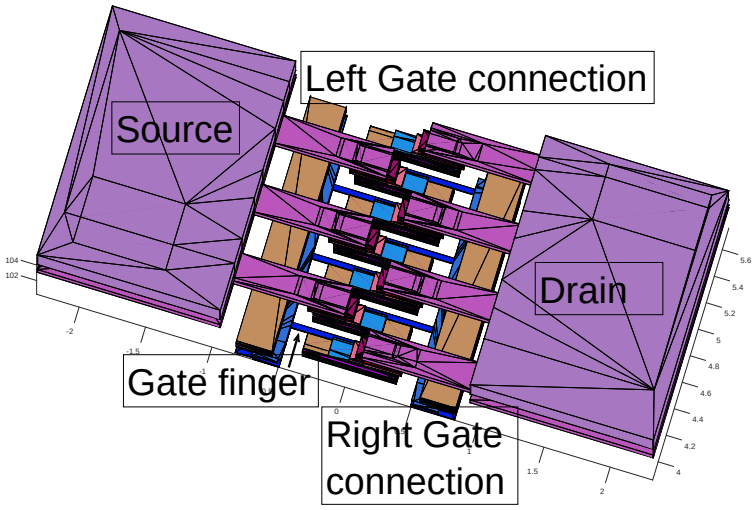

Figure 4.8: 3-D view of portion of FET wired to top level.

4.2.1 DC Characteristics

Fig. 4.9 shows ID-VDS measurements of the NFET and PFET. The thick lines in the figure are for $|VGS|$ and$|VDS| \leq 0.9$ V, which is the suggested nomial operating supply voltage for the process. With $|VDS| = |VGS| = 0.9$ V, NFET has I_{ON} of 32 mA (1.1 $mA/\mu m$) and PFET has I_{ON} of 26 mA (0.9 $mA/\mu m$). PFET current in this condition is 82% of NFET current. While operating in safe condition the NFET can have a maximum current of 36.5 mA at increased bias ($|VD| = 1.3$ V, $|VG| = 0.9$ V). PFET can achieve an almost similar maximum current of 35.5 mA with higher bias ($|VD| = 1.5$ V, $|VG| = 1.1$ V). The ID-VDS

curves for increased bias conditions are shown as dotted lines in Fig. 4.9.

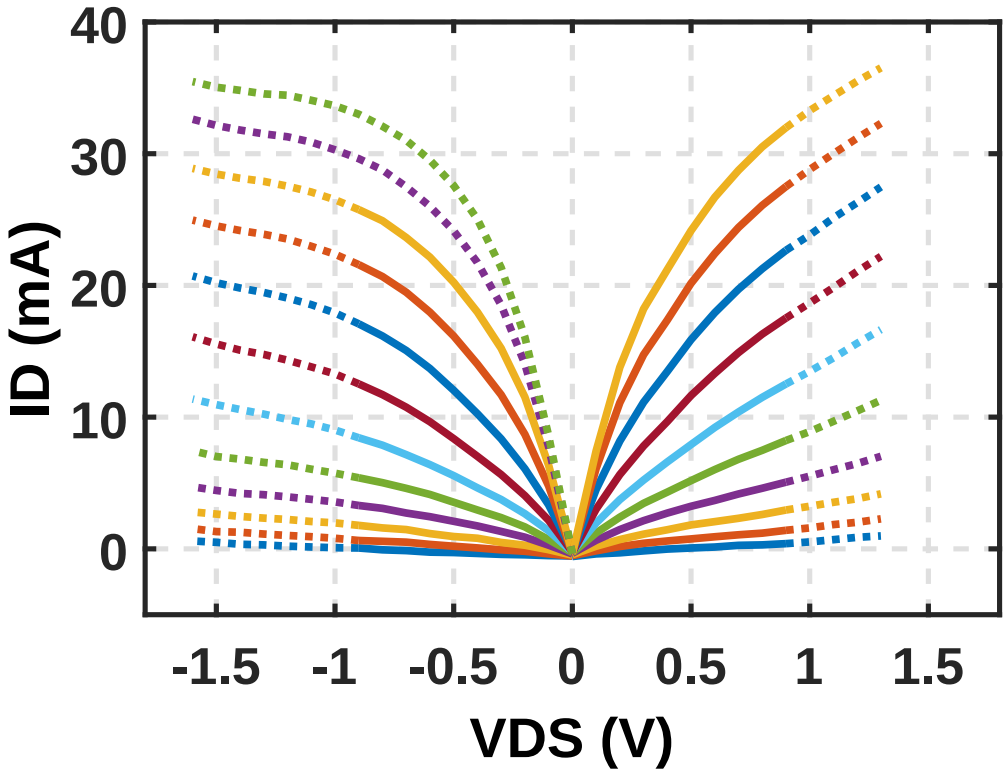

Figure 4.9: Measured $I_D - V_{DS}$ for W = 28.8 μm FETs : NMOS ($|V_{GS}|$ = 0 - 0.9 V) and PMOS ($|V_{GS}|$ = 0 - 1.1 V) with $\Delta|V_{GS}|$ = 0.1 V (Solid lines for $|V_{DS}|\&|V_{GS}| \leq 0.9$ V and dotted lines for $|V_{DS}|\&|V_{GS}| > 0.9$ V .

Due to short channel characteristics the output conductance of the FET is very high. Hence it is difficult to determine a single-valued knee voltage (V_{knee}), as used in many power amplifier calculations. From the measured $I_D - V_D$ (shown in Fig. 4.9) an effective knee voltage (V_{min}) can be roughly estimated to be about 0.4 V for both N and P FETs. With $|VG|$ = 0.9 V and $|VD|$ = 0.4 V, the N/P FETs draw 21 mA / 18 mA. Assuming $|VD|_N$ = 1.1 V and $|VD|_P$ = 1.4 V a simple class-A PA made with the corresponding N/P FET can achieve an output Power ($P_{out} = 1/4(V_{DD} - V_{knee}) \times I_{Dsat}$) of 3.7 mW / 4.5 mW. Thus even though the current (I_{Dsat}) is lower for PFET, the product of V_{DD} - V_{knee}

and I_{Dsat} and hence the output power (P_{out}) is higher for the PFET. A plot of $P_{out}(= 1/4(V_{DD} - V_{knee}) \times I_{DatV_{knee}}$)) vs. V_{knee} (which is essentially a scalar load-pull), estimated using the measured I_D and V_{DS} values from Fig. 4.9 is shown in Fig. 4.10.

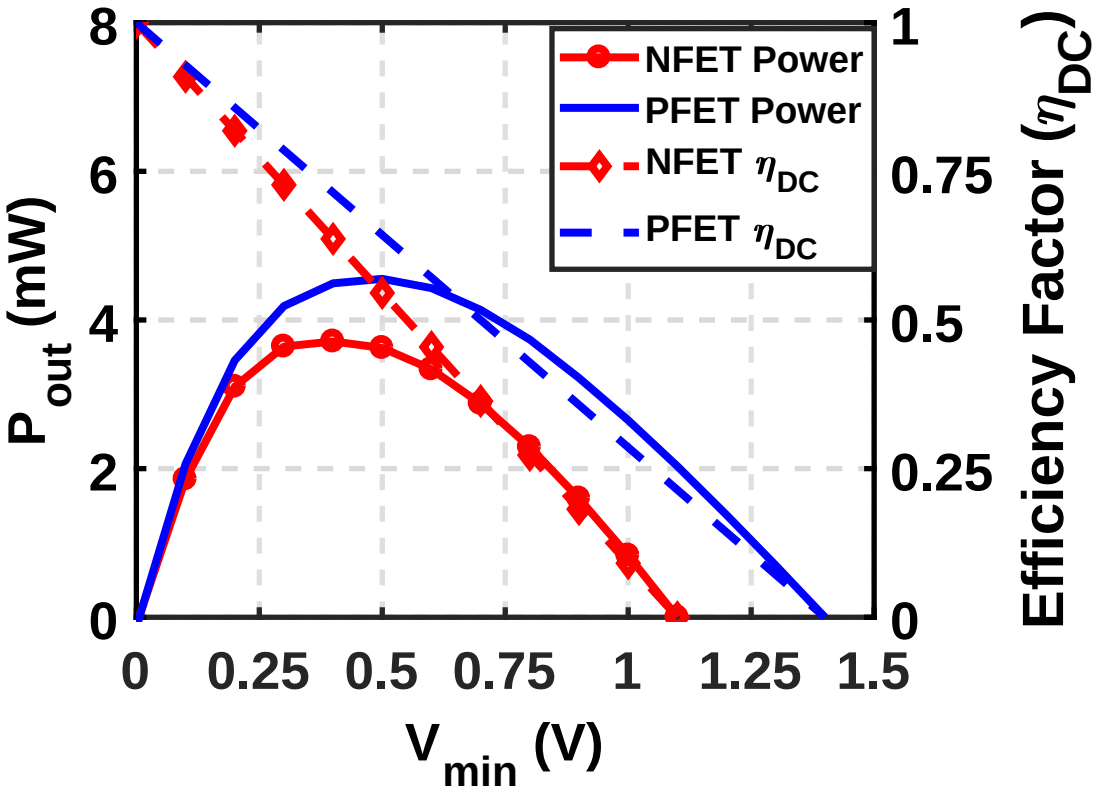

Figure 4.10: Output power and effective efficiency factor (η_{DC}) of an NMOS and PMOS amplifier biased in class-A with constant V_{DD} and varying load or V_{min}.

The maximum efficiency a PA can achieve with a FET of finite V_{knee} is $\eta_{max,class} \times \eta_{DC}$ where η_{DC} is the maximum efficiency of that class of PA with zero V_{knee} and $\eta_{DC} = 1 - V_{knee}/V_{DD}$. Given similar V_{knee} (0.4 V), higher V_{DD} (1.4 V for PFET vs. 1.1 V for NFET) allows PMOS PA to have higher η_{DC}, 72% for PFET compared to 64% for NFET.

The DC transconductances (G_m) derived from the ID-VGS measurements

for the NFET and PFET (W = 28.8 μm) for different values of $|VDS|$ (0 0.9 V) are shown in Fig. 4. 4.11. Both the devices achieve peak G_m at a current density of about 0.6 $mA/\mu m$. The maximum value of G_m for NFET is 56 mS (1.95 $mS/\mu m$) at VGS=0.56 V, VDS = 0.9 V ($I_{D,den}$ = 0.65 $mA/\mu m$) and for PFET is 48 mS (1.65 $mS/\mu m$) at $|VGS|$=0.64 V, $|VDS|$ = 0.9 V ($I_{D,den}$ = 0.58 $mA/\mu m$). The PFET maximum G_m is about 86% of that of the NFET.

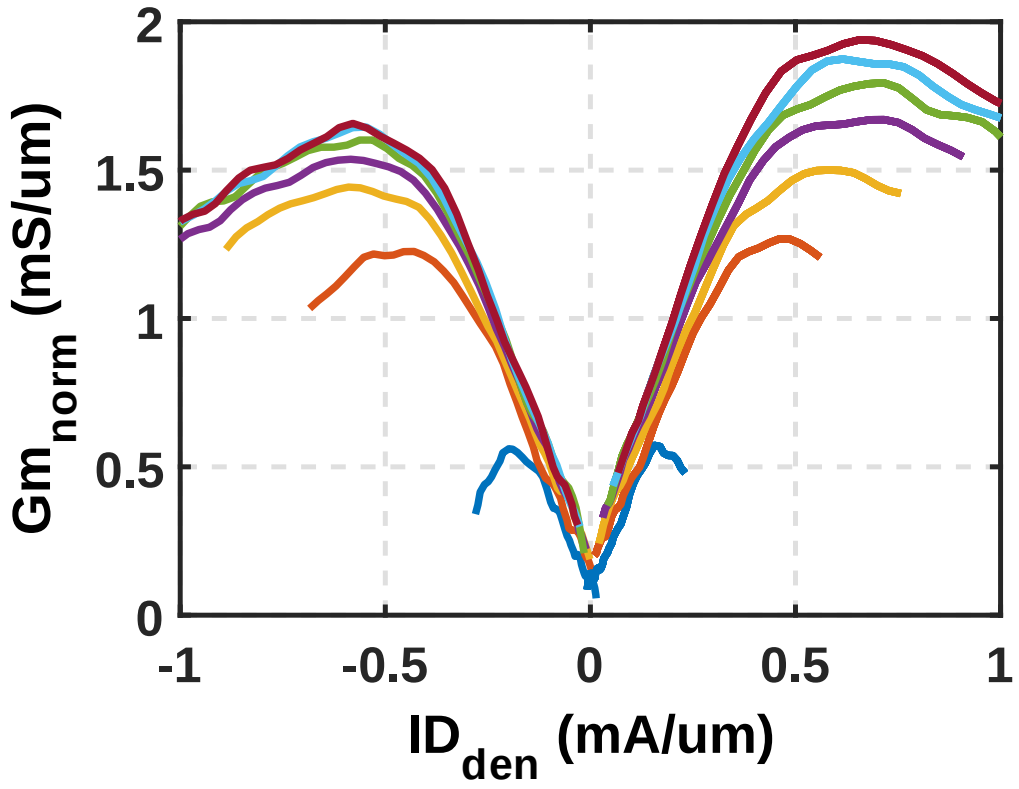

Figure 4.11: Measured DC transconductance (G_m) of NMOS and PMOS FETs (W = 28.8 μm) vs. current density (ID_{den}) for $|V_{DS}|$ = 0 - 0.9 V, $\Delta|V_{DS}|$ = 0.1 V.

The output conductance of NFET is considerably higher than for the PFET (corresponding to the results shown in the ID-VDS plot, where the PFET shows significant current saturation whereas the NFET shows only very weak saturation). The value of Intrinsic gain (analog gain) (G_m/G_{ds}) vs. drain current density for both NFET and PFET (W = 28.8 μm) for different values of $|VDS|$ (0.2 1.2 V)

is shown in Fig. 4. 4.12. At low drain voltages the NFET and PFET have similar gain. The NFET gain saturates around 0.9 V whereas the PFET gain continues to increase beyond that. The values of intrinsic gain at $I_{D,den}$=0.5 $mA/\mu m$ for different $|VDS|$ for both NFET and PFET is shown in the inset of Fig. 4. 4.12.

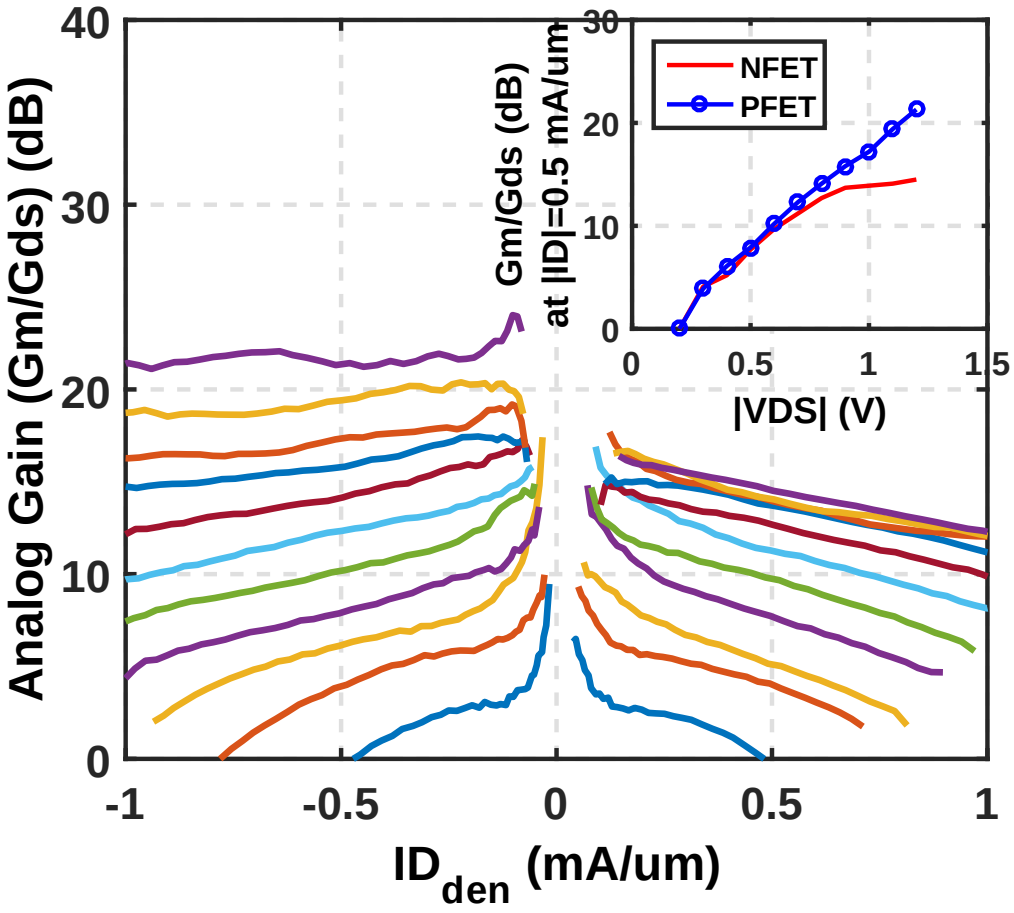

Figure 4.12: Measured intrinsic gain (G_m/G_{ds}) of NMOS and PMOS FETs (W = 28.8 μm) vs. current density (ID_{den}) for $|V_{DS}|$ = 0.2 - 1.2 V, $\Delta|V_{DS}|$ = 0.1 V. Gain at $|ID_{den}|$ = 0.5 mA / μm vs $|V_{DS}|$ for both NFET and PFET shown in inset.

4.2.2 AC Characteristics

NMOS and PMOS (W = 28.8 μm) S-Parameters measurements have been done to extract the equivalent small signal model of the devices. The data was collected at $|V_{DS}|$ = 1.1 V and $I_{D,den}$ = 0.6 $mA/\mu m$ which represents the highest transconductance point. In the measurement data the pads are de-embeded (open-short), moving the reference plane to the top metal layer (LB) of the FET wiring. Thus the measured data includes all the FET interconnect parasitics and represents the device used in the mm-wave amplifier design. The measurement data is verified against simulation. The simulation model contains the FET RC parasitic extacted with the transistor and bottom nine thin metal layer interconnects together. Along with that an EM simulated S-Parameter block for the top three thick metal layer interconnects is also added. S-parameters, MAG, H21 and equivalent circuit parameters ($R_g, g_m, g_{ds}, C_{gs}, C_{gd}, C_ds$) estimated from the Y-parameters [42, 71] are shown in the Fig. figs. 4.13 to 4.18. For mm-wave power amplifier design source resistance (R_S) is also important. But including R_S in the model explicitly makes the analysis complicated. Hence the effect of R_S is absorbed into g_m such that $g_{m,eff} = g_m(1 + g_m.R_S)$.

$$g_m = Re(Y_{21})\,|_{\omega\to 0} \tag{4.2}$$

$$g_{ds} = Re(Y_{22})\,|_{\omega\to 0} \tag{4.3}$$

$$C_{gg} = \frac{Im(Y_{11})}{\omega} \tag{4.4}$$

$$C_{gd} = \frac{|Im(Y_{12})|}{\omega} \tag{4.5}$$

$$C_{gs} = C_{gg} - C_{gd} \tag{4.6}$$

$$R_g = \frac{Re(Y_{11})}{Im(Y_{11})^2} \tag{4.7}$$

$$C_{ds} = \frac{|Im(Y_{22})|}{\omega} - C_{gd} * (1 + g_m\, R_g) \tag{4.8}$$

$$f_t = \frac{g_m}{2\,\pi\,(C_{gs} + C_{gd})} \tag{4.9}$$

$$f_{max} = \frac{f_t}{\left(2\,\sqrt{[(R_S + R_g)\,g_{ds} + 2\,\pi\,f_t\,R_g\,C_{gd}]}\right)} \tag{4.10}$$

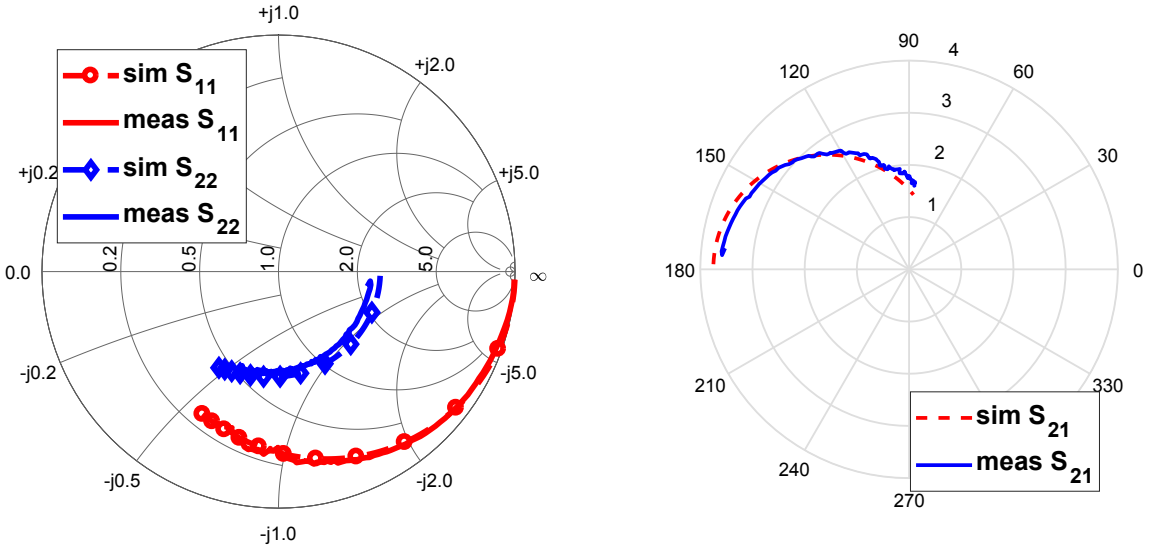

Figure 4.13: Measured and simulated S-parameters for 28.8 μm NFET - a) S_{11} and S_{22} (smith chart), b)S_{21} (polar plot).

The measured and simulated values of g_m, C_{gs}, C_{gd} and R_g agree very well. The measured value of g_{ds} is higher than simulated and while the measured value of C_{ds} is lower than simulated. The value of f_t can be estimated by extrapolation from the current gain ($|h_{21}|$) plot. However due to the noisy nature of the measured data (measured till 110 GHz), estimation of f_{max} by extrapolation the unilateral gain (U) is not attempted. Instead the value of f_{max} is calculated from the value of equivalent circuit parameters estimated.

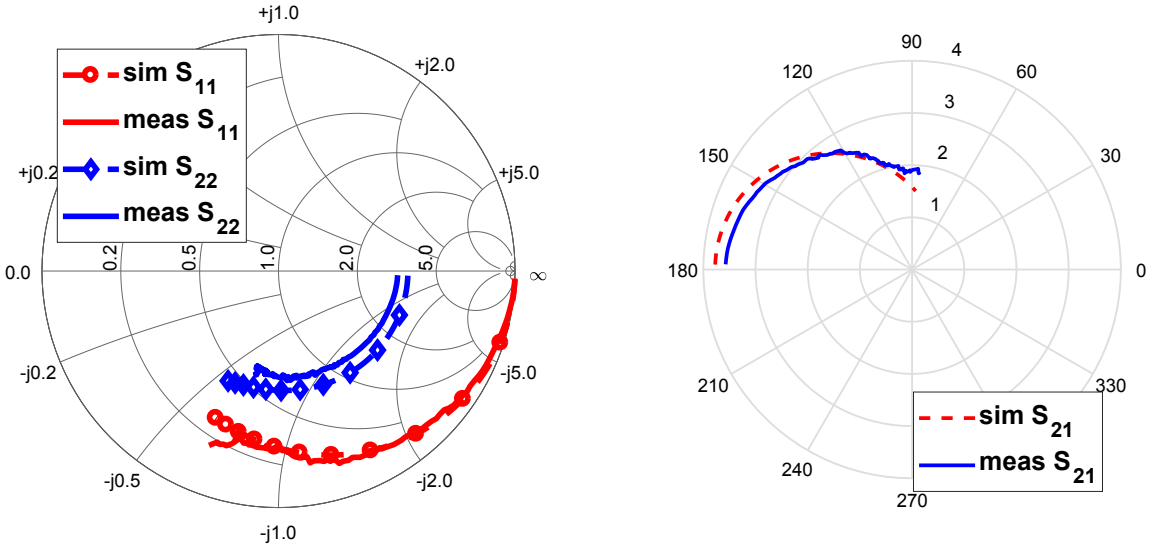

Figure 4.14: Measured and simulated S-parameters for 28.8 μm PFET - a) S_{11} and S_{22} (smith chart), b)S_{21} (polar plot).

Table 4.1 shows the comparison of equivalent circuit parameters estimated from extracted simulations for NFET and PFET. While $g_{m,PFET}$ is only 88% of $g_{m,NFET}$, due to lower values of C_{gs} and C_{gd} for PFET, the estimated value of f_t is similar for both devices ($\approx$ 285 GHz). The slightly lower value of C_{gs} and C_{gd} of PFET partially comes from the fact that the equivalent oxide thickness (T_{ox}) of PMOS device (1.55 nm) is higher than NMOS device (1.4 nm).

For long channel devices, since $R_g\, g_{ds} \ll 2\,\pi\, f_t\, R_g\, C_{gd}$, the analytical expression for f_{max}, ($\frac{f_t}{2\sqrt{[R_g\, g_{ds} + 2\,\pi\, f_t\, R_g\, C_{gd}]}}$ is usually simplified as $\sqrt{\frac{f_t}{8\,\pi\, R_g\, C_{gd}}}$. But as seen from the measurements above the value of g_{ds} is very high for deeply scaled devices. This is due to Channel Length Modulation and Drain Induced Barrier Lowering (DIBL) and is more pronounced for NFETs. For the 32 nm FETs measured here, the value of $R_g\, g_{ds}$ is nearly 44% of $2\,\pi\, f_t\, R_g\, C_{gd}$ (for NFET) and hence cannot be neglected. The value of R_g is similar for both NFET and PFET due to similar lithography. The advantage of higher g_m for NMOS is neutralized by its

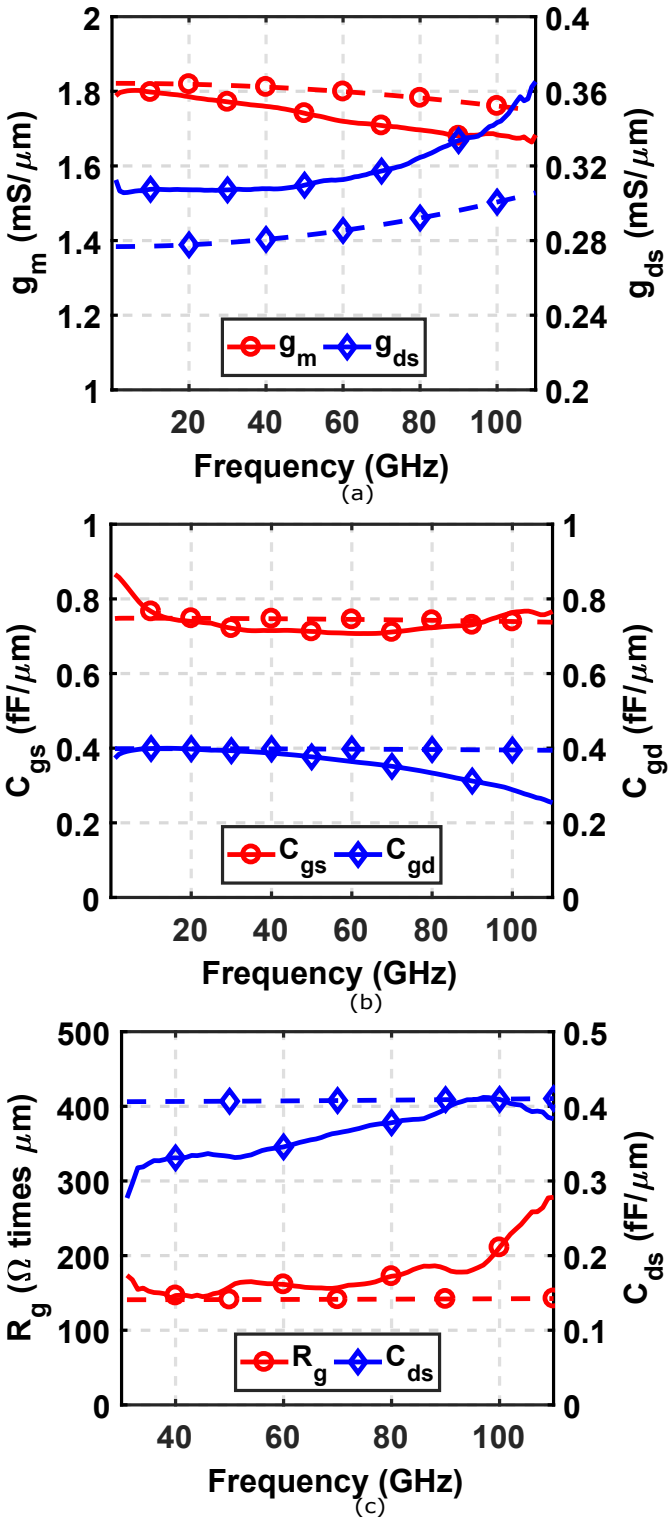

Figure 4.15: Equivalent circuit parameters estimated from measurement and simulation for the 28.8 μm NFET - a) g_m and g_{ds}, b) C_{gs} and C_{gd} and c) R_g and C_{ds} (Solid lines are measurement and dotted lines simulation).

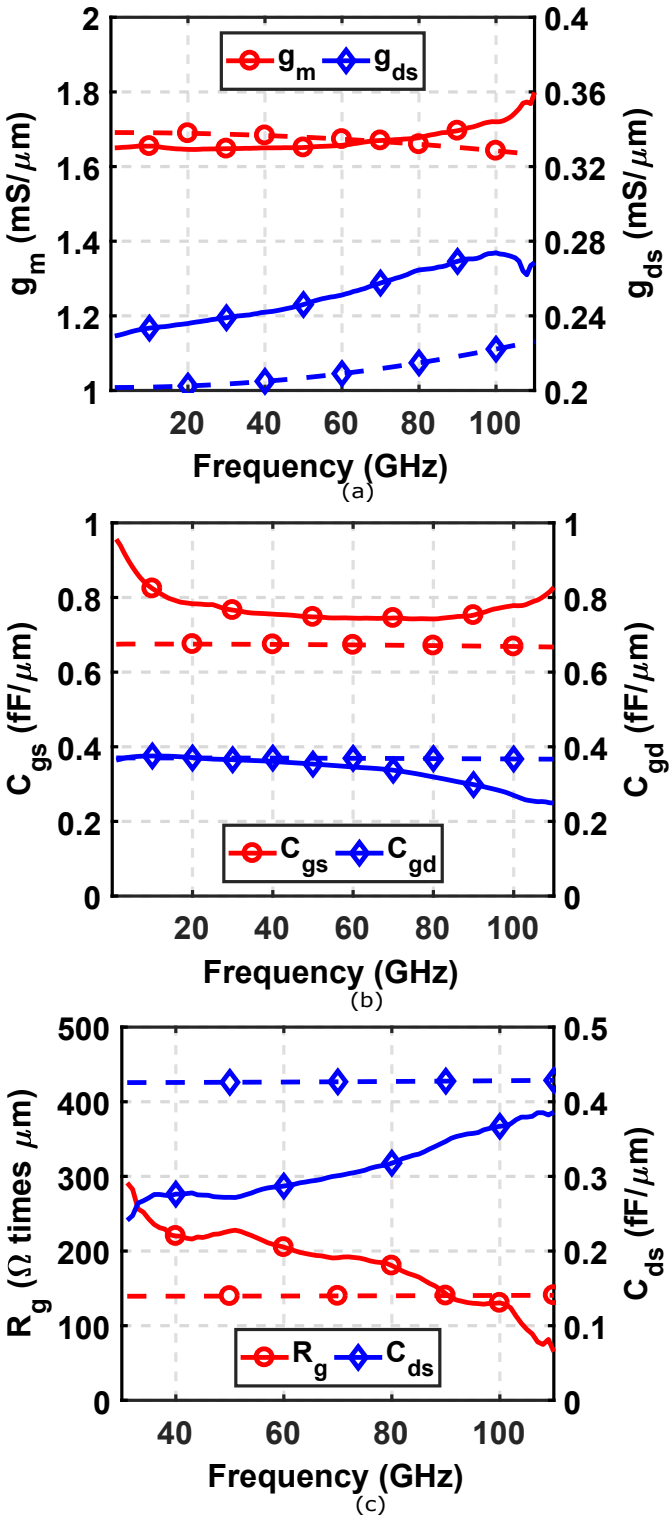

Figure 4.16: Equivalent circuit parameters estimated from measurement and simulation for the 28.8 μm PFET - a) g_m and g_{ds}, b) C_{gs} and C_{gd} and c) R_g and C_{ds} (Solid lines are measurement and dotted lines simulation).

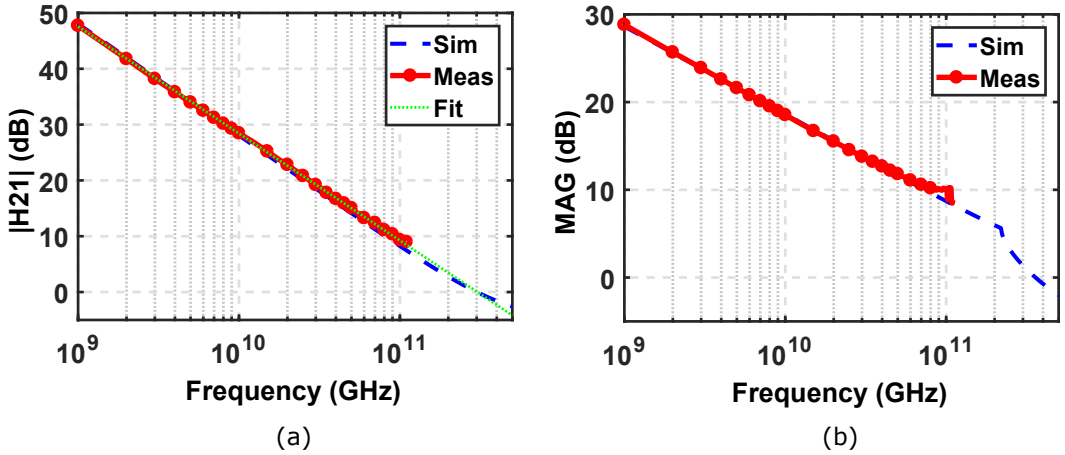

Figure 4.17: Measured and simulated gain for 28.8 μm NFET - a) short circuit current gain ($|h_{21}|$) and b) maximum available gain (MAG). (Solid lines are measurement and dotted lines are simulation. Thin dotted line in (a) is estimated linear fit for f_t calculation.)

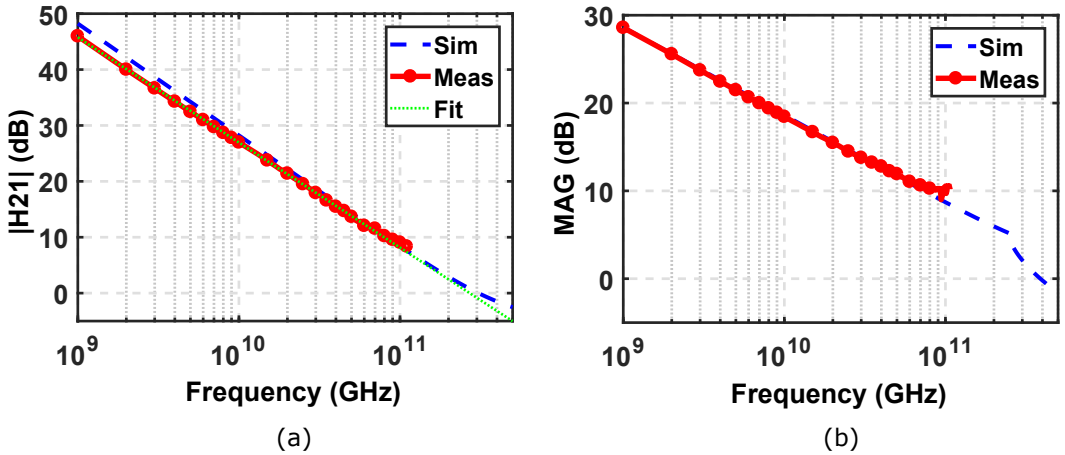

Figure 4.18: Measured and simulated gain for 28.8 μm PFET - a) short circuit current gain ($|h_{21}|$) and b) maximum available gain (MAG). (Solid lines are measurement and dotted lines are simulation. Thin dotted line in (a) is estimated linear fit for f_t calculation.)

Table 4.1: Estimated equivalent circuit parameters for NFET and PFEt (from simulation of extracted 28.8 μm FET)

Parameter	NFET	PFET
R_g ($\Omega.\mu m$)	141	139
g_m (mS / μm)	1.82	1.62
g_{ds} (mS / μm)	0.28	0.19
C_{gs} (fF / μm)	0.75	0.66
C_{gd} (fF / μm)	0.4	0.37
C_{ds} (fF / μm)	0.41	0.42
f_t (GHz)	290	280
f_{max} (GHz)	353	382

f_t is extrapolated from $|h_{21}|$ measurement, f_{max} is calculated from equivalent circuit parameters using 4.10

higher g_{ds} value, resulting in similar f_{max} values for NFET and PFET. In fact for this process the estimated value of f_{max} is higher for PFET than NFET.

4.3 PA Design

A single-stage three-stack PMOS PA was designed for E-band (60 GHz - 90 GHz). At these frequencies, stacking three FETs provides the optimum in terms of output power efficiency trade off [34]. This PA uses inter-stack shunt inductive tuning for efficiency improvement. Also two similar three-stack amplifiers without inter-stack tuning, are made using NMOS and PMOS as a control experiment to compare their performance. These are identical in layout except for the type of FET used.

For short channel MOSFETs reducing the gate resistance (R_g) is very critical in increasing the f_{max}. This requires double contacted gate fingers with sub-micron finger widths. At sub-50 nm processes the resistance of the gate poly outside the active device is a considerable fraction of the total gate resistance. This portion primarily comes from the DRC rules stipulating an exclusion dis-

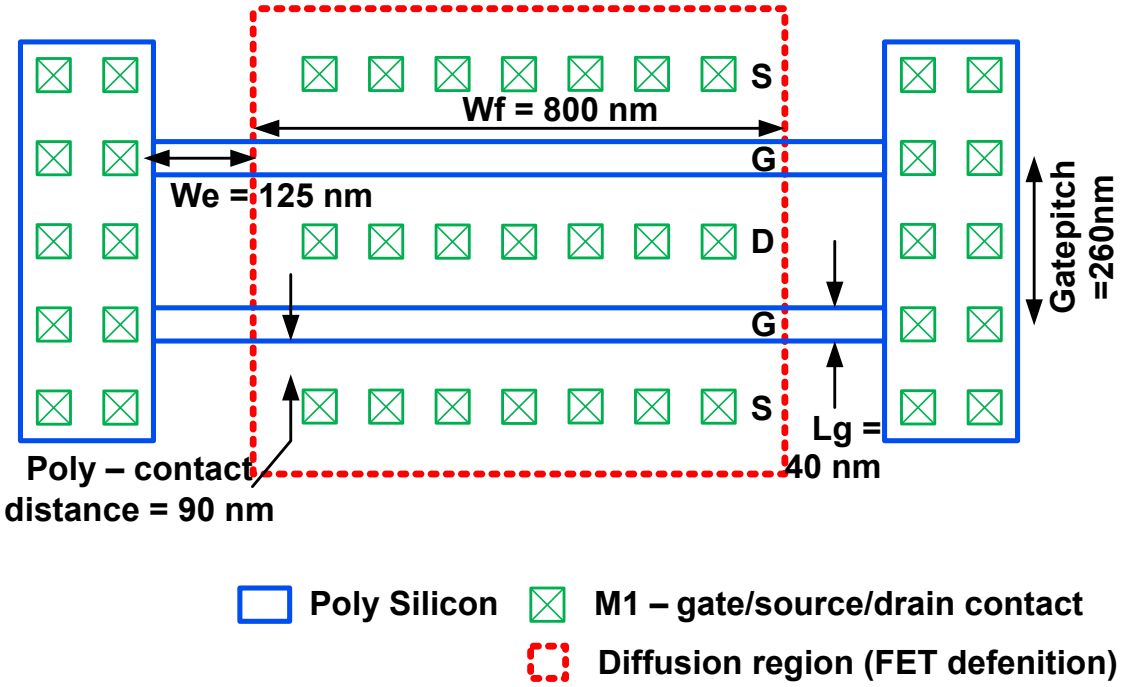

Figure 4.19: Layout of portion of double side gate contacted FET showing finger dimensions (only device layers and contacts shown, metal routings not shown).

tance (W_e) (usually this is roughly three times the gate length) between the active device and the closest via on the gate poly to low resistance metal routing. This resistance does not scale as we reduce the width of the fingers W_f.

$$R_{g_{finger}} = R_{ext} + R_{int} \tag{4.11}$$

$$R_{int} = \frac{1}{12} R_{sheet_{poly}} \left(\frac{W_f}{L_g} \right) \tag{4.12}$$

$$R_{ext} = \frac{1}{2} R_{ext\,one\,side} = \frac{1}{2} R_{sheet_{poly}} \left(\frac{W_e}{L_g} \right) \tag{4.13}$$

$$R_{gtot} = \frac{R_{g_{finger}}}{N_{finger}} = \frac{1}{2} \frac{R_{sheet_{poly}} W_f}{L_g W} \left(W_e + \frac{W_f}{6} \right) \tag{4.14}$$

From 4.14 we can see that for $W_f > 6\,W_e$, as the W_f is decreased the R_g drops as second power. But for $W_f < 6\,W_e$, R_g drops only linearly with W_f. Since

the reduction of W_f amounts to increase in number of fingers ($N_{finger} = \frac{W}{W_f}$), the routing capacitive parasitics, both C_{gs} and C_{gd}, increase nearly linearly. Therefore reducing W_f below $6\,W_e$ is not useful in increasing the f_{max}. For the 32 nm SOI process the $L_{g,drawn}$ is 40 nm and the exclusion distance is about 125 nm (Fig. 4.19). Therefore a gate width of 800 nm should give the maximum f_{max}. This result is also supported by an empirical experiment result in 45 nm SOI process published in [40]. 45 nm SOI has the same $L_{g,drawn}$ and similar lithography. In [40] f_{max} measurements of different gate widths (with single side gate contact) are provided, with $W_f = 400$ nm having the maximum f_{max}. 400 nm single side contact is equivalent to 800 nm double side contact. Also it should be noted that the above mentioned trade-off of $R_g - C_{gd}$ can be broken if we use differential design with C_{gd} neutralization. Also for nanoscale devices the vertical gate stack resistance is also significant, but since it is invariable with device layout (finger width) it is not usually considered while layout optimization.

The maximum pitch between the gate finger strips allowed in this process (260 nm) was chosen for the design. The wider pitch increases the gate-to-(drain/source) contact spacing. This leads to lower gate parasitic capacitances (C_{gd} and C_{gs}). Also relaxed pitch results in higher transconductance (g_m) due to enhanced stress response. Both these effects result in a higher value of f_t [39]. The impact in f_{max} is limited as the effect of increased g_m is negated by the increase in overlap capacitance and gate routing resistance for the double pitch layout. Multiple instances of the unit FET (0.8 $\mu m\times$ 36 fingers) of 28.8 μm were used to build the final device.

The device widths were chosen to have nearly 50 Ω optimum real part of output impedance with the three transistor stack. This avoids the need to have additional impedance transformation at the output and hence allows to have high

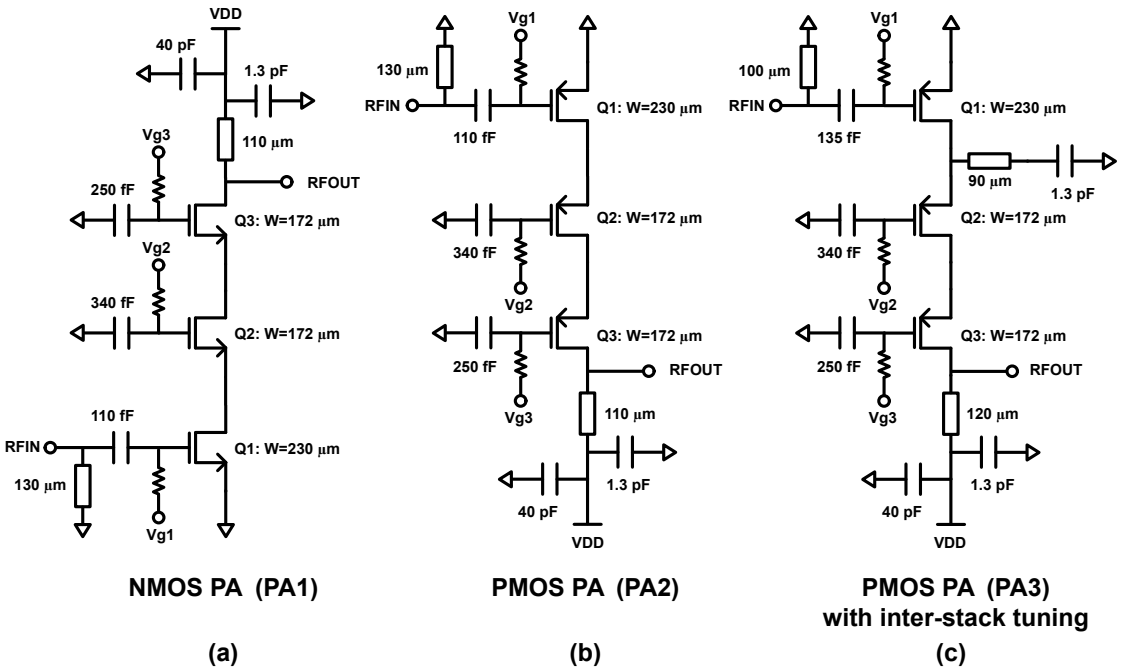

Figure 4.20: Schematic for (a) NMOS PA (PA1), (b) PMOS PA (PA2) and (c) PMOS PA with inter-stack tuning (PA3).

efficiency and bandwidth. Same size devices and matching elements were used for both NMOS PA (PA1) and PMOS PA (PA2) (Fig. 4.20). F_{max} simulations of extracted NMOS and PMOS FETs of 230 μm width with identical layout showed similar f_{max} values for both devices (370 GHz for PMOS, 340 GHz for NMOS). To avoid differences from non-FET elements, exactly same layout was used for both PAs. Both PAs have their sources of the bottom common source FET Q1 connected to ground plane and drains of the top stacked common gate FET Q3 connected to output pads. The NMOS PA uses positive bias voltages and the PMOS PA uses negative voltages. For simplicity only the absolute values of voltages are mentioned in this chapter. PA3 is similar to PA2 but has an additional inter-stack shunt inductance tuning at the drain node of the bottom FET (common source FET). This improves the phase alignment of voltages along the stack and hence enhances the PAE. All the inductances used for matching were implemented using grounded

coplanar waveguides (GCPW). The PAs were designed with about 0.15 mA/μm of quiescent current for maximum linearity and PAE. This condition achieved a small signal gain (S_{21}) within 1 dB of maximum S_{21} achievable across all bias condition. The bias voltages ($|V_{G1}|$ = 0.3 V, $|V_{G2}|$ = 1.6 V, $|V_{G3}|$ = 2.6 V, $|V_{DD}|$ = 3.5 V) are set so that the DC and RF voltage swings are equally distributed between the three FETs in the stack at maximum output power condition.

4.4 Experimental Results

All the three amplifiers (PA1, PA2 and PA3) have similar layout and occupy 440 $\mu m\times$ 280 μm area (0.05 mm^2 excluding the pads) (Fig. 4.21).

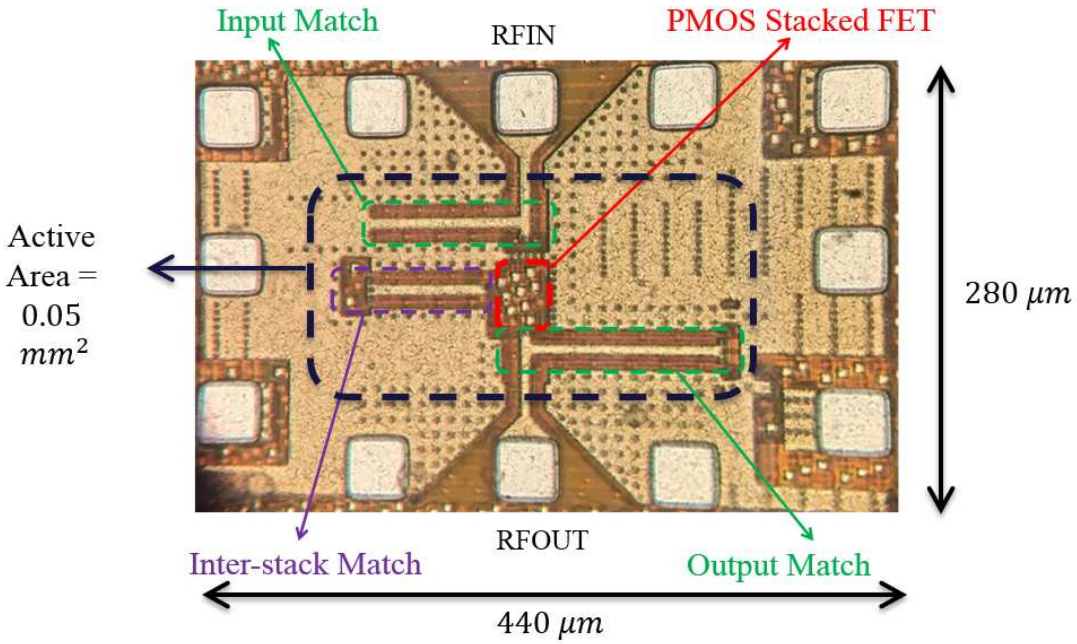

Figure 4.21: Die micro-photograph of 3-stack PMOS PA (PA3) with shunt tuning.

4.4.1 Small Signal Measurements

The PMOS PA with inter-stack tuning (PA3) achieved a maximum S_{21} of 11.6 dB at 75 GHz and a 3-dB gain bandwidth of 27 GHz (35% fractional bandwidth, from 65 to 92 GHz) (Fig. 4.22). Both NMOS PA (PA1) and PMOS PA (PA2) without the inter-stack tuning achieved similar peak S_{21} (> 10 dB) (Fig. 4.23). The shunt inter-stack tuning helps the PA3 to have more than 1 dB higher gain than PA2. The measured maximum gain is lower than simulated by 2 dB and the measured 3-dB bandwidth is greater than simulated by 6 GHz. This could be possibly because of lower than simulated quality factor of passive matching elements used. All the S-parameter measurements are done with quiescent current of 50 mA (0.22 mA/μm) and 3.6 V supply (V_{DS} of 1.2 V per FET) for NMOS and PMOS amplifiers.

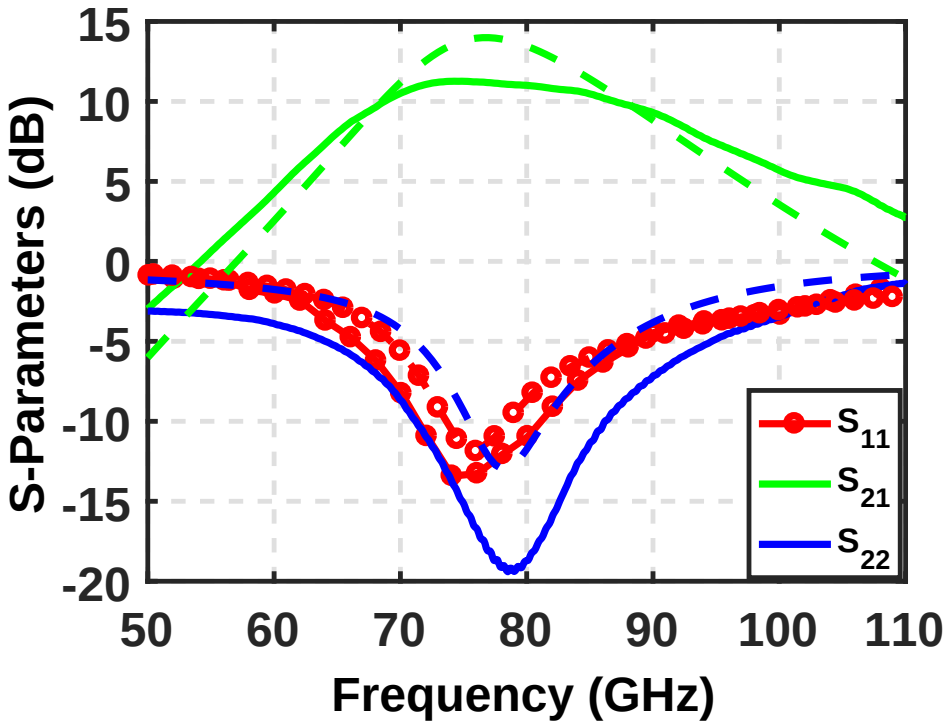

Figure 4.22: Measured (solid line) and simulated (dotted line) S-parameters of 3-stack PFET PA with inter-stack tuning (PA3) ($|V_{G1}| = 0.35$ V, $|V_{DD}| = 3.6$ V.

4.4.2 Large Signal Measurements

For large signal measurements, the PA (PA3) was operated at both low-bias ($I_{D,Quiescent} = 0.18$ mA/μm) and high-bias ($I_{D,Quiescent} = 0.30$ mA/μm) conditions (Fig. 4.24). The low-bias operation achieved maximum output power of 18.7 dBm and peak PAE of 24% at 78 GHz. The high-bias operation achieved maximum output power of 19.6 dBm and peak PAE of 18% at 78 GHz. Large signal frequency sweep measurements showed that the PA achieved more than 20% PAE from below 75 GHz to 81 GHz (Fig. 4.25). With high bias, the PA achieved more than 18 dBm P_{sat} from 70 GHz to 90 GHz. The PA was tested at different drain supply voltages at 78 GHz (Fig. 4.26). The maximum PAE is obtained at V_{DD} = 3.6 V (V_{DS} of 1.2 V per FET) and the maximum output power is obtained at V_{DD} = 4.5 V (V_{DS} of 1.5 V per FET). The measured maximum output power is 1 dB lower than simulation whereas the measured maximum PAE is same as that of simulation. The measured P_{1dB} is 16 dBm in high bias condition and 13 dBm in low bias condition.

Without inter-stack tuning, at 78 GHz, NMOS PA (PA1) achieved a maximum PAE of 15% at 16 dBm and PMOS PA (PA2) achieved 22% at 17 dBm. They respectively achieved a maximum output power of 17 dBm and 19 dBm with high bias (Fig. 4.27 and Fig. 4.28). Under same bias conditions, the NMOS and PMOS PA have similar output power, but PMOS PA has higher PAE.

4.4.3 Reliability Measurements

To check the reliability of the PMOS amplifier, PA3 was operated at an increased supply voltage of 5.1 V for more than 24 hours continuously at peak output power of 19.6 dBm and the output power was measured to be stable within

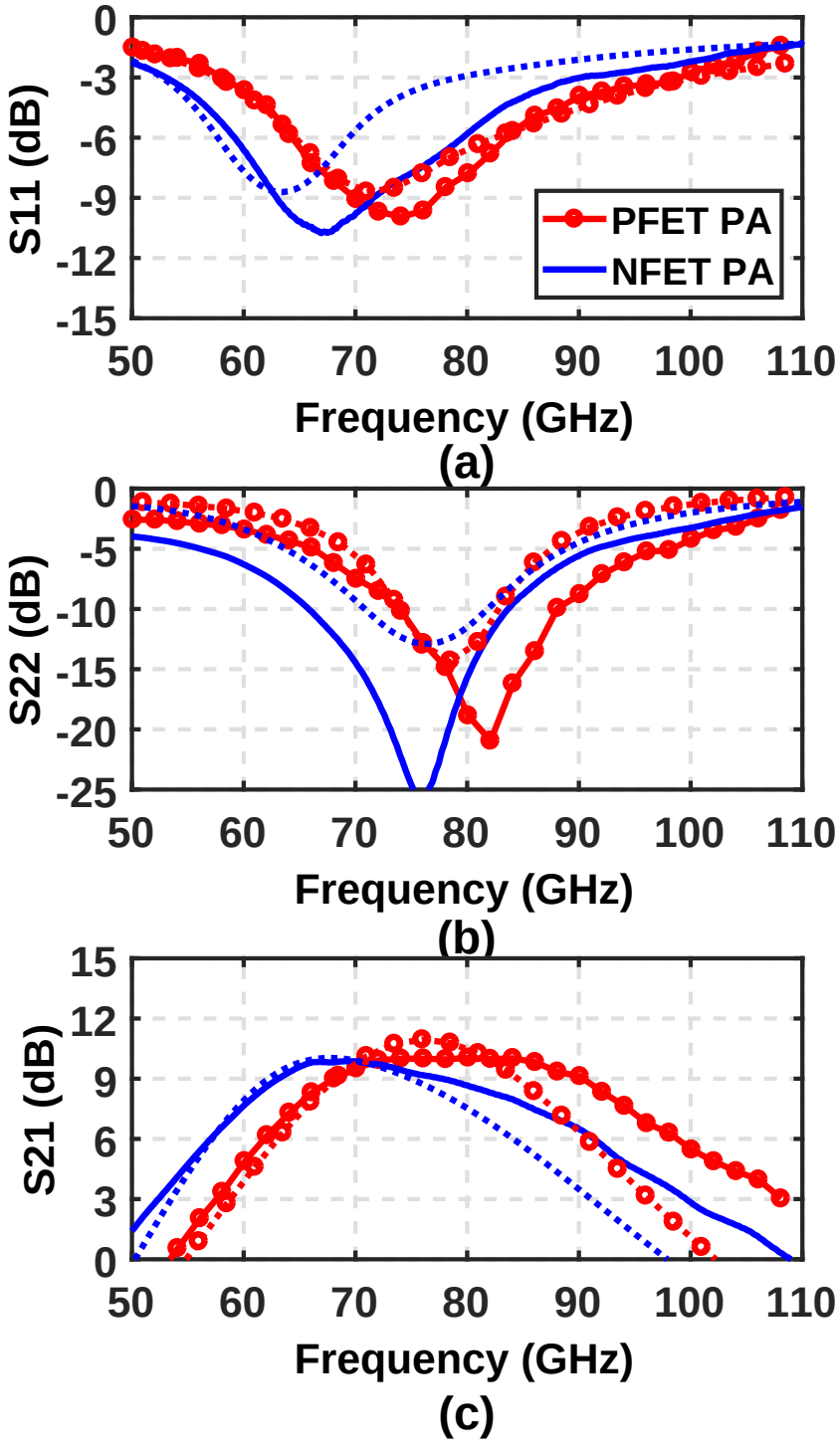

Figure 4.23: Measured (solid line) and simulated (dotted line) S-parameters of 3-stack NFET PA (PA1) and PFET PA (PA2) - a) S_{11}, b) S_{22} and c) S_{21}.

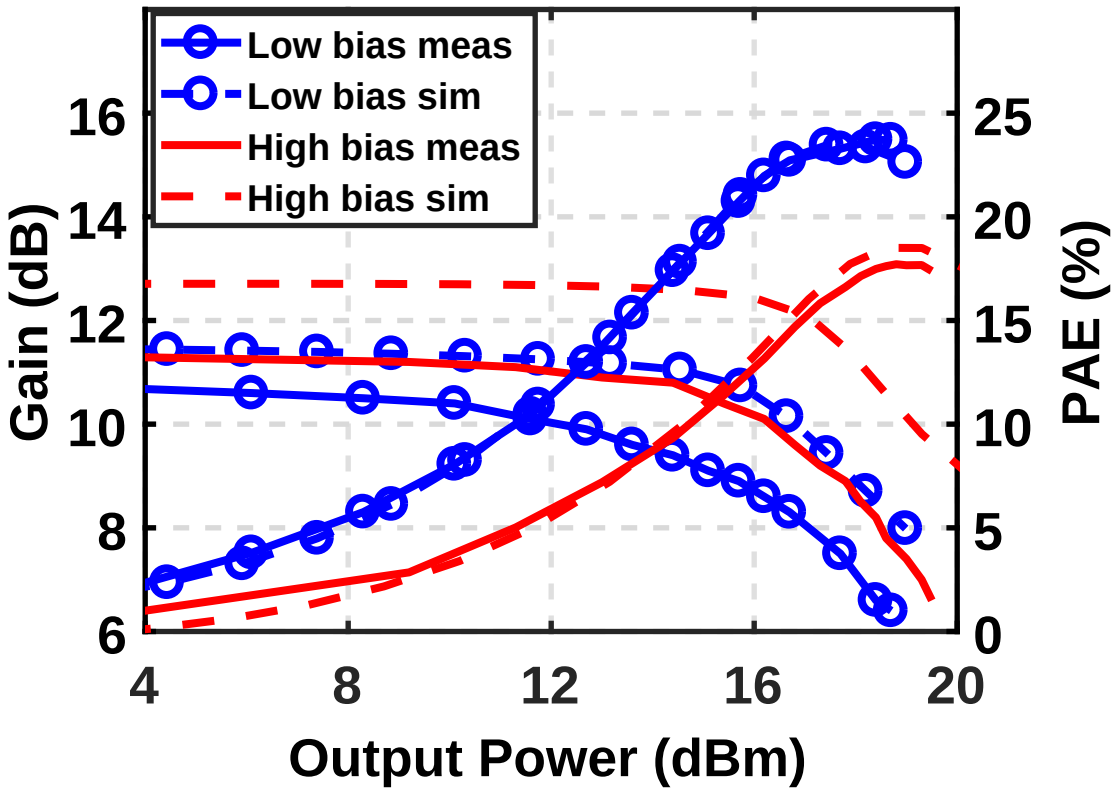

Figure 4.24: Measured and simulated Gain and PAE vs. P_{out} at 78 GHz for PA3 with low and high bias (Low bias : $|V_{G1}| = 0.35$ V, $|V_{DD}| = 3.6$ V, High bias : $|V_{G1}| = 0.4$ V, $|V_{DD}| = 4.5$ V.

+/- 0.1 dB on a probe station. This is within the resolution of the measurement setup for the extended time period (> 24 hours) given the inaccuracies accrued by oxidation of the probe tips, vibrations of the measurement setup and other environmental changes. Assuming 4.1 and the values of n (0.33) and c (8.824 V^{-1}) as given by the process manual, the degradation in current due to HCI for a 24 hour measurement is equivalent to a 10 year measurement at a $|V_{DS}|$ lower by $\frac{n}{c}\, ln\, \left(\frac{10\, x\, 365}{1}\right) = 0.3$ V. Thus the short term measurement with $|V_{DD}| = 5.1$ V ($|V_{DSperFET}| = 1.7$ V) is equivalent to long term use at $|V_{DD}| = 4.2$ V ($|V_{DSperFET}|$ = 1.4 V).

Both DC and AC measurements were conducted to study the reliability

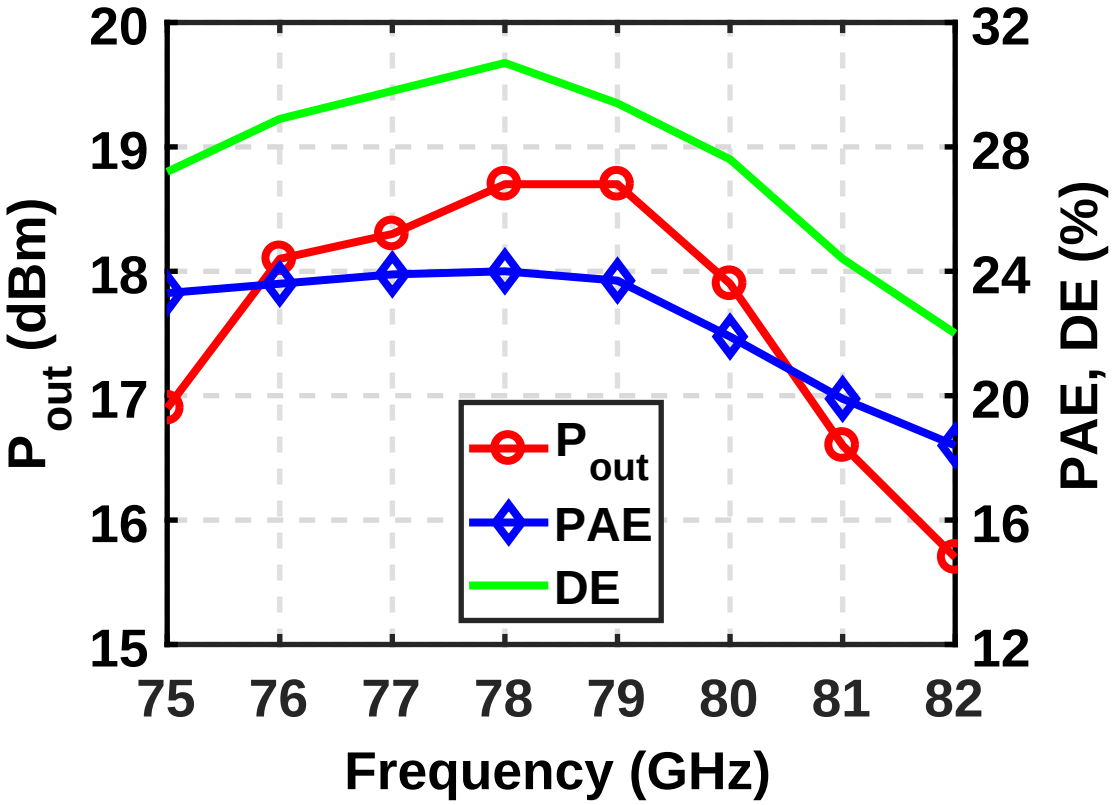

Figure 4.25: Measured maximum P_{out}, PAE and DE vs. frequency at low bias for PA3.

of the NMOS and PMOS. ID-VDS measurements were done on unmatched 28.8 μm and 230 μm NFET and PFET. To differentiate between the degradation due to thermal effects caused by higher power dissipation and the degradation due to voltage stress, the FETs were measured at different drain-dource voltages with gate bias adjuested to have appropriate currents so that the $V_{DS} \times I_D$ is constant. The results showed that NFET and PFET can withstand 1.8 V and 2.1 V of drain-source voltage respectively without irreversible breakdown. The breakdown voltages varied by 0.2 V between multiple samples. Since neither 28.8 μm nor 230 μm devices are matched to 50 Ω, RF reliability measurements on them would be difficult to interpret due to the presence of reflected power. Instead the three-

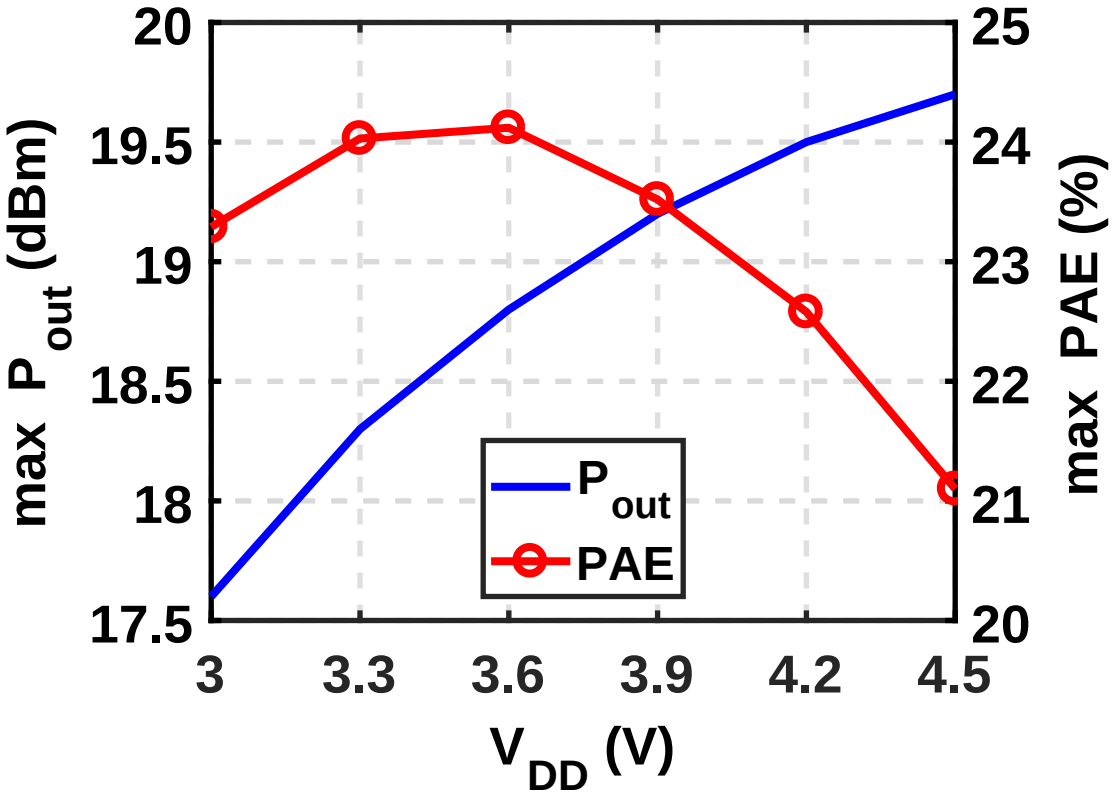

Figure 4.26: Measured maximum P_{out} and PAE vs. V_{DD} at 78 GHz for PA3.

stack PAs (PA1 and PA2) are used for RF measurements. Both the PAs were provided with maximum input power (10 dBm) and tested with different drain supply voltages. Three chips of each flavor (NMOS and PMOS PA) were tested to ensure repeatability and the worst case PMOS and best case NMOS are compared (Fig. 4.29). The PMOS PA (PA2) output power is stable within measurement resolution (0.01 dB) till 4.8 V of drain supply voltage across three-stack ($\approx$ 1.6 V per device). At this point the PA outputted 18.7 dBm of output power. As the supply voltage was raised above 5.1 V (1.7 V per device) the output power started varying slowly with time suggesting FET degradation. Similarly the NMOS PA (PA1) output power was stable till 4.2 V of supply voltage (1.4 V per device) and output power of 16.9 dBm. Above 4.5 V of supply the output power of NMOS PA

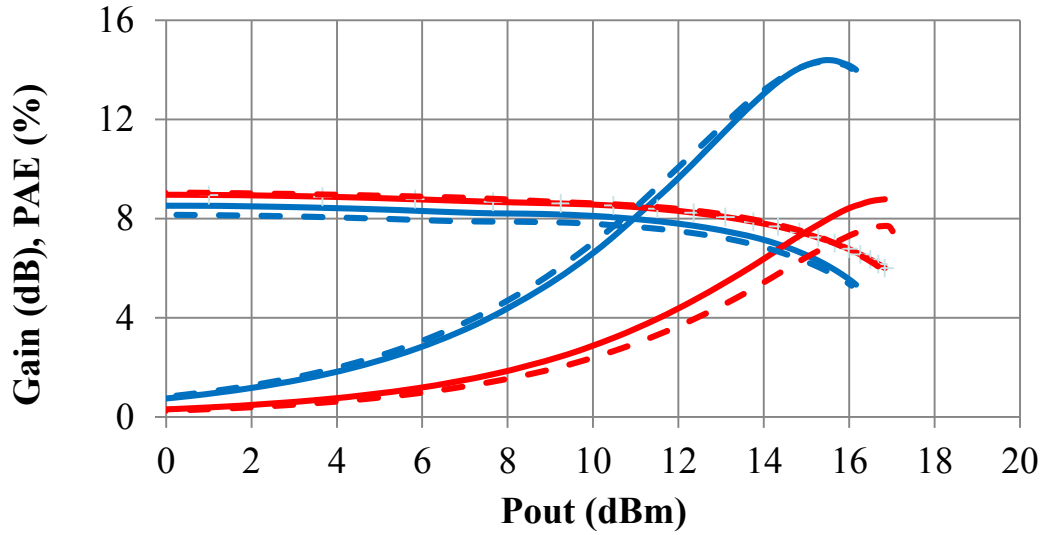

Figure 4.27: Measured Gain and PAE at 78 GHz for two samples of 3-stack NFET PA with low and high bias (Low bias : $|V_{G1}| = 0.3$ V, $|V_{DD}| = 3.0$ V, High bias : $|V_{G1}| = 0.4$ V, $|V_{DD}| = 3.6$ V.

decreased. Since the supply voltage and hence the DC power for the breakdown condition in RF measurement is lower than that of DC measurement it can be inferred that the breakdown is voltage breakdown and not thermal. Also this result confirms the calculated reliable V_{DD} values from the aging calculations in section II. Under RF swing the drainsource voltage can be multiple times the drain-source DC supply voltage. Since it is difficult to effectively realize higher harmonic tuning with high quality factor at mm-wave frequencies sharper drain voltage waveforms with more than two times supply voltage swing are not usually created. Also at higher frequencies, the time the FET suffers voltage stress is less in each cycle. Because of these two reasons higher frequency mm-wave PA can have slightly higher supply voltages than low frequency PA and still be reliable.

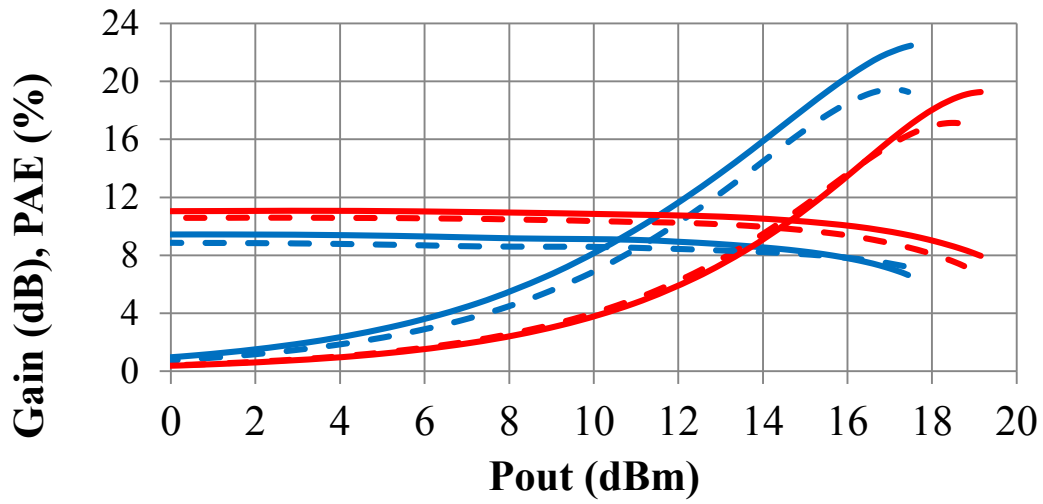

Figure 4.28: Measured Gain and PAE at 78 GHz for two samples of 3-stack PFET PA with low and high bias (Low bias : $|V_{G1}| = 0.3$ V, $|V_{DD}| = 3.6$ V, High bias : $|V_{G1}| = 0.4$ V, $|V_{DD}| = 4.2$ V.

4.5 Comparison with State-of-Art

Table 4.2 provides a comparison with other silicon based mm-wave power amplifiers in E- and V- bands. Most of the amplifiers of comparable output power levels are massively (up to 16-way) on-chip power combined. The PMOS stacked FET PA achieves better efficiency than the NMOS and SiGe PA at the same frequency range (Fig. 4.30).

4.6 Conclusion

For deeply scaled CMOS process nodes, compared to NFET, PFET usually has slightly lower I_{ON} and gain efficiency $\left(= \frac{g_m}{I_D}\right)$, similar analog gain $\left(= \frac{g_m}{g_{ds}}\right)$, $P_{max}\left(= \frac{(V_{DD}-V_{knee})\, I_{D,knee}}{4}\right)$ and $f_{max}\left(= \frac{f_t}{2\sqrt{R_g\, g_{ds}+2\,\pi\, f_t\, R_g\, C_{gd}}}\right)$ and higher $\eta_{DC}\left(= 1 - \frac{V_{knee}}{V_{DD}}\right)$ and lifetime. This is primarily due to the mobility enhancement for PFET by strain engineering and lower effective mass orientation and higher breakdown voltage.

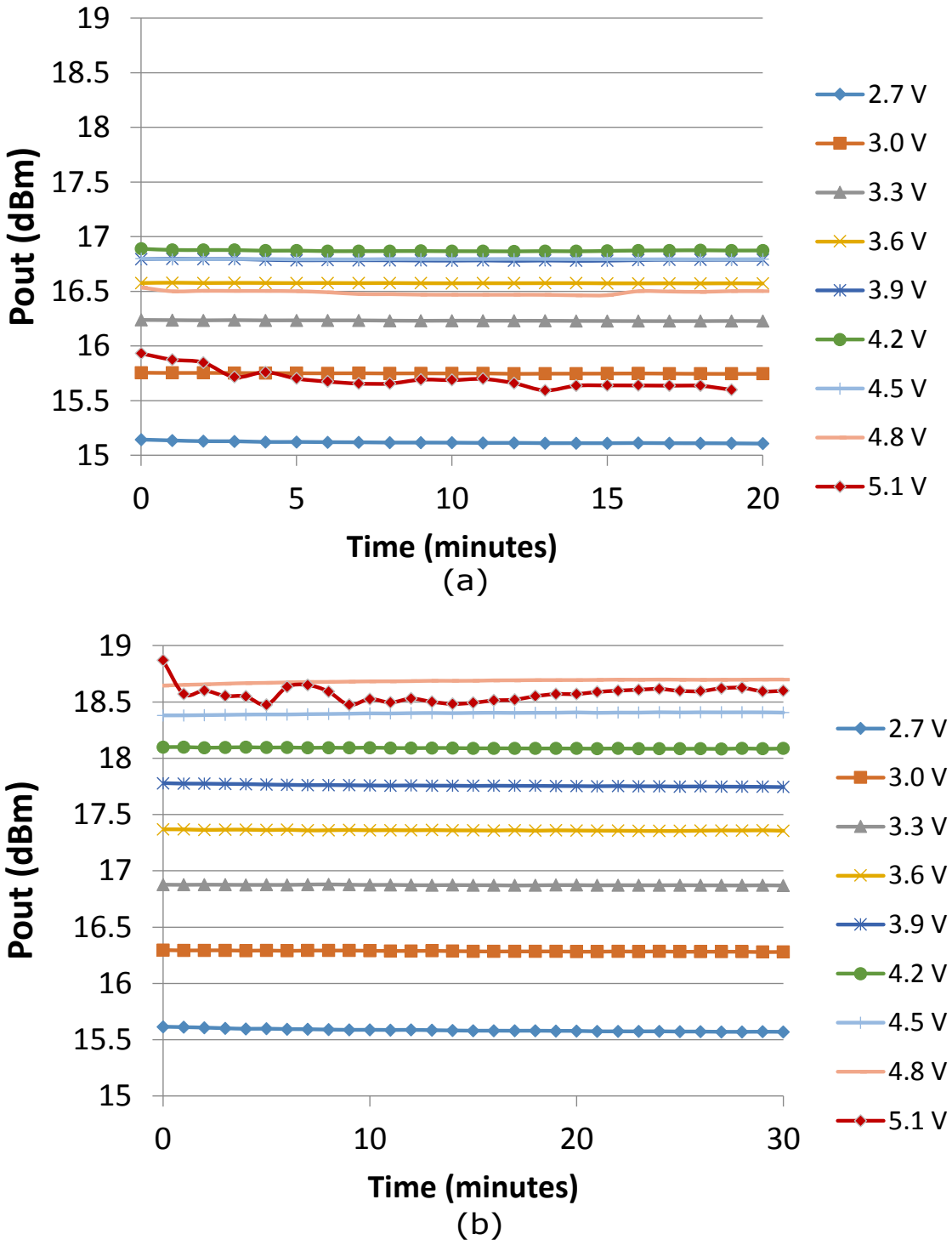

Figure 4.29: Output power vs. time at full power with different supply voltage for a) 3-stack NFET PA (PA1) and b) 3-stack PFET PA (PA2).

Table 4.2: Comparison to previously reported Silicon high power PA in V- and E-band

Ref.	**Tech.**	**Design**	**Freq. (GHz)**	P_{sat} **(dBm)**	**Peak PAE (%)**	**Gain (dB)**	**Chip area** (mm^2)
This work	32 nm SOI CMOS	PMOS, 3-stack, class A	78	19.6	18	11.6	0.12
This work	32 nm SOI CMOS	PMOS, 3-stack, class AB	78	18.7	24	11.2	0.12
[9]	SiGe	16-way power combined	76	27.3	12.4	19.3	6.5
[14]	40 nm CMOS	8-way power combined	80	20.9	22.3	18.1	0.4
[72]	65 nm CMOS	Cascode, 4-way power combined	77	15.8	15.2	20.9	0.4
[13]	SiGe	4-way power combined	62	20.1	18	20.6	0.7
[73]	28 nm UTBB FDSOI	8-way power combined	60	18.8	21	15.4	0.25
[62]	40 nm CMOS	Push-pull, 4-way power combined	60	16.4	23	22.4	0.4
[24]	90 nm SiGe	3-stack HBT	83	23.3	17.1	18.7	1.95

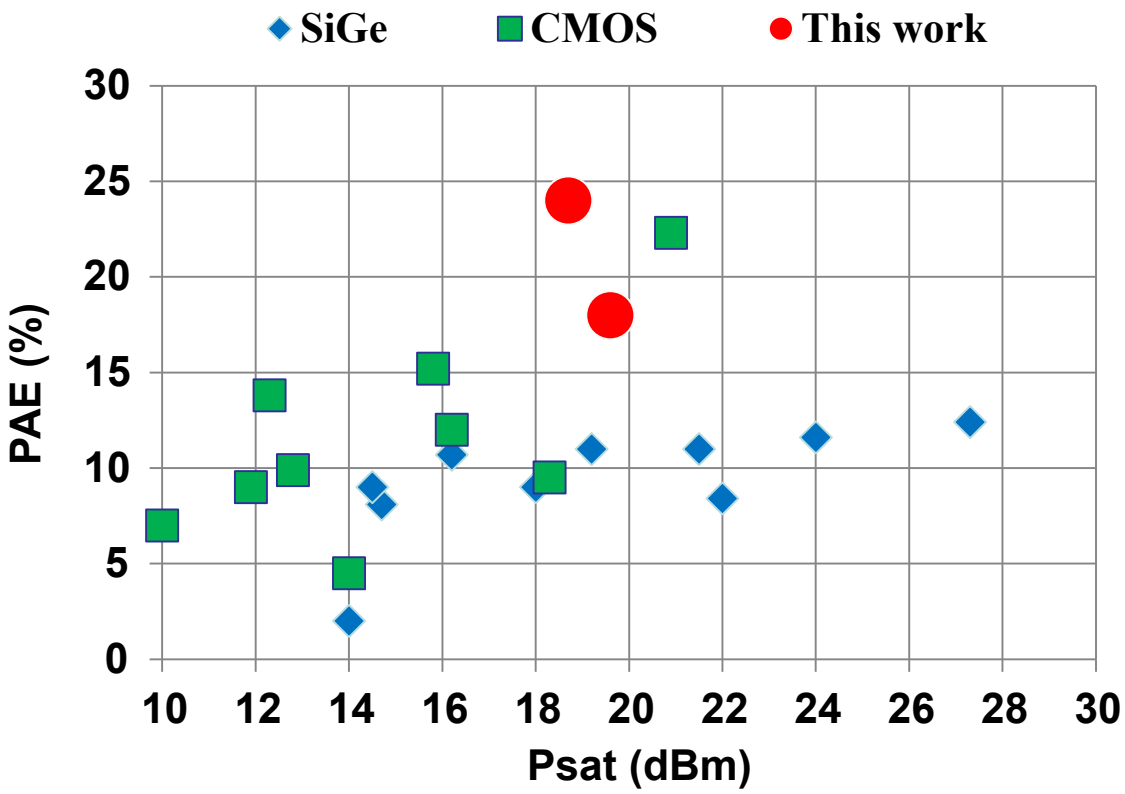

Figure 4.30: P_{sat}, PAE of state-of-art E-band Silicon PA.

We have demonstrated that at very short channel lengths PFETs - instead of conventional NFETs - are promising as a reliable technology for high power, high efficiency, compact mm-wave PAs. We report a single stage, 3-stack PMOS PA with 11.6 dB of gain and 27 GHz (65-92 GHz) 3-dB bandwidth. The PA achieved 19.6 dBm output power and 24% PAE at 78 GHz and occupies only 0.05 mm^2 area (excluding pads) on a 32 nm CMOS SOI process.

Acknowledgment

Chapter 4 is mostly based of of materials used in the following publications

The material in preparation to be submitted to *IEEE Journal of Solid-State Circuits,* J. A. Jayamon, J. F. Buckwalter, and P. M. Asbeck, "Millimeter-wave PMOS Power Amplifier". The dissertation author was the primary investigator and author of this material, and co-authors have approved the use of the material for this dissertation.

The material as it appears in J. A. Jayamon, J. F. Buckwalter, and P. M. Asbeck, "A PMOS mm-wave Power Amplifier at 77 GHz with 90 mw Output Power and 24% Effiency," in *2016 IEEE Radio Frequency Integrated Circuits Symposium (RFIC),* May 2016. The dissertation author was the primary investigator and author of this material, and co-authors have approved the use of the material for this dissertation.

Chapter 5

Conclusions and Future Works

5.1 Dissertation Summary

Highly integrated transceiver chipsets are needed to implement mm-wave wireless applications like radar and high data rate communication links including emerging 5G wireless standards. Deeply scaled CMOS nodes offer fast transistors for mm-wave applications as well as dense digital back end for control, calibration and performance enhancement. High power, high efficiency, broadband, linear, reliable power amplifiers are needed at the front-end of these transmitters. But the non-idealities associated with the nano-scale CMOS devices and their low voltage handling capability make the design of high power mm-wave PA challenging. This thesis demonstrates the viability of FET stacking in CMOS SOI to implement high power PA extended to frequencies as high as 94 GHz. Various other power combining techniques including spatial power combining are used to increase the output power further. Also, design of reliable and efficient PA using high voltage devices (PMOS) available in standard CMOS process and implementation of layout techniques for performance and lifetime enhancement are studied and demonstrated. A

W-band 3-stack PA design in 45 nm CMOS SOI process is presented in chapter 2. This PA achieved a measured output power of 17 dBm and PAE of 9% at 90 GHz. This design increased the frequency of operation of stacked FET from previously demonstrated 45 GHz to 94 GHz and represented twice the previously reported highest power at 94 GHz. Stacking three FETs helps to increase the voltage swing three-fold and also increases the output impedance to nearly 50 Ω. This avoids any additional impedance transformation matching networks at the output and hence allows broadband and efficient operation of the PA. Multiple channels of a similar PA implemented as a pseudo-differential amplifier are made into a 2 x 4 array. The CMOS chip along with a differential microstrip antenna array, deposited on a quartz wafer placed on top of the CMOS chip, with the antennas electromagnetically coupled to the PA output, is used to implement a spatial power combined PA-antenna array at 94 GHz. The array achieved a measured EIRP of 33 dBm and estimated output power of 24 dBm. This represents the highest output power from a CMOS chip at 94 GHz. The CMOS chip has the signal distribution networks (including Wilkinson dividers, baluns and grounded CPW transmission lines), driver amplifiers (twelve-stage) and on-chip differential antenna feed in addition to the PAs. The radiation measurements of the PA-antenna array with modulated signals of 256 QAM (375 MS/s - 3 Gbps) with digital pre-distortion are demonstrated. A multigate-cell style distributed layout of stacked FET PA with superior thermal dissipation properties is presented in Chapter 3. This PA implemented in 45 nm CMOS SOI process achieved a measured output power of 24.8 dBm and 29% at 28 GHz. This represents the highest reported power from a CMOS PA at 28 GHz without use of elaborate power combining schemes. The PA achieved this power level while occupying a chip area of less than 0.1 mm^2. This translates to 300 mW at 3 W/mm^2 power density and highlights the merit of stacked FET design as

the most area efficient power combining scheme. The distributed layout technique makes the PA very reliable and negligible performance degradation was noticed after multiple days of wafer probed measurements with continuous operation at full output power of 24.8 dBm. Broadband modulated signal measurements showed 36 Gbps of data rate at 28 GHz (7.2 GS/s of single carrier 32 QAM signals with average output power of 17 dBm and 14% PAE and 6 GS/s of single carrier 64 QAM signals with average output power of 14 dBm and 9.3% PAE). The measurements were done without use of any digital pre-distortion techniques. This represents the widest bandwidth modulation measurement reported for any high power mm-wave PA. These properties make this PA an excellent choice for the 5G wireless transmitter frontend. Chapter 4 presents a comparison study of performance and reliability of NMOS and PMOS FET available in deeply scaled CMOS processes. With device measurements of 32 nm CMOS SOI FETs it is shown that the PFETs available in the process are as fast as and more reliable than NFETs. A 3-stack PMOS PA is designed at E-band and measured. The PA achieved a maximum output power of 19.6 dBm and maximum PAE of 24% at 78 GHz. This is the first exclusively PMOS mm-wave PA reported and has the highest reported PAE in CMOS for an E-band PA. Also this represents the highest reported power from a CMOS PA at this frequency without use of elaborate on-chip power combining.

5.2 Future Work

As discussed in the chapter 4, the better performance of PFETs compared to NFETs is expected to hold with continued scaling, at least till the 7 nm gate-length CMOS node when NFETs potentially can have III-V channels and have considerably higher transconductances. Identical NMOS and PMOS PAs, using

the multigate design discussed in chapter 3, implemented in 45 nm CMOS SOI, achieved similar maximum measured output power at 28 GHz. It is shown in chapter 4 that the 32 nm PMOS PA achieved higher output power than the comparable NMOS PA. Simualtions using 14 nm CMOS SOI FinFET show even higher output power improvement for PMOS compared to similar NMOS. Due to the small pitch for gate fingers as well as the increased parasitic capacitance between the contacts and the 3D gate fins, the FinFET devices suffer from increased extrinsic parasitics. Also the nominal supply voltage is only 0.8 V for the 14 nm process. Therefore the multigate style layout which would avoid the intermediate node contacts for the stack FET design using PFETs could be highly advantageous in a FinFET process. But for thin fin on SOI, heat removal could be an issue and careful thermal analysis needs to be done to compare the heat dissipation characteristics of such a structure compared to conventional style with contacts in between. Also compound structures using NFETs and PFETs together could be implemented, which makes use of higher voltage handling capability of PFETs for higher power or complementary gate capacitance variation of NFET and PFET for linearization. Differential implementation of the multigate cell with drain-gate capacitance neutralization could potentially increase both the output power and gain of the multigate PA. Since interstack inductive tuning is difficult to implement in multigate-cell, capacitance accleration (C_{DS} feedback) could increase the efficiency of the PA as well as enable the architecture to be used at higher mm-wave frequencies upto W-band. The 28 GHz multigate PA has close to class-B like efficiency roll-off with power back-off. Implementing Doherty or similar load-modulation techniques with the unit multigate PA could lead to better than class-B PAE characteristics and would be extremely useful in transmitters using high PAPR modulation schemes for the emerging wireless communication standards.

Bibliography

[1] M. Yang, V. W. C. Chan, K. K. Chan, L. Shi, D. M. Fried, J. H. Stathis, A. I. Chou, E. Gusev, J. A. Ott, L. E. Burns, M. V. Fischetti, and M. Ieong, "Hybrid-orientation technology (HOT): opportunities and challenges," *IEEE Transactions on Electron Devices*, vol. 53, no. 5, pp. 965–978, May 2006.

[2] P. Packan, S. Cea, H. Deshpande, T. Ghani, M. Giles, O. Golonzka, M. Hattendorf, R. Kotlyar, K. Kuhn, A. Murthy, P. Ranade, L. Shifren, C. Weber, and K. Zawadzki, "High performance Hi-K + metal gate strain enhanced transistors on (110) silicon," in *2008 IEEE International Electron Devices Meeting*, Dec 2008, pp. 1–4.

[3] C. H. Lin, B. Greene, S. Narasimha, J. Cai, A. Bryant, C. Radens, V. Narayanan, B. Linder, H. Ho, A. Aiyar, E. Alptekin, J. J. An, M. Aquilino, R. Bao, V. Basker, N. Breil, M. Brodsky, W. Chang, L. Clevenger, D. Chidambarrao, C. Christiansen, D. Conklin, C. DeWan, H. Dong, L. Economikos, B. Engel, S. Fang, D. Ferrer, A. Friedman, A. Gabor, F. Guarin, X. Guan, M. Hasanuzzaman, J. Hong, D. Hoyos, B. Jagannathan, S. Jain, S. J. Jeng, J. Johnson, B. Kannan, Y. Ke, B. Khan, B. Kim, S. Koswatta, A. Kumar, T. Kwon, U. Kwon, L. Lanzerotti, H. K. Lee, W. H. Lee, A. Levesque, W. Li, Z. Li, W. Liu, S. Mahajan, K. McStay, H. Nayfeh, W. Nicoll, G. Northrop, A. Ogino, C. Pei, S. Polvino, R. Ramachandran, Z. Ren, R. Robison, I. Saraf, V. Sardesai, S. Saudari, D. Schepis, C. Sheraw, S. Siddiqui, L. Song, K. Stein, C. Tran, H. Utomo, R. Vega, G. Wang, H. Wang, W. Wang, X. Wang, D. Wehelle-Gamage, E. Woodard, Y. Xu, Y. Yang, N. Zhan, K. Zhao, C. Zhu, K. Boyd, E. Engbrecht, K. Henson, E. Kaste, S. Krishnan, E. Maciejewski, H. Shang, N. Zamdmer, R. Divakaruni, J. Rice, S. Stiffler, and P. Agnello, "High performance 14nm SOI FinFET CMOS technology with 0.0174 squm embedded DRAM and 15 levels of Cu metallization," in *2014 IEEE International Electron Devices Meeting*, Dec 2014, pp. 3.8.1–3.8.3.

[4] S. Natarajan, M. Agostinelli, S. Akbar, M. Bost, A. Bowonder, V. Chikarmane, S. Chouksey, A. Dasgupta, K. Fischer, Q. Fu, T. Ghani, M. Giles, S. Govindaraju, R. Grover, W. Han, D. Hanken, E. Haralson, M. Haran,

M. Heckscher, R. Heussner, P. Jain, R. James, R. Jhaveri, I. Jin, H. Kam, E. Karl, C. Kenyon, M. Liu, Y. Luo, R. Mehandru, S. Morarka, L. Neiberg, P. Packan, A. Paliwal, C. Parker, P. Patel, R. Patel, C. Pelto, L. Pipes, P. Plekhanov, M. Prince, S. Rajamani, J. Sandford, B. Sell, S. Sivakumar, P. Smith, B. Song, K. Tone, T. Troeger, J. Wiedemer, M. Yang, and K. Zhang, "A 14 nm logic technology featuring 2nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 squm SRAM cell size," in *2014 IEEE International Electron Devices Meeting*, Dec 2014, pp. 3.7.1–3.7.3.

[5] S. Y. Wu, C. Y. Lin, M. C. Chiang, J. J. Liaw, J. Y. Cheng, S. H. Yang, S. Z. Chang, M. Liang, T. Miyashita, C. H. Tsai, C. H. Chang, V. S. Chang, Y. K. Wu, J. H. Chen, H. F. Chen, S. Y. Chang, K. H. Pan, R. F. Tsui, C. H. Yao, K. C. Ting, T. Yamamoto, H. T. Huang, T. L. Lee, C. H. Lee, W. Chang, H. M. Lee, C. C. Chen, T. Chang, R. Chen, Y. H. Chiu, M. H. Tsai, S. M. Jang, K. S. Chen, and Y. Ku, "An enhanced 16 nm cmos technology featuring 2nd generation FinFET transistors and advanced Cu/low-k interconnect for low power and high performance applications," in *2014 IEEE International Electron Devices Meeting*, Dec 2014, pp. 3.1.1–3.1.4.

[6] H. T. Friis, "A note on a simple transmission formula," *Proceedings of the IRE*, vol. 34, no. 5, pp. 254–256, May 1946.

[7] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.

[8] E. Johnson, "Physical limitations on frequency and power parameters of transistors," in *1958 IRE International Convention Record*, vol. 13, March 1965, pp. 27–34.

[9] H. C. Lin and G. M. Rebeiz, "A 70 - 80 GHz SiGe amplifier with peak output power of 27.3 dBm," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 7, pp. 2039–2049, July 2016.

[10] Y. H. Hsiao, Z. M. Tsai, H. C. Liao, J. C. Kao, and H. Wang, "Millimeter-wave CMOS power amplifiers with high output power and wideband performances," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 12, pp. 4520–4533, Dec 2013.

[11] W. Tai, L. R. Carley, and D. S. Ricketts, "A 0.7 w fully integrated 42 GHz power amplifier with 10% pae in 0.13 um SiGe BiCMOS," in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, Feb 2013, pp. 142–143.

[12] A. Chakrabarti and H. Krishnaswamy, "High-power high-efficiency class-e-like stacked mmwave PAs in SOI and bulk CMOS: Theory and implementation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 8, pp. 1686–1704, Aug 2014.

[13] Y. Zhao and J. R. Long, "A wideband, dual-path, millimeter-wave power amplifier with 20 dBm output power and PAE above 15BiCMOS," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 9, pp. 1981–1997, Sept 2012.

[14] D. Zhao and P. Reynaert, "An e-band power amplifier with broadband parallel-series power combiner in 40-nm CMOS," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 2, pp. 683–690, Feb 2015.

[15] T. O'Sullivan, M. Urteaga, R. Pierson, and P. M. Asbeck, "InP HBT millimeter-wave power amplifier implemented using planar radial power combiner," in *2008 IEEE MTT-S International Microwave Symposium Digest*, June 2008, pp. 293–296.

[16] M. P. DeLisio and R. A. York, "Quasi-optical and spatial power combining," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 3, pp. 929–936, Mar 2002.

[17] Y. A. Atesal, B. Cetinoneri, M. Chang, R. Alhalabi, and G. M. Rebeiz, "Millimeter-wave wafer-scale silicon BiCMOS power amplifiers using free-space power combining," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, no. 4, pp. 954–965, April 2011.

[18] A. K. Ezzeddine and H. C. Huang, "The high voltage/high power FET (HiVP)," in *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium, 2003*, June 2003, pp. 215–218.

[19] M. Shifrin, Y. Ayasli, and P. Katzin, "A new power amplifier topology with series biasing and power combining of transistors," in *IEEE 1992 Microwave and Millimeter-Wave Monolithic Circuits Symposium Digest of Papers*, June 1992, pp. 39–41.

[20] M. Rodwell, S. Jaganathan, and S. T. Allen, "Series-connected microwave power amplifiers with voltage feedback and method of operation for the same," Aug. 31 1999, uS Patent 5,945,879.

[21] T. Sowlati and D. M. W. Leenaerts, "A 2.4 GHz 0.18 um CMOS self-biased cascode power amplifier," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 8, pp. 1318–1324, Aug 2003.

[22] 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 Transactions on Microwave Theory and Techniques, vol. 58, no. 1, pp. 57–64, Jan 2010.

[23] 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 amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 4, pp. 1543–1556, April 2013.

[24] K. Datta and H. Hashemi, "Performance limits, design and implementation of mm-wave SiGe HBT class-E and stacked class-E power amplifiers," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 10, pp. 2150–2171, Oct 2014.

[25] A. Balteanu, I. Sarkas, E. Dacquay, A. Tomkins, G. M. Rebeiz, P. M. Asbeck, and S. P. Voinigescu, "A 2-bit, 24 dbm, millimeter-wave SOI CMOS power-DAC cell for watt-level high-efficiency, fully digital m-ary QAM transmitters," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 5, pp. 1126–1137, May 2013.

[26] A. Agah, J. A. Jayamon, P. M. Asbeck, L. E. Larson, and J. F. Buckwalter, "Multi-drive stacked-FET power amplifiers at 90 GHz in 45 nm SOI CMOS," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 5, pp. 1148–1157, May 2014.

[27] E. Torkildson, C. Sheldon, U. Madhow, and M. Rodwell, "Nonuniform array design for robust millimeter-wave MIMO links," in *GLOBECOM 2009 - 2009 IEEE Global Telecommunications Conference*, Nov 2009, pp. 1–7.

[28] B. Hanafi, O. Grbz, H. Dabag, J. F. Buckwalter, G. Rebeiz, and P. Asbeck, "q-band spatially combined power amplifier arrays in 45-nm CMOS SOI," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 6, pp. 1937–1950, June 2015.

[29] S. M. Bowers and A. Hajimiri, "Multi-port driven radiators," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 12, pp. 4428–4441, Dec 2013.

[30] A. Alizadeh, M. Frounchi, and A. Medi, "On design of wideband compact-size Ka/Q-band high-power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 6, pp. 1831–1842, June 2016.

[31] A. Margomenos, A. Kurdoghlian, M. Micovic, K. Shinohara, D. F. Brown, A. L. Corrion, H. P. Moyer, S. Burnham, D. C. Regan, R. M. Grabar, C. McGuire, M. D. Wetzel, R. Bowen, P. S. Chen, H. Y. Tai, A. Schmitz, H. Fung, A. Fung, and D. H. Chow, "GaN technology for E, W and G-band applications," in *2014 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)*, Oct 2014, pp. 1–4.

[32] F. Golcuk, T. Kanar, and G. M. Rebeiz, "A 90 - 100 GHz 4 x 4 SiGe BiCMOS polarimetric transmit/receive phased array with simultaneous receive-beams capabilities," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 8, pp. 3099–3114, Aug 2013.

[33] W. Shin, B. H. Ku, O. Inac, Y. C. Ou, and G. M. Rebeiz, "A 108 - 114 GHz 4 x 4 wafer-scale phased array transmitter with high-efficiency on-chip antennas," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 9, pp. 2041–2055, Sept 2013.

[34] 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," in *2013 IEEE Topical Conference on Biomedical Wireless Technologies, Networks, and Sensing Systems*, Jan 2013, pp. 79–81.

[35] T. Yamaguchi, T. Sawai, M. Nishida, and M. Sawada, "Ultra-compact 1 W GaAs SPDT switch IC," in *1999 IEEE MTT-S International Microwave Symposium Digest (Cat. No.99CH36282)*, vol. 1, June 1999, pp. 315–318 vol.1.

[36] M. Ahn, C. H. Lee, and J. Laskar, "CMOS high power SPDT switch using multigate structure," in *2007 IEEE International Symposium on Circuits and Systems*, May 2007, pp. 3283–3286.

[37] D. Wang, R. Wolf, A. Joseph, A. Botula, P. Rabbeni, M. Boenke, D. Harame, and J. Dunn, "High performance SOI RF switches for wireless applications," in *2010 10th IEEE International Conference on Solid-State and Integrated Circuit Technology*, Nov 2010, pp. 611–614.

[38] J. Cai, T. Ning, P. Oldiges, A. Chou, A. Kumar, W. Rausch, W. Haensch, and G. Shahidi, "SOI series MOSFET for embedded high voltage applications and soft-error immunity," in *2008 IEEE International SOI Conference*, Oct 2008, pp. 21–22.

[39] S. Lee, B. Jagannathan, S. Narasimha, A. Chou, N. Zamdmer, J. Johnson, R. Williams, L. Wagner, J. Kim, J. O. Plouchart, J. Pekarik, S. Springer, and G. Freeman, "Record RF performance of 45-nm SOI CMOS technology," in *2007 IEEE International Electron Devices Meeting*, Dec 2007, pp. 255–258.

[40] J. O. Plouchart, "Applications of SOI technologies to communication," in *2011 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)*, Oct 2011, pp. 1–4.

[41] O. Inac, M. Uzunkol, and G. M. Rebeiz, "45-nm CMOS SOI technology characterization for millimeter-wave applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 6, pp. 1301–1311, June 2014.

[42] B. Razavi, R.-H. Yan, and K. F. Lee, "Impact of distributed gate resistance on the performance of MOS devices," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 41, no. 11, pp. 750–754, Nov 1994.

[43] B. Cetinoneri, Y. A. Atesal, A. Fung, and G. M. Rebeiz, "*w* -band amplifiers with 6-db noise figure and milliwatt-level 170-200 GHz doublers in 45-nm CMOS," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 3, pp. 692–701, March 2012.

[44] B. Hanafi, "Design of silicon power ampliers and arrays for millimeter wave applications," 2014.

[45] J. Chang, K. Kim, S. Lee, and S. Nam, "24 GHz stacked power amplifier with optimum inter-stage matching using 0.13 um CMOS process," in *2011 3rd International Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*, Sept 2011, pp. 1–3.

[46] M. Seo, B. Jagannathan, J. Pekarik, and M. J. W. Rodwell, "A 150 GHz amplifier with 8 dB gain and +6 dBm p_{sat} in digital 65 nm CMOS using dummy-prefilled microstrip lines," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 12, pp. 3410–3421, Dec 2009.

[47] J. H. Chen, S. R. Helmi, R. Azadegan, F. Aryanfar, and S. Mohammadi, "A broadband stacked power amplifier in 45-nm CMOS SOI technology," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 11, pp. 2775–2784, Nov 2013.

[48] S. Y. Mortazavi and K. J. Koh, "A 28 GHz inverse class-F power amplifier with coupled-inductor based harmonic impedance modulator," in *2015 IEEE Custom Integrated Circuits Conference (CICC)*, Sept 2015, pp. 1–4.

[49] A. Sarkar and B. A. Floyd, "A 28 GHz harmonic-tuned power amplifier in 130-nm SiGe BiCMOS," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 2, pp. 522–535, Feb 2017.

[50] J. Curtis, A. V. Pham, M. Chirala, F. Aryanfar, and Z. Pi, "A Ka-band doherty power amplifier with 25.1 dBm output power, 38 percent peak PAE and 27 percent back-off PAE," in *2013 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, June 2013, pp. 349–352.

[51] C. F. Campbell, M.-Y. Kao, and S. Nayak, "High efficiency Ka-band power amplifier MMICs fabricated with a 0.15 um GaN on SiC HEMT process," in *2012 IEEE/MTT-S International Microwave Symposium Digest*, June 2012, pp. 1–3.

[52] H. c. Park, S. Daneshgar, Z. Griffith, M. Urteaga, B. S. Kim, and M. Rodwell, "Millimeter-wave series power combining using sub-quarter-wavelength baluns," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 10, pp. 2089–2102, Oct 2014.

[53] K. Datta and H. Hashemi, "Watt-level mm-wave power amplification with dynamic load modulation in a SiGe HBT digital power amplifier," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 2, pp. 371–388, Feb 2017.

[54] K. Dasgupta, K. Sengupta, A. Pai, and A. Hajimiri, "A mm-wave segmented power mixer," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 4, pp. 1118–1129, April 2015.

[55] A. K. Sahoo, S. Fregonese, M. Weis, N. Malbert, and T. Zimmer, "A scalable electrothermal model for transient self-heating effects in trench-isolated SiGe HBTs," *IEEE Transactions on Electron Devices*, vol. 59, no. 10, pp. 2619–2625, Oct 2012.

[56] A. Ghosh, "The 5G mmwave radio revolution." *Microwave Journal*, vol. 59, no. 9, 2016.

[57] J. P. John, J. Kirchgessner, R. Ma, D. Morgan, I. To, and V. P. Trivedi, "Si-based technologies for mmwave automotive radar," in *2016 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)*, Oct 2016, pp. 1–4.

[58] Z. Griffith, M. Urteaga, P. Rowell, and R. Pierson, "71 - 95 GHz (23 - 40 percent PAE) and 96 - 120 GHz (19 - 22 percent PAE) high efficiency 100 - 130 mW power amplifiers in InP HBT," in *2016 IEEE MTT-S International Microwave Symposium (IMS)*, May 2016, pp. 1–4.

[59] K. Datta and H. Hashemi, "High-breakdown, high- $f_{\max}$ multiport stacked-transistor topologies for the W -band power amplifiers," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 5, pp. 1305–1319, May 2017.

[60] J. Jayamon, O. Gurbuz, B. Hanafi, A. Agah, J. Buckwalter, G. Rebeiz, and P. Asbeck, "Spatially power-combined W-band power amplifier using stacked CMOS," in *2014 IEEE Radio Frequency Integrated Circuits Symposium*, June 2014, pp. 151–154.

[61] J. Kim, H. Dabag, P. Asbeck, and J. F. Buckwalter, "q -band and W -band power amplifiers in 45-nm CMOS SOI," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 6, pp. 1870–1877, June 2012.

[62] S. Kulkarni and P. Reynaert, "A 60 GHz power amplifier with AM-PM distortion cancellation in 40-nm CMOS," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 7, pp. 2284–2291, July 2016.

[63] I. Sarkas, A. Balteanu, E. Dacquay, A. Tomkins, and S. Voinigescu, "A 45nm SOI CMOS class-D mm-wave PA with 10 Vpp differential swing," in *2012 IEEE International Solid-State Circuits Conference*, Feb 2012, pp. 88–90.

[64] J. A. del Alamo, D. A. Antoniadis, J. Lin, W. Lu, A. Vardi, and X. Zhao, "III-V MOSFETs for future CMOS," in *2015 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)*, Oct 2015, pp. 1–4.

[65] R. Carter, J. Mazurier, L. Pirro, J. U. Sachse, P. Baars, J. Faul, C. Grass, G. Grasshoff, P. Javorka, T. Kammler, A. Preusse, S. Nielsen, T. Heller, J. Schmidt, H. Niebojewski, P. Y. Chou, E. Smith, E. Erben, C. Metze, C. Bao, Y. Andee, I. Aydin, S. Morvan, J. Bernard, E. Bourjot, T. Feudel, D. Harame, R. Nelluri, H. J. Thees, L. M-Meskamp, J. Kluth, R. Mulfinger, M. Rashed, R. Taylor, C. Weintraub, J. Hoentschel, M. Vinet, J. Schaeffer, and B. Rice, "22 nm FDSOI technology for emerging mobile, internet-of-things, and RF applications," in *2016 IEEE International Electron Devices Meeting (IEDM)*, Dec 2016, pp. 2.2.1–2.2.4.

[66] J. Watts, K. Sundaram, K. W. J. Chew, S. Lehmann, S. N. Ong, W. H. Chow, L. H. Chan, J. Mazurier, C. Schwan, Y. Andee, T. Feudel, L. Pirro, E. Erben, E. Nowak, E. Smith, E. M. Bazizi, T. Kammler, R. Taylor, B. Rice, and D. Harame, "RF-pFET in fully depleted SOI demonstrates 420 GHz ft," in *2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, June 2017, pp. 84–87.

[67] S. M. Sze and K. K. Ng, *Physics of semiconductor devices*. John wiley & sons, 2006.

[68] R. S. Muller, T. I. Kamins, M. Chan, and P. K. Ko, *Device electronics for integrated circuits*. Wiley New York, 2003, vol. 986.

[69] S. Mittl and F. Guarn, "Self-heating and its implications on hot carrier reliability evaluations," in *2015 IEEE International Reliability Physics Symposium*, April 2015, pp. 4A.4.1–4A.4.6.

[70] S. Lombardo, E. Wu, and J. Stathis, "Comprehensive model for progressive breakdown in nFETs and pFETs," in *2016 IEEE International Electron Devices Meeting (IEDM)*, Dec 2016, pp. 31.6.1–31.6.4.

[71] C. Enz and Y. Cheng, "MOS transistor modeling for RF IC design," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 2, pp. 186–201, Feb 2000.

[72] J. Oh, B. Ku, and S. Hong, "A 77 GHz CMOS power amplifier with a parallel power combiner based on transmission-line transformer," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 7, pp. 2662–2669, July 2013.

[73] A. Larie, E. Kerherv, B. Martineau, L. Vogt, and D. Belot, "A 60 GHz 28 nm UTBB FD-SOI CMOS reconfigurable power amplifier with 21dBm P1dB and 74 mW PDC," in *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, Feb 2015, pp. 1–3.