Ka-Band and W-Band Millimeter-Wave Wideband Linear Power Amplifier Integrated Circuits at 30 GHz and 90 GHz with Greater Than 100 mW Output Powers in Commercially-Available 0.12 μm Silicon Germanium HBT Technology

by

Michael S. Chang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering)
in The University of Michigan
2017

Doctoral Committee:

Professor Gabriel M. Rebeiz, Univ. of California, San Diego, Co-Chair
Professor Amir Mortazawi, Co-Chair
Professor Michael P. Flynn
Professor Kim A. Winick

Michael S. Chang

changm@umich.edu

ORCID iD: 0000-0002-0701-4996

To my family, friends, teachers & mentors, and Kaiann.

Acknowledgements

This work has been sponsored by Alfred Hung at Army Research Labs (ARL) and Thomas Kenny at DARPA. Access to the commercial foundry technology was sponsored and provided by TAPO (Trusted Access Program Office). Additional IBM/GF SiGe 8HP foundry-run support has been provided by MOSIS and IBM (now GlobalFoundries). My graduate studies have been also been generously sponsored by the Agilent Technologies Graduate Fellowship from the Wireless Semiconductor Division at Hewlett Packard/Agilent Technologies (now Avago/Broadcom) as well as the University of Michigan EECS/ECE department.

I thank my committee for their participation, feedback, guidance, and support throughout the doctoral process: Professors Gabriel Rebeiz, Amir Mortazawi, Michael Flynn, and Kim Winick. Kaiann & I have had the wonderful opportunities to be students in their classrooms throughout our years in the EECS undergraduate & graduate programs.

The following people have been generous with their time and efforts to support parts of my research projects:

- Fellow group-mates Yusuf Atesal, Berke Cetinoneri, and Ramadan Alhalabi, at UC San Diego for their collaboration and extended work into power-combined power-amplifiers arrays based on the W-band PA work.

- Byung-Wook Min and Christopher Galbraith for their assistance in the lab with my Ka-band measurements and for finding, purchasing, and setting up many pieces of key equipments after the major lab move from Michigan.

- Jason May and Sang Young Kim for their assistance in the lab with my W-band measurements and for helping to obtain and setup the W-band waveguide probe station.

- Prof. Kamal Sarabandi for W-band lab equipment from the EECS Radlab at University of Michigan.

- Don Kimball at UCSD for helping secure key pieces lab equipment for my measurements when our lab was still being built.

- Prof. Peter Asbeck's group for sharing the waveguide measurement probe station in CalIT^2 lab and allowing me to stay on the station for months to complete my many W-band measurements.

- Ulrich Pfeiffer, formerly at IBM, for valuable technical discussions when I first started working in the new IBM 8HP process.

I learned a great deal working with my industry mentors and managers throughout my time with the Wireless Semiconductor Division (WSD) & Integrated Circuits Business Division (ICBD) at Hewlett Packard, Agilent Technologies, and Avago Technologies (now Broadcom): Dave Der-Wei Wu, Marshall Maple, Issy Kipnis, Domingo Fugerero, Henrik Morkner, S.L. Shyh-Liang Fu, Phil Fischer, Detlef Daniel, Jim McMahon, Lan Nguyen, Maurice Smallwood, and Jim White. They taught me a great deal about practical wireless MMIC & RFIC circuit design, power amplifier design, modeling, lab measurements & bench tuning, and life in Silicon Valley.

Thank you to the University of Michigan EECS Department and ECE program chairs for all their support. I am grateful to the EECS and Radlab staff who looked out for me and provided me with generous assitance: Beth Stalknaker, Michelle Chapman, Jose-Antonio D. Rubio, and Steven Pejuan. They care about the students in the graduate programs very much, and it always shows.

Much appreciation goes to Yusuf Atesal, Berke Cetinoneri, Tumay Kanar, Sang Young Kim, Timothy Lee, Jason May, Chenhui Niu, Chirag Patel, Bhaskara Rupakula, Woorim Shin, Berktug Ustundag, Yang Yang, and Tiku Yu for their assistance with administering the RFIC servers, CAD tools, and PCs with me and for always keeping me well-connected, remotely, to UCSD.

Graduate school life is full of work, but it is also full of many interesting times with wonderful people. I am thankful for the intellectual discussions, advice & assistance, friendship, and many great times spent with my fellow TICSG/Tixer group mates: Abbas Abbaspoor-Tamijani, Yusuf Atesal, Ramadan Al-Halabi, Andrew Brown, Jose Cabanillas Costa, Berke Cetinoneri, Mohammed El-Tanani, Jennifer Edwards, Kamran Entesari, Christopher Galbraith, Alex Grichener, Philip Timothy Hancock, Jeong-Geun Kim, Sang-Young Kim, Kwang-Jin Koh, Balaji Lakshminarayana, Bryan Hung, Jason May, Denis Mercier, Jeremy Muldavin, Byungwook Min, Kiran Nimmagada, Sang-June Park, Hojr Sedaghat Pisheh, Isak Reines, Jad Rizk, Bernhard Schoenlinner, Woorim Shin, Guan Leng Tan, Carson White, Mehmet Uzunkol, Tauno Vaha-Heikkila, and Tiku Yu.

A special thanks goes to my relatives and friends who housed me during my multiple trips and long stays in San Diego for my measurements: Christie Chang & Brian Sato and

Sungill Kim & Elaine Sung-Hae Kim. They made La Jolla and San Diego a home away from home for me.

Gratitude to long-time friends Renee & Nicholas Rice, Alice & Leonard Wong, and Seekit & Suszan Leong for welcoming me into their homes during my conference trips and other short trips to the SF Bay Area; and in Ann Arbor, Gavin & Annie Sy and Joe & Jen Brunett for all the years of great company (Go Blue!). Thanks also to the Fu family and Leeann & Peter for the many wonderful meals & memorable times together in Ann Arbor.

I must express special thanks to Gabriel Rebeiz— advisor, teacher, mentor, and friend—, for all his generous support, personal & professional, to both Kaiann and myself through the many years. Being part of the TICSG family has been a wonderful experience.

Finally, life is not complete without Kaiann and my loving family, Man-Feng, Ming-Ming, and Jennifer. They have always fully supported me in all my endeavors and provide me the motivation to tackle the next challenge. Love to them always...

Michael Chang
Ann Arbor, Michigan
November 13, 2017

Table of Contents

List of Tables

List of Figures

List of Appendices

Appendix

List of Abbreviations

β_F Forward current gain from base to collector.

η_C Collector efficiency.

λ_0 Free-space wavelength.

1D One-dimesional.

2D Two-dimesional.

$\mathbf{2}f_0$ Twice the center operating frequency.

8HP GlobalFoundries 0.12 µm silicon germanium HBT process technology with 0.13 µm FETs.

9HP GlobalFoundries 90 nm silicon germanium process technology.

12SOI GlobalFoundries 45 nm SOI CMOS process technology (now 45RFSOI).

AC Alternating current.

ACPR Adjacent channel power ratio.

ADS Advanced Design System CAD tool by Keysight Technologies (formerly Agilent Technologies).

AM Top-most, aluminum metal layer in GF 8HP.

BEOL Back-end of line process in IC fabrication (above substrate).

BJT Bipolar junction transistor.

BV Breakdown voltage.

$BV_{\mathbf{CEO}}$ Collector-to-emitter breakdown voltage with open base.

$BV_{\mathbf{CBO}}$ Collector-to-base breakdown voltage with open emitter.

$BV_{\mathbf{CER}}$ Operational collector-to-emitter breakdown voltage with external base resistance R.

BW Bandwidth.

CAD Computer-aided design.

CB Common base topology or collector-base junction when subscripted.

CE Common emitter topology.

CMOS Complementary metal-oxide semiconductor.

CMR Common-mode rejection.

CMRR Common-mode rejection ratio.

CPW Co-planar waveguide.

D-band 110–170 GHz frequency range.

DC Direct current.

DEC Decade.

deg Degrees.

DT Deep-trench isolation.

DUT Device under test.

EIRP Effective isotropic radiated power.

EM Electromagnetics.

ESD Electrostatic discharge.

ESR Equivalent series resistance.

EVM Error vector magnitude.

f_0 Center operating frequency.

FEOL Front-end of line process in IC fabrication (within substrate).

FET Field-effect transistor.

FEM Finite-element method.

f_{MAX} Maximum frequency of oscillation.

f_T Transition frequency.

GaAs Gallium arsenide.

GF GlobalFoundries.

GFUS GlobalFoundries United States (formerly IBM).

HBT Heterojunction bipolar transistor.

HEMT High electron mobility transistor.

HiCUM High-current model.

III-V Compound III-V semiconductor technology.

IP3 3rd-order intercept point.

I_B DC base current in bipolar transistor.

I_C DC collector current in bipolar transistor.

IC Integrated circuit.

InP Indium phosphide.

IM Intermodulation.

I/O Input/output.

ITU International Telecommunication Union.

ITRS International Technology Roadmap for Semiconductors.

IV Current and voltage.

Ka-band 26.5–40 GHz frequency range.

Ku-band 12–18 GHz frequency range.

LC Inductor-capacitor network.

L_e Emitter inductance.

Lith Lithography.

LNA Low-noise amplifier.

$\mathbf{M}_{234}$ Stacked copper line made from layers M2, M3, and M4.

M1 Copper metallization level in GF 8HP (closest to substrate).

M2 Copper metallization level in GF 8HP (2nd metal layer).

M3 Copper metallization level in GF 8HP (3rd metal layer).

M4 Upper copper metallization level in GF 8HP (4th metal layer).

MAG Maximum available gain.

MESFET Metal-semiconductor field-effect transistor.

mil One-thousandth of an inch or 25.4 μm.

MIM Metal-insulator-metal (capacitor with nitride dielectric).

MMIC Monolithic microwave integrated circuit.

MN Matching network.

mmW Millimeter-wave.

MOM Metal-oxide-metal (capacitor).

MOSFET Metal-oxide semiconductor field-effect transistor.

MQ Copper metal layer used as transmission-line ground shields.

m_{th} Mutual thermal coupling coefficient.

MSG Maximum stable gain.

NF Noise figure.

NFET n-channel field-effect transistor.

NIST National Institute of Standards and Technology.

Np Nepers.

npn Bipolar transistor with n-doped emitter, p-doped base, and n-doped collector.

OCT Octave.

OFDM Orthogonal frequency-division multiplexing.

pnp Bipolar transistor with p-doped emitter, n-doped based, and p-doped collector.

p-cell Parameterized cell.

p-sub p^- type silicon substrate.

$\mathbf{P_{1dB}}$ Output-referred power at the 1 dB gain compression point (relative to small-signal gain).

PA Power amplifier.

PAE Power-added efficiency.

PAR Peak-to-average ratio.

PCB Printed circuit board.

PFET p-channel field-effect transistor.

$\mathbf{P_{in}}$ RF input power.

PMEPR Peak-to-mean envelope power ratio.

POH Power-on hours.

$\mathbf{P_{out}}$ RF output power.

$\mathbf{P_{sat}}$ Saturated RF output power (maximum output power).

PVT Process, voltage, and temperature.

Q Quality factor.

$\mathbf{Q_0}$ Unit-cell transistor (5 μm length).

Q1 First gain stage power transistor.

Q2 Second gain stage power transistor.

Q3 Third gain stage power transistor.

Qb Bias-reference power transistor.

QAM Quadrature amplitude modulation.

QW Quarter-wavelength.

rad Radians.

RAM Random access memory.

R_{th} Thermal resistance.

RF Radio frequency.

RFC RF choke (ideal very large inductor).

RFIC Radio frequency integrated circuit.

RLC Resistor-inductor-capacitor network.

RMS Root-mean square.

SC Short-circuited.

SCR Space-charge region.

SMT Surface-mount technology.

S-parameters Scattering parameters.

Si Silicon.

SiGe Silicon germanium.

SiO_2 Silicon dioxide.

SOA Safe operating area.

SOC System-on-chip.

SOI Silicon-on-insulator.

SPICE Simulation program with integrated circuit emphasis.

SS Small-signal.

STI Shallow-trench isolation.

TaN Tantalum nitride (resistor).

TCAD Technology computer aided design.

T-line Transmission line.

TSV Through-silicon via.

V-band 40–75 GHz frequency range.

VBIC Vertical bipolar inter-company model.

V_{CC} DC collector-supply voltage.

VNCAP Vertical natural capacitor.

W-band 75–110 GHz frequency range.

Z_0 Characteristic impedance.

Abstract

This dissertation presents several fully-integrated power amplifiers (PAs) within Ka-band (26.5–40 GHz) and W-band (75–110 GHz) with > 100 mW output powers in commercially-available 0.12 μm silicon germanium HBT-transistor technology (f_T = 200 GHz, BV_{CEO} = 1.8 V).

Linear PAs are by far the highest power-consuming component in wireless communication circuits. As wireless standards push into millimeter-wave frequencies for high data rates, challenges created by point-to-point propagation require transceiver phased arrays for antenna-beam steering to maintain optimal channel links. Integration of high-output-power, linear silicon PAs is key in enabling compact and efficient transmitter arrays.

Linear power amplification is especially difficult in silicon at millimeter-wave frequencies because limited transistor gains, low transistor breakdown voltages, and high-loss, high-parasitic passive-circuit components typically result in low output powers and poor power efficiencies. Moreover, at 90–100 GHz, integrated-circuit PAs are sensitive to process skews and HBT modeling errors, as transistor model-to-hardware correlations degrade after 60 GHz.

These SiGe PAs demonstrate record Class-A/AB linear output powers by pushing limits of voltage, power, and frequency while maintaining reliability-aware design.

To overcome voltage limitations of impact ionization, base-ballasting techniques using the lowest reported HBT base-biasing resistance of 20 Ω push operational collector voltages well above the native BV_{CEO}. Additional novel inverted microstrip quarter-wave transmission-line structures, created below the RF global ground plane and terminated by resonant MIM capacitors, present the necessary high-impedance at the fundamental frequency and Class AB harmonic terminations to the HBT base terminal.

This is the first SiGe work to pack unit-cell SiGe HBT transistors at maximum density in 2D arrays to synthesize large power transistors, circumventing needs for transistor-level power splitting and power combining at W-band. Measured data is used to model thermal coupling and provide custom HBT thermal impedance models for stability.

Resonant capacitor structures are used throughout the 3-stage PA to provide ultra-compact, low-impedance RF references and to synthesize, along with short transmission lines, very low values of DC-blocked, shunt inductance required for interstage power matching of

large power HBTs.

Performance impacts from HBT-modeling errors at high frequencies and foundry-process skews are mitigated using a unit-cell design approach throughout the PA, from device level to system level. Identical or similar-characteristic components are used throughout the PA as unit cells: transistors, matching-network capacitors, resonant capacitors, resistors, transmission lines, bias circuits, and matching-network topologies. Additionally, fine bias-adjustment provisions for each gain stage allow tuning of critical DC operating points as collector voltages are pushed higher.

For integration into arrays, the designs integrate wideband capacitor-filtering arrays for stability and follow reliability guidelines for 100,000 power-on-hours operation at 100°C. A single Ka-band chip is tested for > 1000 hours and a W-band chip for > 500 hours with no measurable performance degradation.

The W-band 90-GHz power amplifier achieves a maximum saturated output power of 19.9 dBm (98 mW), 14.6 dB gain with 20% fractional 3-dB bandwidth from 79–97 GHz, and peak power-added efficiency (PAE) of 15.4%. The Ka-band 33-GHz power amplifier achieves a maximum saturated output power of 24.2 dBm (260 mW), 14.6 dB gain with 55% fractional 3dB-bandwidth from 20.5–36 GHz, and peak PAE of 22.3%.

In an extended collaboration, the W-band PA has been successfully integrated within a 3×3 SiGe transmit array using quasi-optical free-space combining techniques to achieve 3.2 watts EIRP, the highest reported power at 90–98 GHz in SiGe technology.

Chapter 1

Introduction

1.1 Overview

This written dissertation presents several fully-integrated power amplifiers (PAs) fabricated in commercially-available 0.12 μm silicon germanium (SiGe) foundry technology within Ka-band (26.5–40 GHz) and W-band (75–110 GHz). A standalone Ka-band PA was designed first and published in [7]. Subsequently, a standalone W-band PA was designed and published in [8]. Finally, a W-band PA array was designed in collaboration with colleagues at Univ. of California, San Diego and published in [9].

This dissertation presents the two standalone PA designs in reverse order: Chapter 2 presents the W-band standalone PA first, in order to review the most challenging and encompassing design details first; Chapter 3 presents the Ka-band PA. The W-band design builds upon the Ka-band PA and includes additional design techniques required to overcome the challenges of operating at higher frequencies. The W-band and Ka-band PA design also have minor variants in breakouts that are measured, as well.

To avoid repeating identical concepts and design approaches in both Chapter 2 and Chapter 3, the design details shared between W-band and Ka-band are presented primarily in Chapter 2. The PA array is covered in Chapter 4 along with potential future research opportunities and extensions.

The following sections in this introduction further outline the dissertation work and summarize novel contributions made to the field of SiGe millimeter-wave (mmW) power amplifiers. General Class AB power amplifier theory and SiGe foundry process information is found in the later portions of this chapter.

1.1.1 Background

GlobalFoundries (GF) 0.13 μm SiGe BiCMOS[1] 8HP became commercially available in August, 2005.[2] The 200 mm (8 inch) foundry process was formally known as IBM 8HP before its acquisition by GlobalFoundries in July, 2015.

8HP SiGe technology has been instrumental in enabling mmW integrated circuits (ICs) research in silicon around popular mmW frequency bands, 24, 30, 60, 77, and 90 GHz, for integrated communication systems and radars [10, 11]. Compared to the prior 0.18 μm SiGe BiCMOS 7HP technology process, the 8HP heterojunction bipolar transistor (HBT) f_T & f_{MAX} have been increased from 120 GHz & 100 GHz to 200 GHz & 240 GHz, respectively. Along with PAs, other transmit and receive circuits (e.g., LNA, mixers, switches, and digital control circuits) were also being developed in SiGe with the ultimate goal of enabling fully-integrated mmW transmit and receive chips in silicon. The 8HP HBTs' high f_T/f_{MAX} (> 200 GHz) enable active-gain circuit operation at mmW frequencies up to 100 GHz, but low gains and breakdown voltages arising from fundamental device & materials properties make achieving large-signal output power at mmW frequencies particularly challenging.

In the research community, the primary goal of the initial mmW SiGe power amplifier work was to obtain as much output power as possible in highly linear operation— Class A or AB amplification. The initial research was focused primarily on pushing the limits of both output power and frequency to explore achievable limits with SiGe HBTs with limited operational collector-emitter breakdown voltages BV_{CER} between $BV_{\mathrm{CEO}} = 1.8$ V to $BV_{\mathrm{CBO}} = 6.0$ V. Typically, the f_T & f_{MAX} should be at least 10–20 times the operating frequency for high RF performance with low circuit complexity. Thus, designing circuits above 20 GHz in 8HP requires careful design choices & optimizations due to tradeoffs in limited available gain with other critical design parameters such as linearity, efficiency, and noise.

A major challenge of the 8HP PA design was the inaccuracy of large-signal simulations due to transistor model limitations and severe simulation convergence issues at high collector voltages. In the early-stage SiGe processes, the HBT models are less refined and hence have limited capability in predicting implemented hardware performance, particularly at high mmW frequencies > 60 GHz where models are difficult to extract accurately from direct measurements. The design kit's RF device models are normally modeled and fitted using extracted parameters at lower frequencies which are then scaled & extrapolated to cover frequencies up to f_{MAX} and beyond. Although operation above $BV_{\mathrm{CEO}} = 1.8$ V was known to be possible, simulations and optimizations of large-signal designs were problematic for

[1] BiCMOS is a technology designation for a process that contains both bipolar and CMOS transistors.

[2] Depending on author and devices used, GF 8HP SiGe BiCMOS is regularly cited as either 0.12 μm or 0.13 μm technology. The HBT emitter width is 0.12 μm while the CMOS FETs have 0.13 μm gate lengths.

collector voltages > 1.7 V for common-emitter configurations using the original HBT models. Thus, part of the difficulty of pushing output power higher was the inability to precisely determine the optimal operating conditions at high collector voltages and currents, especially pertaining to RF interstage matching and optimal output load-line impedances.

As the limits of a new IC technology are explored at higher and higher mmW frequencies, PA development typically advances in distinct phases in the PA research community:

1. mmW frequency and power: PA development begins with Class A/AB operation to explore the transistor and technology limits in frequency and power for highly linear designs. These PA designs focus on determining basic design limits in the new process technology for critical decisions such as power-transistor sizing.

2. Power combining: often, there is also parallel development of on-chip power combining of the Class A/AB PAs to increase maximum output power at the expense of overall power efficiency. New foundry process developments enable lower loss and more complex on-chip planar combining structures.

3. Efficiency: once the limits of linear Class A/AB are explored and understood, other classes of PA operation are investigated to further improve power efficiency and output power, usually at the expense of linearity. Advanced stages of development lead to highly-efficient switch-mode classes of operation (e.g., Class E, F, and variants). Efficiency-enhancing architectures, such as Doherty and switched-load architectures, are often attempted to boost efficiency further.

4. Linearity: finally, as PA development matures with different architectures and approaches, linearity comes into focus as PAs are integrated into transmitter systems with system-level linearity requirements. As PAs are pushed toward higher output powers and efficiencies near the transistor operational limits, non-linearities arise from large-signal compression, gain variations, and intermodulation (IM) distortions even for linear-mode PAs. (The large-signal non-linearity is especially difficult to predict accurately in simulations because of transistor-model limitations when they are operating at their current and voltage extremes.) To mitigate this loss of linearity from targeting high output powers, advanced linearization techniques are applied:

    (a) Non-linearities are compensated directly within the multi-stage PA block. Linearization techniques are used to internally compensate linear-mode PAs (e.g., Class A, AB, B, C) without additional system-level blocks and complexity in the transmit chain.

(b) In contrast, additional *external* signal-processing blocks (and even additional PAs) can be added around the PA to compensate for its non-linearities. The signal-processing techniques include feed-forward correction, pre-distortion, envelope elimination & restoration (also known as polar, EER, and Kahn transmitters), and out-phasing [12]. These techniques are required to linearize extremely nonlinear PAs, such as switch-mode PAs; and come at the cost of significant, added system-level complexities.

This dissertation work was part of the initial exploration research phase where the primary goal was to achieve highly linear, high output power (17–20 dBm in Class A/AB) with reasonable gain (> 10 dB) and efficiency (PAE > 10%). Because of the high currents and voltages within the PA, an additional goal was to follow foundry guidelines for reliability across all active and passive devices. Meeting these combined goals allows the resulting PA to be successfully integrated into a larger transmit chain: as an element in an wafer-scale transmitter array, as a highly linear driver to an off-chip III-V high-power PA, or as a standalone output PA in a highly linear transmit chain.

The research community had investigated various techniques to produce high output power, linear mmW PAs, such as differential topologies [13–20], common-emitter [15, 21–24], cascode [13, 16–19, 24, 25, 25–28], distributed amplifiers [26], fully-distributed PA design with corporate feeds and power combining [22], and and power-combining of the output transistor [19–22, 22, 26].

This dissertation work extends the maximum powers and frequencies of linear Class A/AB PAs by introducing novel IC design elements, applying a unit-cell design approach everywhere possible while still optimizing internals of the PA design.

Efforts are made to preserve RF gain at 90 GHz for practical PAE (> 10%) goals in order to remain suitable for integration into larger PA array, demonstrated in extended work in Chapter 4. Many design aspects are also centered around pushing operational collector voltages past values (1.7 V) where simulations begin to fail because of HBT model limitations. Best efforts are made to be compliant with reliability guidelines for 100,000 power-on-hour operation at 100°C for commercial electronics applications.

1.1.2 Design Progression

The primary dissertation work has been completed across a total of three SiGe tapeouts in GlobalFoundries (GF) SiGe BiCMOS 8HP, then known as IBM 8HP. A fourth tapeout included collaboration with Yusuf Atesal, Berke Cetinoneri, and Ramadan Alhalabi at University of California, San Diego, to integrate the W-band power amplifiers into a large

3×3 single-chip, wafer-scale PA array with quasi-optical spatial power combining.

The following sections document the progression of the tapeouts and design process from Ka-band to W-band to extended 3×3 PA-array collaborations.

1. 30 GHz Ka-Band PA: First Iteration at 1.4–2.1 V

The initial Ka-band design is presented in Chapter 3 and published in [7]. This was the first tapeout in GF 8HP at the University of Michigan and the first silicon hardware parts— the PA was designed without prior silicon samples for measurement characterization of passive & active devices. The design relied solely on simulations of HBT foundry models, foundry-measured data sheets for some limited HBT structures, and custom electromagnetic (EM) simulations for the passive structures (transmission lines, interconnects, parasitics, inductors, and capacitors).

A particular challenge centered around the modeling of power transistors based on custom layouts of aggregated transistors. GF 8HP does not offer dedicated power-transistor cells or standard transistor cells greater than 18 μm emitter lengths. Part of the design optimization involved inspecting measured-versus-modeled HBT DC and RF data within the foundry-supplied data sheets to make critical decisions on HBT reactive matching and operating points since there was still significant model-to-hardware error at mmW frequencies.

One of the goals of the initial Ka-band design was to determine how large the output transistor could be while obtaining enough gain for the PA to have sufficient PAE. Larger output transistors are capable of delivering more power, but the matching (impedance transformation) is more difficult, and gain drops.

The baseline Ka-band design #1 is designed around a conservative Class A design with $V_{CC} = 1.4$ V nominal collector voltage since simulations issues in the early HBT models make large-signal design difficult above 1.4–1.5 V. The independent current & voltage biasing provisions in each gain stage allows operating and tuning the design into Class AB at the measurement bench. With the bias provision, higher output powers and efficiencies are successfully obtained at 30 GHz in the lab beyond 1.4 V all the way to $V_{CC} = 2.1$ V. The active bias circuits present 40 Ω external DC resistances to the HBT power cells in each gain stage.

The 2-stage common-emitter Class A PA with Q1 = 30 μm & Q2 = 60 μm custom power-cell HBTs is designed at a conservative 1.4 V supply voltage for better model accuracy since there are significant simulation convergence issues for large-signal simulations at voltages above 1.5 V arising from high-current limitations in the HBT's early VBIC transistor model. All matching networks and active bias networks are integrated on chip. The active bias circuits present 40 Ω external DC resistances to the HBT power cells in each gain stage.

Bias provisions are included to adjust collector currents in each gain stage that allows tuning the design into Class AB at the measurement bench as well as push the collector voltages higher for more output power and efficiency. With the bias provision, higher output powers and efficiencies in Class AB operation are successfully obtained at 30 GHz in the lab beyond 1.4 V all the way to $V_{CC} = 2.1$ V.

At nominal 1.4 V biasing, the first iteration of the Ka-band design achieves $P_{\rm sat} = 19.4$ dBm differential output power at 32 GHz, $P_{\rm 1dB} = 16.4$ dBm, and peak-PAE = 11.2%. The small-signal gain is 13.0 dB with a 3dB-bandwidth from 26–40 GHz (42%).

With increased collector voltages to 2.1 V and re-tuned collector currents for optimized biasing, the same first-iteration Ka-band PA achieves $P_{\rm sat} = 21.3$ dBm differential output power at 30 GHz, $P_{\rm 1dB} = 15.5$ dBm, and peak-PAE = 12.9%. The small-signal gain is 15.5 dB with a 3dB-bandwidth from 26–40 GHz (42%).

2. 30 GHz Ka-Band PA: Improved Design at 1.5–2.2 V

The second Ka-band design iteration builds upon observations from the extensive Ka-band DC & RF measurements and carries the approach forward to more aggressive PA designs. A key confirmation is that the PA design, with proper independent bias adjustments, can be successfully tuned & operated at higher collector voltages with high RF performance even though large-signal simulations would not converge for simulations above 1.7 V collector voltages.

The second Ka-band design is more aggressive in terms of transistor sizing, power matching, and low external base biasing resistance to push collector voltages. The second Ka-band iteration is designed around 1.7 V nominal operation and uses larger aggregated transistors (up to 90 μm in total emitter length), and the designs are optimized more for large-signal operation farther into gain compression than the first Ka-band design. External DC base resistances are further reduced from 40 Ω down to an extremely low 20 Ω to push operational collector even higher for higher output powers. The extremely low external bias resistance requires that many more wideband MIM capacitor arrays be added for additional bias and transistor stability.

Using feedback from the initial Ka-band design and comparing measurements to simulations, active and passive models are adjusted in back-modeling to better match the measured RF and DC characteristics. In particular, custom thermal models for the aggregated HBT power-cell arrays help to better predict RF & DC performance to optimize the PA further. The improved custom models are used to further optimize the PA at higher collector voltages (1.7 V) where higher output powers are possible, but model inaccuracies and simulation convergence are still highly problematic.

The improved 2-stage PA design uses much larger (+60%) custom HBT power-cell transistors (Q1 = 45 μm & Q2 = 90 μm) and is biased deeper into Class AB operation, resulting in higher efficiency. To achieve higher output powers, the PA interstage & output matching are optimized for operation at higher power powers (around P_{1dB}, further into gain compression).

To push collector voltages closer to BV_{CBO}, external base resistances of the integrated active bias networks are decreased from 40 Ω to 20 Ω, compared to the first design, to further reduce impact ionization effects in the PA DC biasing and allow for higher PA output voltage swings. This is the lowest reported external base resistance to date for integrated active biasing within SiGe mmW PAs. Significantly more on-chip capacitance is added to improve wideband bias stability with integration into a larger system in mind.

At 2.2 V collector voltages, the second iteration of the Ka-band design achieves P_{sat} = 24.2 dBm differential output power at 33 GHz, P_{1dB} = 14.5 dBm, and peak-PAE = 22.3%. The small-signal gain is 14.6 dB with a 3dB-bandwidth from 20.5–36 GHz (54%). The deeper Class AB design trades off linearity for P_{sat} and PAE improvements, as evidenced by the lower P_{1dB}. Small-signal peak gain shifts lower in frequency than the first Ka-band iteration because the PA is matched to operate at high output powers near P_{1dB} and Class AB operation, but the large-signal response is still at 33 GHz.

3. 90 GHz W-Band PA at 1.7–2.3 V

The W-band PA is presented in Chapter 2 and was published in [8]. Leveraging the two Ka-band iterations and custom model improvements from back-modeled simulation results from measured lab data, the 90 GHz W-band PA utilizes the same design approach and concepts at three times the frequency. Scaling the design to 90 GHz requires modifications and new design aspects to address the limited gain (less than 5 dB per gain stage) available.

The W-band design is even more aggressive by pushing PA design to 3× the frequency and operating the HBTs near their active-gain and voltage limits. The W-band design shares similar design approaches and PA topologies. Elements of the Ka-band design are re-used and adjusted accordingly for challenges at 90 GHz. The same 5 μm unit cell is re-used in different aggregated power-cell layouts. The same core active-bias circuit, high-pass matching-network structures, collector transmission lines, and base-biasing inverted transmission line are adopted and optimized for 90 GHz operation.

The extensive DC & RF measurement data from the Ka-band provided valuable insights into model-to-hardware correlations for the W-band design. DC data provided mutual thermal coupling details of the transistor-packing configurations. Knowing the weaknesses in the HBT RF models at high collector voltages and currents, especially around the collector-base

junction, helps to center the large-signal design in the W-band circuits.

Although similar design approaches and some unit cells are carried over from the Ka-band PAs, all W-band PA circuits and layouts are completely re-worked from scratch— no Ka-band layouts are re-used directly.

Major differences between W-band and Ka-band implementations include the following:

- The output transistor size is reduced because of the lack of gain.
- An additional gain stage is added (3 common-emitter gain stages).
- The custom inverted microstrip transmission-line structure is modified to be used as a quarter-wavelength transformer at 90 GHz for base-bias injection (while the Ka-band design use the custom line in series with spiral inductors as non-resonant inductive structures).
- The external DC bias resistance is dropped further to just 20 Ω to push higher operational collector voltages.
- Resonant LC structures (transmission-line stubs terminated in self-resonant MIM capacitors) are utilized for compact interstage matching network that provide the required DC blocking and RF matching reactances.
- Resonant MIM capacitors are used to set the reference low-impedance nodes that are then transformed by the quarter-wave feed lines to high impedances at the bias insertion points.
- Much larger wideband bypass capacitors arrays with many more instances on-chip: de-Q'd MIM capacitor arrays provide low-impedance bypass outside of the operating bandwidth and transistor & bias stability at lower frequencies (< 30 GHz).

The low HBT gain at 90 GHz forces the W-band PA to have another gain stage (3 stages) over that of the Ka-band designs. Specific local-density layout limitations on the deep-trench (DT) layer are lifted (that were present in the Ka-band design) which allows HBT unit-cell transistors to be abutted in 2D arrays rather than 1D arrays separated by > 38 μm. The lifting of the DT rules allows extremely compact 2D aggregated power-cells to be formed, which aids greatly in reducing inter-cell wiring losses and parasitics at 90 GHz. Unmodeled increases in mutual thermal coupling must be modeling and estimated, however, to predict proper DC & RF[3] performance in the large, compact power transistors.

[3]RF, in this text, refers to AC signals and not a specific frequency range lower than mmW (i.e. RF vs. mmW). The terms RF and AC are used interchangeably here, depending on convention.

At the time of publication in [8], this was the highest frequency and highest power silicon PA. It remains one of the highest power non-power-combined linear PAs in silicon technology at 90 GHz.

At optimized bias currents and 2.3 V collector voltages, the baseline design achieves P_{sat} = 19.9 dBm differential output power at 90 GHz, P_{1dB} = 18.8 dBm, and peak-PAE = 15.4%. The small-signal gain is 14.6 dB with a 3dB-bandwidth from 79–97 GHz (20%). The measured noise figure of the PA at 1.7 V is 8.0 dB at 90 GHz.

Other design variations (breakouts) are detailed further in the measurement results in Section 2.11.

4. 90 GHz W-Band 3 × 3 PA Array

Chapter 4 discusses a specific design extension of the W-band PA work presented in [9]. 9 W-band PA elements are fully integrated into a 48 mm^2 SiGe chip and bonded with arrays of antennas on quartz to form a quasi-optical, spatially-combined 3 × 3 PA array. Each element in the array is based on the standalone W-band PA presented in Chapter 2. This PA array work was done in collaboration with Yusuf Atesal, Berke Cetinoneri, Ramadan Alhalabi, and Gabriel Rebeiz at University of California, San Diego.

The personal, primary contribution to this collaboration was to transfer the W-band PA design, determine suitable bias points for reliable array operation, help re-tune design elements to center the design to 94 GHz with an additional gain stage, and assist in stabilizing the low-frequency response of active biasing networks that have long DC and RF-bias feed lines. In addition, extensive RF % DC measurement data from the standalone W-band PA work provided valuable comparisons with the new HiCUM HBT model used in the PA array simulations.

The 3 × 3 PA array operates at 2.0 V & 1.6 A current and achieves a measured EIRP of 33–35 dBm (3.16 W) at 90–98 GHz. (EIRP measurements above 98 GHz are limited by the available source power of the measurement system.) The total on-chip SiGe PA power output of all 9 PA elements is 21–21 dB, and the total radiated power is 17.5–19.5 dBm.

1.1.3 Summary of Contributions

This section summarizes key contributions of the power amplifier research presented in this dissertation.

- This work demonstrates power limits of standalone, highly linear Class A/AB PAs in SiGe at Ka-band and W-band without additional power combining. The PAs are designed with suitable gain (> 10 dB) and PAE (> 10%) for integration into PA

arrays or all-silicon transmitters. The designs follow reliability guidelines for 100,000 power-on-hours (11 years) lifetime at 100°C, a typical target for commercial products.

- At the time of publication, the Ka-band PA was the highest output power silicon PA at 30 GHz with $P_{\text{sat}} = 19.4$ dBm[4] & 11.2% PAE. A previously unpublished iteration of the improved Ka-band PA raised the power to $P_{\text{sat}} = 21.3$ dBm & 13% PAE.

- A second iteration of the Ka-band PA uses much larger transistors (45 μm and 90 μm) and is optimized for powers closer to P_{1dB} operation at 1.7 V. The measured results are $P_{\text{sat}} = 24.2$ dBm @ 33 GHz with $V_{CC} = 2.2$ V, peak-PAE = 22.3%, $P_{\text{1dB}} = 14.5$ dBm, maximum small-signal gain = 14.5 dBm, and 3dB-bandwidth = 20.5–36.5 GHz (16 GHz or 56% BW).

- At the time of publication, the W-band PA was the highest frequency SiGe PA and also the highest output power silicon PA at 90 GHz. To this day, it is still one of the highest output power linear Class A/AB PA in 0.12 μm SiGe BiCMOS at 90 GHz with $P_{\text{sat}} = 19.9$ dBm with 15% PAE. Previously unpublished results from two design variations are also presented: $P_{\text{sat}} = 20.3$ dBm & 15.5% PAE @ 83 GHz; $P_{\text{sat}} = 19.4$ dBm & 12.4% PAE @ 94 GHz.

- At the time of publication, the W-band 3×3 PA array was the highest output power measured of any SiGe PA system, power-combined or otherwise. The spatially-combined array outputs an EIRP of 33–35 dBm at 90–98 GHz.

- This work presents a scalable circuit topology and unit-cell design approach that was successfully scaled from 30 GHz to 90 GHz and integrated into a wafer-scale array. The techniques can be further scaled in advancing SiGe HBT technologies up to $\sim 0.45 \times f_{\text{MAX}}$ frequencies.

- Extremely low external HBT base resistance designs of 20–40 Ω and associated active-biasing circuits are successfully implemented to push operational collector voltages past $BV_{\text{CEO}} = 1.8$ V.

- Unit-cell resonant structures are successfully applied across the entire PA chips. Although designs may shift in small-signal center frequency by 2–3% with process variations, the wideband PA large-signal performance remains well-centered after some

[4]All output powers reported in this section are differential values derived from adding 3 dB to single-ended measurements due to limited available source power and measurement equipment constraints. The PAs have differential inputs and outputs but no measurement baluns on-chip or off-chip for fully differential measurements due to limited available source power to drive the PAs fully into compression at P_{sat}.

tuning via biasing adjustments. Process variations and model deficiencies do not de-tune the PA performance characteristics since critical, sensitive elements track together with $< 0.1\%$ relative matching.

- The PAs achieve high performance despite known modeling and simulation limitations. The original transistor models, based on the VBIC compact transistor model, have convergence issues above 1.5–1.7 V when the PA is pushed into gain compression. Optimizing the biasing and large-signal matching is particularly challenging, and bias voltage & current tuning provisions are incorporated in each gain stage to allow for proper compensation at the measurement bench.

- This work demonstrates that a well-optimized common-emitter topology can deliver as much if not more output power than other Class A/AB linear power amplifier topologies, including cascode implementations within the same technology.

- This work demonstrated that maximum output power is achieved at 2.3 V for common-emitter PAs in SiGe 8HP. Beyond 2.3 V, because of impact ionization effects, there are no additional gains in maximum output power. Certain bias conditions can even result in degradation of maximum achievable output power. These observations could only be found by testing silicon samples in lab and was not characterizable via simulation due to model limitations at these operating limits.

- Custom models are extracted for the mutual thermal coupling of 5 μm HBT unit-cells within densely-packed in 2D arrays. The mutual thermal coupling coefficients are found for the custom aggregated power-cells of each gain stage and applied to HBT models in order to better predict and optimize the DC & RF performance.

- At W-band, this research was the first work demonstrating a power transistor created from densely-packed, minimally-spaced 5 μm npn cells in a 2D array with overlapping deep-trench edges. Ballast resistors, which are typically required for preventing destructive thermal runaway from current hogging on the individual emitter fingers, are shown to be unnecessary in this implementation. This allows for improved RF gain & power since ballast resistors may be omitted with appropriate transistor sizing and interconnects within the aggregated PA cell can be minimized to reduce additional wiring parasitics.

- The PAs are designed specifically with back-modeling in mind. The designs use judiciously-chosen unit cells for transistors, MIM capacitors, bias networks, and certain transmission-line structures. Very similar structures in interstage matching and input

& output matching are also chosen such that complexity of back-modeling is reduced. This methodology also ensures that modeling errors and process biases affect the PA in a way that does not severely degrade performance. For example, the design uses similarly-parameterized transmission lines, nearly-identical bias networks, similarly-scaled power transistor layouts, and similarly-configured wideband bypass capacitor arrays.

The PA back-modeling is used to refine models for more accurately simulating large-signal performance, especially at higher collector voltages above 1.7 V, and allowing design optimization in follow-on designs. For example, the back-modeling refines the HBT transistor model of the custom power transistors created by aggregating densely-packed 2D arrays of smaller HBT unit cells. Custom modeling around the foundry-supplied model accounts for unmodeled mutual thermal coupling and adjusts collector-to-base capacitance C_{CB} to fit observed impact-ionization & thermal effects in the measured PA. By performing back-modeling in-situ at the PA level rather than an individual device level, the back-modeling is simplified to a more manageable problem focused on parameters that have greatest impact on the performance metrics of interest. The approach of using unit cells and similar circuit structures greatly reduces the complexity of back-modeling. These resulting model refinements were used for improving the subsequent PA designs— the second Ka-band iteration and then the scaled W-band design.

- The PA design includes integrated on-chip active biasing with low DC impedances presented to the base terminals of the power transistors. The 20 Ω external base resistance is the lowest resistance used to date with on-chip active biasing for SiGe mmW PAs. This low external base resistance allows transistor collector DC voltages to be biased above $BV_{\mathrm{CEO}} = 1.8$ V and peak RF/AC voltages to approach $BV_{\mathrm{CBO}} = 6.0$ V resulting in higher output powers in common-emitter configuration.

  Active on-chip biasing with current-mirrored references allows for fine bias tuning in the lab environment, even at high collector voltages, e.g., $V_{\mathrm{CC}} = 2.3$ V, where setting the bias currents becomes extremely sensitive. Fine tuning is required during measurements because the HBT transistors are being used in a region where they are not accurately modeled (i.e., densely packed in a 2D array and at high collector voltages & currents). Since each gain stage's power transistor size and layout is different, each stage must have provisions to tune the bias points independently.

- A novel stacked M_{234} (stacked M2, M3, and M4 metals) transmission line is designed for the quarter-wavelength base-biasing feeds using metal layers M1, M2, M3, and

M4, which are thin lower-level copper layers not typically meant for RF transmission lines. The custom microstrip line runs beneath the global RF ground shield, allowing it to route underneath transmission lines above the ground shield, facilitating compact, differential layout while presenting the necessary DC and RF external impedances to the power transistor's base terminal: a very low external DC resistance (20 Ω) and a high RF impedance (> 120 Ω) via quarter-wave transformation network.

- For each of the Ka-band and W-band designs, a die was tested for over 500 power-on hours from 1.7–2.3 V at room temperature under various RF input powers with no measured degradation in RF performance. The design follows foundry guidelines for 100,000 power-on hours at 100°C, e.g., transmission-line sizing, metal-via density for transistors and capacitors, and power-transistor wiring.

- Noise figure is measured for the W-band PA; this is the first known noise measurement done on a SiGe W-band PA. The measurements show that the PA has comparable noise figure (8–9 dB NF) to dedicated low-noise amplifiers near 90 GHz in the same SiGe 8HP technology. Although the high power dissipation of the PA makes it unpractical to use as a standalone low-noise amplifier (LNA), this measurement demonstrates that the PA contributes little noise to a transmitter chain.

- The output power at 1 dB gain compression, P_{1dB}, is characterized over a large bandwidth. These were the first published SiGe mmW PA works in W-band and Ka-band to provide P_{1dB} results over frequency. P_{1dB} provides an important reference to expected large-signal linearity. (Its measurement is especially time-consuming at W-band with manually-tuned frequency & power sources.)

- The PA design is ready for integration into a larger system, as demonstrated by the 3×3 PA array. Numerous bypass capacitor arrays provide wideband stability. Additionally, matching networks and biasing networks are designed to help stabilize the transistor response at low frequencies (below 10 GHz) where the HBTs intrinsically have high RF/AC gain. Integrated active-biasing networks provide stable bias references for the power transistors.

  The PA architecture has built-in electrostatic discharge (ESD) protection on the RF input and output pads because of the shunt-stub high-pass matching network designs. The PA can drive off-chip loads such as antennas while being ESD-compliant at the RF I/O. Typically, ESD protection on RF I/O is omitted due to severe degradation of RF performance, especially at mmW frequencies.

Table 1.1: Comparison of GlobalFoundries SiGe BiCMOS technology nodes; HP = high-performance HBT, HB = high-breakdown HBT [1–5]

		HP				HB			
Process	**Lith (μm)**	f_T **(GHz)**	f_{MAX} **(GHz)**	$BV_{\mathbf{CEO}}$ **(V)**	$BV_{\mathbf{CBO}}$ **(V)**	f_T **(GHz)**	$f_{\mathbf{MAX}}$ **(GHz)**	$BV_{\mathbf{CEO}}$ **(V)**	$BV_{\mathbf{CBO}}$ **(V)**
5HPE	0.35	43	79	3.3	11.0	17	38	9.6	23.0
6HP	0.25	47	88	3.35	10.5	27	70	5.7	14.0
7HP	0.18	120	100	1.8	6.4	27	57	4.25	12.5
7WL	0.18	60	120	3.3	11.0	29	90	6.0	16.0
8WL	0.13	100	200	2.5	8.75	54	170	4.7	16.0
8HP	0.13	210	265	1.8	6.0	60	225	3.55	12.0
8XP	0.13	250	330	1.65	5.65	67	225	3.30	11.7
9HP	0.09	300	360	1.7	5.30	135	350	2.5	8.0

Silicon Foundry Technology

GlobalFoundries' fifth-generation 90 nm SiGe BiCMOS 9HP process (f_T = 300 GHz & f_{MAX} = 380 GHz) recently became available for limited-access design prototyping. The faster GF 9HP HBT transistors provide approximately 3 dB more available gain at 100 GHz compared to GF 8HP and allow the 9HP HBTs to be used for PAs above 110 GHz in D-band. SiGe foundry technology continues to advance in 4-year development cycles. Table 1.1 outlines the current five generations of GlobalFoundries SiGe BiCMOS technology [5] Table 1.2 presents the current ITRS roadmap for future SiGe HBT technologies [29].

Table 1.1 compares key transistor metrics of five generations of GlobalFoundries' SiGe BiCMOS technology.

SOI CMOS development, driven by the need for extremely low-powered digital-logic ICs for mobile devices, has seen aggressive foundry technology development. Commercial CMOS & SOI processes now achieve similar f_T/f_{MAX} values to SiGe processes [30–32]. SOI processes have been used to design successful mmW PAs. The low-resistivity of bulk CMOS makes mmW design significantly more challenging that SOI, but successful designs have also been demonstrated [33–35].

In FET devices, for both bulk CMOS and SOI devices, the f_T metric scales approximately

[5]Each generation is prefixed by the same number: 5 (1st), 6 (2nd), 7 (3rd), 8 (4th), and 9 (5th). 6HP is a laterally-scaled version of 5HPE with similar HBT performance (f_T).

Table 1.2: ITRS technology roadmap for SiGe HBT development

Years	**Lith (nm)**	**Emitter Length (nm)**	f_T **(GHz)**	$f_{\mathbf{MAX}}$ **(GHz)**	$BV_{\mathbf{CEO}}$ **(V)**	$BV_{\mathbf{CBO}}$ **(V)**	**Peak f_T Current Density (mA/μm^2)**
2014–2016	115	1150	282	432	1.70	4.9	14.6
2017–2020	88	880	394	557	1.65	4.5	23.3
2021–2024	66	660	502	766	1.55	4.3	33.1
2025–2028	44	440	641	1087	1.55	4.1	71.0
2029	22	220	760	1470	1.05	3.9	118.5

inversely with transistor nFET gate length [36]:

$$f_T \propto \frac{1}{L_{gate}} \tag{1.1}$$

The results of device scaling on f_T and supply voltage (V_{DD}) is summarized for bulk CMOS and SOI devices in Table 1.1.3. Peak f_T of bulk CMOS nFETs is roughly half that of SOI.

From an f_T figure-of-merit standpoint, 120 nm SiGe HBTs have achieved comparable performance to nFETs two technology nodes ahead in 65 nm bulk CMOS [37]. A process's peak f_T/f_{MAX} is usually quoted for the core transistor without interconnects and other wiring parasitics. Operation near these quoted f_T/f_{MAX} figures-of-merit can be more easily approached with digital circuits and some mixed-signal circuits where interconnect wiring is done primarily on low-level metals. However, when transistor devices are examined in the use context as RF devices that require wiring to top-level metals of low-loss transmission lines, f_T of SOI devices drop significantly more than those of SiGe HBTs. 130 nm and 90 nm SiGe HBTs have been measured to have considerably higher f_T/f_{MAX} than 32 nm SOI once wiring to top-level metal is considered [38].

A significant challenge for SOI PAs is the low DC drain-source voltages allowed on the FETs. For the same f_{MAX} as a BiCMOS process, the allowed drain-source voltages are much smaller in SOI. For instance, nFET devices in 14–45 nm SOI CMOS operate at $V_{\mathrm{DD}} = 0.8$–1.0 V and RF swings across any two FET junctions are limited to $2 \times V_{\mathrm{DD}} = 1.6$–2.0 V maximum for long term reliability [39, 40]. To overcome the low voltage limitation, stacked FET topologies, usually in nonlinear switch-mode PA configurations, are used to increase the output power out of the FETs. Additional output power combining can also be applied to the stacked-FET switch-mode PA.

Table 1.3: Bulk CMOS & SOI Scaling Effects

Technology	**Lithography (nm)**	f_T **(GHz)**	**Supply Voltage (V)**
Bulk CMOS	130	90	1.2
	90	120	1.2
	65	160	1.2
	55	190	1.2
	40	260	1.1
	28	310	1.0
SOI	90	250	1.2
	65	320	1.2
	45	485	1.0
	32	550	0.9
	22	900	0.8
	14	1400	0.8
	12	1600	0.75

For the same area, prototyping costs of 90 nm SiGe BiCMOS and 45 nm SOI CMOS technologies are currently nearly double the cost of 130 nm SiGe BiCMOS. GF SiGe 8HP and other 130 nm ($f_T \approx 200$ GHz) SiGe technologies are still widely used today for research in switch-mode (Class E/F) mmW PAs and especially in large-area, highly integrated RF circuits at 24 GHz, 30 GHz, 60 GHz, 77 GHz, and 90 GHz [41–50]. Linear-mode Class A/AB mmW PA research has moved to 90 nm SiGe BiCMOS (9HP) and frequencies above 110 GHz, extending mmW linear PA techniques developed from W-band & 130 nm SiGe BiCMOS (8HP).

Relevancy to Other SiGe mmW PA Trends

This dissertation work demonstrates the limits of linear Class A/AB output powers and frequencies in GF 8HP without power combining. Subsequent research in the mmW silicon PA field includes

- Extending highly-linear Class A/AB mmW PAs techniques to higher frequencies ($>$ 100 GHz) within new 90 nm SiGe BiCMOS processes such as GF 9HP ($f_T/f_{\mathrm{MAX}} =$ 300/360 GHz).
- Developing more power-efficient and area-efficient on-chip, off-chip, and multi-chip power-combining techniques.
- Using switch-mode (nonlinear) PA classes & topologies to increase output power and

efficiency of mmW PAs using SiGe BiCMOS (130 nm & 90 nm) and advanced SOI CMOS technologies (45 nm and 32 nm).

Because the design methodology presented in this dissertation work is scalable, the PA design techniques can be applied to new technology nodes that push even higher frequencies of PA operation. Later follow-on research by lab colleagues using 90 nm SiGe (GF 9HP), adopting a similar PA design technique presented here, extends Class A/AB linear output powers to 25.3–27.3 dBm at 68–88 GHz [51] and 20–20.8 dBm at 114–127 GHz (D-band) [52] by using large-scale 16-way and 8-way reactive on-chip power combining, respectively.

After various works, including the ones presented here, successfully demonstrated the output-power & PAE limits of highly-linear Class A/AB design techniques in mmW PAs ($\sim$ 30–100 GHz), the focus below 100 GHz has shifted to achieving even higher output powers with a strong emphasis on efficiency at the expense of PA linearity. To achieve high output powers ($>$ 20 dBm) with high efficiencies (PAE $>$ 30%) at mmW, highly nonlinear switch-mode PA classes of operation (e.g., Class E, Class F) have been employed. Switch-mode PA classes are designed to operate with constant-envelope, constant-amplitude signals (e.g., PSK) that use the nonlinear PA near its saturated output power level where PAE is maximum. To use switch-mode PAs for signals with amplitude variation, however, significant system-level complexity must be added around the PA block to linearize the response; some techniques even require multiple PAs. Issues of making a highly nonlinear PA usable are pushed into the supporting baseband-to-RF transmitter architecture. Digital assistance and linearization techniques such as digital pre-distortion, feed-forward, and envelope elimination & regeneration must be applied around the nonlinear PA block [11, 53].

The high-volume, commercial feasibility of fully-integrated mmW silicon PAs in high data rate communications systems (e.g. 5G wireless) has yet to be determined. Class A/AB SiGe PAs are highly linear, but require additional lossy power-combining techniques to achieve higher output powers at the expense of efficiency. Switch-mode Class E SOI CMOS and SiGe PAs, on other other hand, require complex & sensitive linearization schemes and transmitter optimizations to make the PA usable with envelope-modulated signals like high-order m-ary QAM (quadrature amplitude modulation) and OFDM (orthogonal frequency-division multiplexing).

Linear Class A/AB SiGe PAs are still quite relevant in the development of future mmW systems. For instance, Class A/AB designs are more easily integrated into multi-element transmit arrays since they do not require many supporting circuits to linearize the response. In addition, SOI CMOS designs have yet to prove their long-term reliability in PAs because of severe hot-carrier degradation effects.

Ultimately, decisions around partitioning of a mmW transceiver will determine what

technology or technologies are chosen. In applications where high linearity is a priority over efficiency (base stations and access points) III-V GaAs, GaN, or InP chips may still be required for the final high-power PA stage. In this case, highly linear Class A/AB SiGe PAs still may be preferred to drive III-V PA rather than more non-linear Class E SiGe stacked-HBT or Class E SOI stacked-FET solutions. In applications where efficiency is the priority (mobile, low power), SOI Class E PAs have the advantage of efficiency and integration with mixed-signal & digital circuits on the same substrate/chip.

1.2 Introduction to mm-Wave SiGe PAs

Millimeter-wave (mmW) frequencies are defined by the International Telecommunication Union (ITU) as the extremely high frequency (EHF) band covering radio frequencies from 30–300 GHz, which corresponds to free-space wavelengths between 1–10 mm. For this dissertation, power amplifier (PA) designs are implemented in commercially-available silicon geranium HBT technology within two millimeter-wave (mmW) frequency bands: Ka-band (26.5–40 GHz, λ_0 = 11.1–7.5 mm) and W-band (75–110 GHz, λ_0 = 4.0–2.73 mm). Popular W-band applications include collision-avoidance radar (94 GHz), medical & security imaging (85–110 GHz), automotive radar (76–81 GHz), point-to-point wireless communications (26–80 GHz), and last-mile wireless backhauls (75–95 GHz) [1]. Popular Ka-band applications include mobile communications (including possible new 5G cellular radio standards), satellite communications, radar, imaging, and point-to-point wireless backhauls. Atmospheric windows within Ka-band and W-band provide frequencies where there is lower attenuation from atmospheric oxygen and water vapor gases: 0.1 dB/km around 30 GHz and 0.4 dB/km around 90 GHz at sea level [54].

Silicon germanium BiCMOS technology offers major advantages in the system integration of complex transceivers because of its ability to utilize large areas and integrate many types of circuits, including power amplifiers. SiGe BiCMOS allows for low-cost, high-volume integration of RF, analog, mixed-signal, and digital circuitry all on a single chip. Sophisticated power control, built-in self-test, and calibration & tuning can also be fully integrated.

Since higher frequencies in mmW correspond to shorter wavelengths, compact transmission-line structures and multi-element arrays can be integrated easily onto a single silicon substrate. The guided wavelength within silicon dioxide (SiO_2) is roughly half that of the free-space wavelength:

$$\lambda_{guided} = \frac{\lambda_0}{\sqrt{\epsilon_r}} \tag{1.2}$$

where $\epsilon_r = 4.1$ for silicon dioxide.

Distributed transmission-line & passive structures (e.g., inductors) typically dominate the area of mmW IC designs, so higher frequency mmW circuits shrink the size of the silicon chip. Additionally, the larger reticle areas of advanced silicon processes (e.g., 18 × 20 mm^2 for 8HP, 25 × 30 mm^2 for 45 nm 12SOI/45RFSOI) allow very high integration of mmW circuits, from RF to mixed-signal to digital circuits within a single chip, which is not possible in III-V technologies. The large wafers of silicon processes (200 mm/8 inches in 8HP; 300 mm/12 inches in 12SOI) also allows wafer-scalable designs to utilize entire wafers for applications such as large-scale multi-element phased arrays in silicon. A 200 mm wafer can support stitching up to 60 reticles while a 300 mm wafer supports up to 100. Examples of recent successful wafer-scaled phased arrays in silicon can be found in [55].

Advanced silicon technologies, including SiGe and SOI CMOS, create opportunities to open up mmW frequencies to high-volume consumer wireless applications where bandwidth & capacity are needed urgently. Prior to the recent advances in mmW silicon technology, III-V technologies (e.g. GaAs/InP HBT, HEMT, MESFET) were the only practical options to design for commercial applications above 24 GHz.

For telecommunication systems, moving from sub-6 GHz to mmW frequencies provides much-needed bandwidths to support the rapid growth of mobile data consumption, currently driven by streaming video and cloud-based data storage & services. The larger bandwidths available at higher frequencies provide larger channel capacity, as shown by Shannon's capacity theorem for an additive white Gaussian noise (AWGN) channel:

$$C = BW \log_2 (1 + \text{SNR}) \tag{1.3}$$

where C is the channel capacity in bits-per-second (bps), BW is the bandwidth in Hz, and SNR is the signal-to-noise ratio (unit-less). The increased available bandwidths at mmW frequencies is utilized to increase, significantly, maximum data rates per user and the total number of users supported within a channel.

For wireless communications systems above 60 GHz, the PA phase-noise and linearity requirements are a somewhat relaxed since the greater bandwidth available allows for fewer bits per Hertz and hence lower-order constellations [56]. For radar and imaging systems, moving to higher mmW frequencies provides more resolution in ranging and imaging needed to support increased automation, e.g., autonomous vehicles and automated manufacturing.

1.2.1 mmW Power Amplifiers

A power amplifier is a large-signal amplifier that pushes beyond pure small-signal operation with weak nonlinearities toward the limits of a transistor where strong nonlinearities and

clipping effects dominate. RF power amplifiers often push the transistor to the extremes of allowable voltage and current ranges ($V_{\max}$, $I_{\max}$), limited by transistor breakdown voltages and electromigration current limits.

Power amplifiers are normally the highest power-consuming block in RF transceivers. However, obtaining high output powers in silicon processes is extremely challenging because of silicon's inherent lower bandgap energy (Si: 1.12 eV, Ge: 0.66 eV) relative to III-V materials (GaAs: 1.424 eV, InP: 1.344 eV). Fully-integrated SiGe power amplifiers (PAs) are particularly attractive for phased-array and point-to-point communication systems where a full silicon system-on-chip solution can replace complex GaAs-based T/R modules [57]. This results in lower-cost and higher-reliability solutions for mmW systems. SiGe technology has demonstrated the capability of generating significant millimeter-wave linear (Class A/AB) power amplification within W-band (75–110 GHz) with reasonable efficiencies [14, 21, 26, 58–60].

High output powers and high efficiencies are difficult to achieve in SiGe PAs at mmW frequencies because of low available transistor gain and low breakdown voltages in HBTs. Typically, commercial designs of high-performance RF circuits at frequency f_0 target technologies with f_T & $f_{\mathrm{MAX}} > 20\times$ the operating frequency. High f_T/f_0 ratios allow RF circuit solutions to have low complexity and high performance in terms of gain, linearity, noise, and efficiency. For mmW circuits in SiGe, the limited gain is an important design challenge, especially for PAs, which typically require gain compression to achieve maximum output powers & PAE.

The maximum available HBT common-emitter small-signal gain can be estimated from the f_{MAX} [56]:

$$G_{\mathrm{MAG}}(dB) \approx 20\log_{10}\left(\frac{f_{\mathrm{MAX}}}{f_0}\right) \tag{1.4}$$

This estimation is valid above frequencies where the HBT is unconditionally stable (Rollett[6] $K > 1$). MAG drops at a 20 dB/dec slope (-6 dB/oct). For GF 8HP, the MAG estimation in Eq. (1.4) is valid above $\sim$ 20–40 GHz, depending on bias conditions & parasitics, or equivalently, > 0.10–$0.15 \times f_{\mathrm{MAX}}$. At lower frequencies where $K < 1$ and transistor stability is not guaranteed, maximum stable gain (MSG) is quoted. MSG increases at 10 dB/dec slope (-3 dB/oct) with decreasing frequencies below $\sim$30 GHz in 8HP.

Low transistor gain at mmW frequencies necessitates additional PA gain stages, which generally leads to more complexity &area (more biasing, more interstage matching), increased power consumption, more nonlinearity, less efficiency, less design margin for pro-

[6]See Appendix A

cess/voltage/temperature (PVT) variations. Low transistor gains also leave fewer options to trade off any excess transistor gains for increases in circuit bandwidths.

GF SiGe 8HP integrates high-breakdown (HB) HBT devices alongside high-performance (HP) HBTs by selectively doping collectors with lighter doping profiles. High breakdown devices have collector-emitter breakdowns between $BV_{\mathrm{CEO}} = 3.55$ V to $BV_{\mathrm{CBO}} = 12.0$ V depending on the external bias conditions. The reduced collector doping decreases the f_T of high-breakdown devices considerably to 60 GHz. Although the f_{MAX} of the high breakdown device is still 225 GHz, the low f_T indicates that the device is not able to provide significant current gains at mmW frequencies > 24 GHz, which limits the range of useful RF output impedances that the HBT can drive beyond 24 GHz.

1.2.2 Power Amplifier Metrics

This section covers some key system-level metrics of power amplifiers such as bandwidth, gain, P_{1dB}, P_{sat}, PAE, and linearity.

Frequency and Bandwidth

Gain

The large-signal gain of a PA— gain when output powers are high relative to small-signal levels— can be more or less than small-signal gains, depending on the PA class and design details. Purely Class A PAs have large-signal gains that are close to small-signal gains until the large-signal gain begins compressing near P_{1dB}) output levels. The large-signal gain decreases as the PA begins to saturate. Class AB amplifiers may have large-signal gains that are larger than their small-signal gains, depending on the PA biasing points (current, mostly) and resulting conduction-angle reduction as well as gain expansion with increasing input powers.

P_{1dB}

P_{1dB} is the output power[7] when the small-signal gain compresses by 1 dB. For Class AB amplifiers that experience gain expansion, P_{1dB} is the output when when the peak small-signal gain compresses by 1 dB, which gives the most conservative P_{1dB} values.

Fig. 1.1 illustrates the P_{1dB} point relative to the gain (slope) of the P_{out} versus P_{in} curve of the amplifier.

[7]Sometimes OP_{1dB} is used to differentiate between input & output P_{1dB}, but all references in this text will refer to P_{1dB} as the *output* 1 dB gain-compression point.

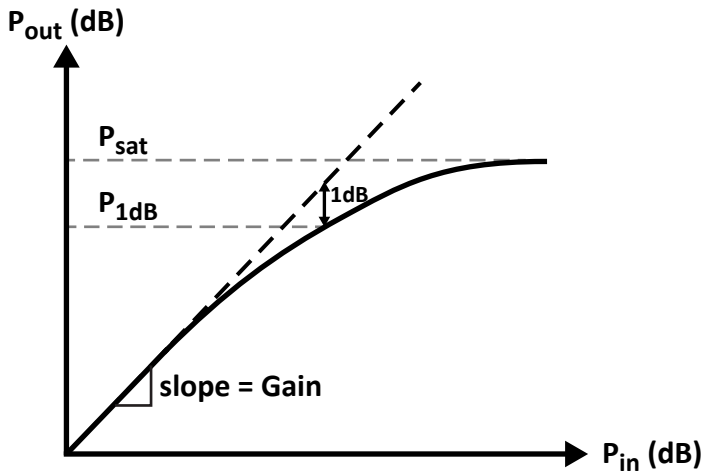


Figure 1.1: Illustration of P_{1dB}, gain, gain compression, and P_{sat} for a conceptual P_{out} versus P_{in} amplifier response.

P_{1dB} can provide a rough estimate of linearity performance in terms of third-order intermodulation products (IM3) [61].

$$\mathrm{IIP}_3 = \mathrm{IP}_{\mathrm{1dB}} + 9.6 \text{ dB} \tag{1.5a}$$

$$\mathrm{OIP}_3 = \mathrm{OP}_{\mathrm{1dB}} + 10.6 \text{ dB} \tag{1.5b}$$

$$\mathrm{IP}_{\mathrm{1dB}} = \mathrm{OP}_{\mathrm{1dB}} - [G_{\mathrm{ss}}(\mathrm{dB}) - 1 \text{ dB}] \tag{1.5c}$$

where G_{SS} is the small-signal gain— the amplifier under small RF input powers such that there is no gain compression.

The estimate is only a rough rule-of-thumb for simple weakly nonlinear systems up to third-order nonlinearities since amplifiers and cascaded amplifiers with strong nonlinearities (clipping, higher-order nonlinearities and harmonics re-mixing and interacting, higher-order nonlinearities, memory effects) affects the ratio and requires dedicated Volterra series modeling for the circuit interactions, but looking at output P_{1dB} for linear Class A/AB amplifiers gives a rough idea of nonlinearity. Near P_{1dB}, the large-signal nonlinearities are different than the small-signal nonlinearities due to saturation and hard clipping. Thus, OIP3 can be smaller than $P_{\mathrm{1dB}} + 10$.

Linearity at PA output powers must ultimately be measured with real, modulated signals and error-vector magnitude (EVM) measurements or adjacent channel power ratio (ACPR) from spectral re-growth when PA is operating near maximum output powers. Small-signal

extrapolations of small-signal two-tone IP3 does not fully capture the actual linearity of the PA with real signals near compression.

$P_{\mathbf{sat}}$

P_{sat} is the maximum output power, which occurs when the Class A/AB PA is in heavy gain compression: typically > 3 dB gain compression compared to small-signal values. At P_{sat} output power, increasing the PA's RF input power does not result in additional output power— the output power saturates.

Efficiency

Peak PAE is the power-added efficiency at its maximum value (across output powers), which usually occurs past P_{1dB} output powers in Class A/AB PAs. In highly linear Class A/AB PAs, as input power P_{in} is increased from small-signal levels, P_{out} and PAE also increase. PAE peaks, depending on design, at a particular gain-compression point. Increasing the RF input power to drive a PA further into compression beyond peak PAE results in more P_{out} output power, but less PAE due to the dropping gain (from amplifier compression).

Many common definitions for power amplifier efficiency, η, are used depending on the system considerations and priorities [6], and a few of the most commonly quotes metrics are discussed here.

The simplest metric of efficiency is η in Eq. (1.6) which only considers the overall DC to RF-output-power conversion efficiency. The DC power dissipation is that of the entire PA, biasing circuits and all gain stages. RF input-power requirements are not considered.

$$\eta = \frac{P_{\text{RF,out}}}{P_{DC}} \tag{1.6}$$

The power-added efficiency (PAE) in Eq. (1.7) is the most relevant efficiency metric for mmW PAs since amplifier gain is very limited at high frequencies (i.e., < 10 dB per gain stage), and the required input powers are non-negligible. PAE takes into account the RF input power requirements to drive the PA into high output-power operation.

$$\eta_{\text{PAE}} = \frac{P_{\text{RF,out}} - P_{\text{RF,in}}}{P_{DC}} \tag{1.7}$$

PAE is directly affected by the system gain, and Eq. (1.7) can be re-written in terms of operating power gain factor (non-dB), G, as shown in Eq. (1.8).

$$\text{PAE} = \eta_{\text{PAE}} = \frac{P_{\text{RF,out}} - P_{\text{RF,in}}}{P_{DC}} = \frac{P_{\text{RF,out}} - \frac{P_{\text{RF,out}}}{G}}{P_{DC}} = \eta_C \left(1 - \frac{1}{G}\right) \tag{1.8}$$

An alternative definition of efficiency is overall or total efficiency, which also takes into account the power required for the RF input. Unlike PAE, which can be negative, the lower bound on total efficiency is 0% at low PA gains.

$$\eta_{\text{overall}} = \eta_{\text{total}} = \frac{P_{\text{RF,out}}}{P_{DC} + P_{\text{RF,in}}} \tag{1.9}$$

When PA powers gains are high, the required RF input powers become negligibly small, and the various efficiencies all converge to

$$\eta = \frac{P_{\text{RF,out}}}{P_{DC}} \approx \eta_{\text{PAE}} \approx \eta_{\text{total}} \approx \eta_{\text{overall}} \qquad \text{(high gain)} \tag{1.10}$$

For multi-stage mmW PAs with transistors operating near their frequency limits, low gain impacts the overall cascaded, system PAE. The cascaded PAE of a three-stage power amplifier is approximately [20]

$$\eta_{\text{PAE,cascaded}} \approx \frac{1}{\frac{1}{\eta_{\text{PAE3}}} + \frac{1}{G_3\eta_{\text{PAE2}}} + \frac{1}{G_3G_2\eta_{\text{PAE1}}}} \tag{1.11}$$

where G_3 is the power-gain factor (non-dB) of the third gain stage (i.e., 6 dB gain equates to $G = 4$) and η_{PAE2} is the PAE efficiency factor of the first gain stage (i.e., 10% PAE equates to $\eta_{\text{PAE1}} = 0.10$). The PAE of the final output stage has the largest affect on the cascaded 3-stage PAE and is the upper bound of the realizable cascaded PAE.

Since most of the overall PA system efficiency is typically limited by the output-transistor stage, examining the efficiency of the output transistor provides valuable insight into how efficient the final-stage output transistor is operating with respect to the output load line. Collector efficiency (or drain efficiency, for FETs) is a measure of the output transistor's efficiency, ignoring RF input power requirements, biasing circuit power dissipations, and all power dissipation in preceding gain stages. Collector efficiency η_C is given in Eq. (1.12).

$$\eta_C = \frac{P_{\text{RF,out}}}{P_{DC,\text{output transistor}}} \tag{1.12}$$

The figure of merit only considers the output transistor and the losses of the output transistor and output-matching network. The maximum efficiency, theoretically, is $\eta_C = 0.5$ or 50% for purely Class A operation with no output-matching losses. Moving the PA into Class AB and higher-order operation increases the theoretical limits of η_C closer to 1.00 (100%).

Table 1.4: Impact of communication signal types with various peak-to-mean-envelope power ratios (PMEPR) on maximum achievable for ideal linear Class A power amplifiers [6].

Signal	**PMEPR (dB)**	**Class A PAE (%)**
FSK (MSK, GMSK)	0	50
QPSK	3.6	21.9
$\pi/4$ DQPSK	3.0	25.1
OQPSK	3.3	23.4
8-PSK	3.3	23.4
64-QAM	7.8	8.3

1.2.3 Linearity

Signals with constant-amplitude envelopes— no amplitude variations—, such as continuous wave (CW), frequency modulation (FM), frequency-shift keying (FSK), and minimum-shift keying (MSK) do not require high PA linearity since the signals only use phase modulation, and power amplifiers can be optimized around this key operating condition. Examples of constant-envelope signals include frequency-modulated continuous-wave (FMCW) automotive and short-ranged radar at 24 GHz and 77 GHz.

Linear amplification is required, however, when the signal amplitude has an envelope. Most modern high data-rate digital communications systems employ signals that contain both amplitude and phase modulation, such as QPSK, m-ary QAM, and OFDM. Unlike the constant-envelope case, amplifier linearity becomes critically important to the system performance.

Effects of PMEPR[8] (peak-to-mean-envelope power ratio) on amplifier efficiency is shown in Table 1.2.3.

Even systems that use constant-envelope modulation like QPSK or PSK may need significant power amplifier linearity if multiple channels are aggregated. Multichannel communications systems like OFDM may use orthogonal subcarriers that have constant-envelope modulations on the subcarriers, but the composite signal of the combined subcarriers results in a signal waveform that contains an amplitude-modulate envelope. Every doubling of combined sub-carriers increases the PMEPR/PAR by +3 dB [62]. In other words, for PAs,

[8]PMEPR is also known as peak-to-average power ratio (PAR) to circuit designers.

additional power back-off (away from the onset of compression) is given by

$$\text{PMEPR (dB)} = 10 \log_{10} N \tag{1.13}$$

where N independently-modulated carriers are combined into a multi-carrier signal.

For a fixed data rate in a system, PA linearity requirements reduce as the operating frequency increases because available bandwidth typically increases with frequency, as well. Lower-order, less complex modulation schemes with less power-back-off requirements (linearity) can be used to achieve the same data rates.

1.2.4 Linear Classes of Operation

The system-level transceiver design and intended signal modulation help determine a PA configuration and operating class: biasing, input drive, output loads, and harmonic terminations. For this work, the goal is to design highly linear Class A/AB amplifiers as outlined previously in Section 1.1.1.

In linear-mode amplifiers (Class A, AB, B— non switch-mode), the choice PA DC biasing affects the overall performance.

Class A offers the best linearity, small-signal gain & bandwidth, but lowest power efficiency. Biasing a PA into Class AB offers better efficiency at the expensive of gain and linearity relative to pure Class A. One means of increasing the overall PAE of a 3-stage PA is to decrease the DC biasing of the final output stage into Class AB (more efficient, but less gain & linearity) while the earlier gain stages are maintained in Class A (less efficient, more gain, and more linear).

Other classes of PAs can offer even higher efficiencies over Class AB, but their small-signal gains become very small and will drop significantly as input signal levels decrease. Small-signal gains drop as quiescent biasing drops deeper into Class AB with further-reduced conduction angles. For signals with amplitude envelopes, the nonlinearity must be must be corrected via supplemental circuits like envelope elimination & restoration (EER), feed-forward, outphasing, pre-distortion, and digital signal processing. If a multi-stage PA is designed to have high efficiency but very low total gain ($<$ 10 dB gain from PA input to output), then the problem of generating enough input power to drive the PA is being "pushed back" into a preceding amplifier driver. The PA driver would need to deliver more input power to the low-gain PA (versus a high-gain PA), and this impacts the overall achievable PAE of the entire transmitter chain— PA and input drivers.

Class A

In Class A operation, the HBT operates as a current-dependent current source. Of all PA classes, pure Class A offers the best small-signal gain, bandwidth, and linearity & harmonic distortion at the cost of power efficiency. Increasing the quiescent DC current or decreasing the input drive power decreases the distortion levels and harmonics seen at the PA output. Class A is useful for pushing a highly linear PA to high frequencies where transistor gain falls off quickly with frequency (toward f_T/f_{MAX}[9]).

Amplifier linearity is required when the signal contains amplitude modulation (varying amplitude envelopes) either alone or in combination with phase modulation schemes.

Other PA classes may have to sacrifice small-signal gain and linearity in order to preserve gain under large-signal conditions.

Purely Class A designs operating far from output compression can rely on traditional linear amplifier design approaches based on small-signal S-parameters. Such designs are common in measurement lab amplifiers where high linearity is a priority and minimal wave distortion is desired. Efficiency is not a particular concern. Gain, noise, and linearity within are usually more important rather than output power and PAE. Most useful power amplifiers in mobile communications systems, however, put great emphasis on output power and PAE (and require that the peak signal envelope reach close to the saturated output power levels).

With little gain at W-band, Class A approach preserves the most small-signal gain. Whether small signal gain and linearity at small input levels are required depends on the system design. If the system requires uniform gain over wide input powers, Class A or mild Class AB are appropriate. If the system operates the PA at a fixed high output level (or always near compression or saturation), then deeper Class AB or Class B may be more appropriate.

The large-signal design of purely Class A power amplifiers closely mirrors small-signal approaches using S-parameters as long as device constraints (maximum voltage, current, and power dissipation) are properly taken into account.

Class AB and Overdriven Class A/AB

Decreasing the standing quiescent current such that a PA operates in Class AB trades off some linearity & small-signal gain for higher efficiency.

Conduction angle is between 360° and 180°, and the output transistor is one for more than half the cycle buy less than the full cycle of a sinusoidal input signal. The output conduction angle is a function of the input-signal power level. Near peak amplitudes, the output signal

[9]In this text, all instances of f_T/f_{MAX} are to designate "f_T or f_{MAX}" and not "f_T divided by f_{MAX}".

is distorted, compared to purely linear Class A.

Gain compresses as input powers are increased because the output voltage begins to saturated and clip.

In Class AB operation, the quiescent bias point is reduced such that the negative side of the swing cuts off the transistor (i.e., current is turned off). The conduction angle is a function of drive amplitude as well as quiescent bias point [63]. BJTs biased down into deep Class AB tend to show low small-signal gains and then a great deal of gain expansion at higher input drives. Class AB is a slightly nonlinear amplifier, depending on the operating regime. The second harmonic dominates the harmonic content outside of the fundamental.

As operation pushes deeper into Class AB (at reduced conduction angles closer to Class B), the power gain of the device drops as well as the linearity (distortion and harmonics). Deep Class AB[10] (closer to Class B) operation typically trades off 3–6 dB of small-signal gain, for instance, but efficiency over wide input powers and peak efficiency can be increased over operation closer to Class A since the DC quiescent current is reduced.

Conduction Angle and Harmonic Content

Harmonic content arises from the chosen conduction angle as well as input drive level (input power).

Fig. 1.2 illustrates the amplitude of harmonics in output currents generated through various conduction angles[11] [63].

Purely Class A operation occurs when the conduction angle is a full 360°: the transistor acts as a linear controlled collector-current source and is always on & conducting an output collector current. The Class A output transistor is always conducting during the full cycle of the input signal. Class AB is defined anywhere between 360° to 180° (Class B). From a full conduction angle of 360° (Class A) to a half conduction angle 180° (Class B), the primary harmonic content at the collector of a Class A, B, and AB PAs is the 2nd harmonic. Class C amplifiers have conduction angles below 180°, and the 3rd, 4th, and 5th harmonic amplitudes increase significantly while DC amplitude decreases. By terminating the harmonics properly, the output waveforms can be shaped back into sinusoidal waveforms at the fundamental frequency to increase the overall amplifier efficiency.

If the amplifier is overdriven (clipping) or biased into Class AB (reduced conduction angle operation), significant harmonic content can be generated, and modified design approaches

[10]"Light Class AB" and "deep Class AB" operation are relative, non-exact terms describing biasing points in relation to purely Class A and fully Class B. Deeper Class AB operation refers to Class AB that is pushing closer towards Class B.

[11]See Appendix A for the formulas representing each harmonic.

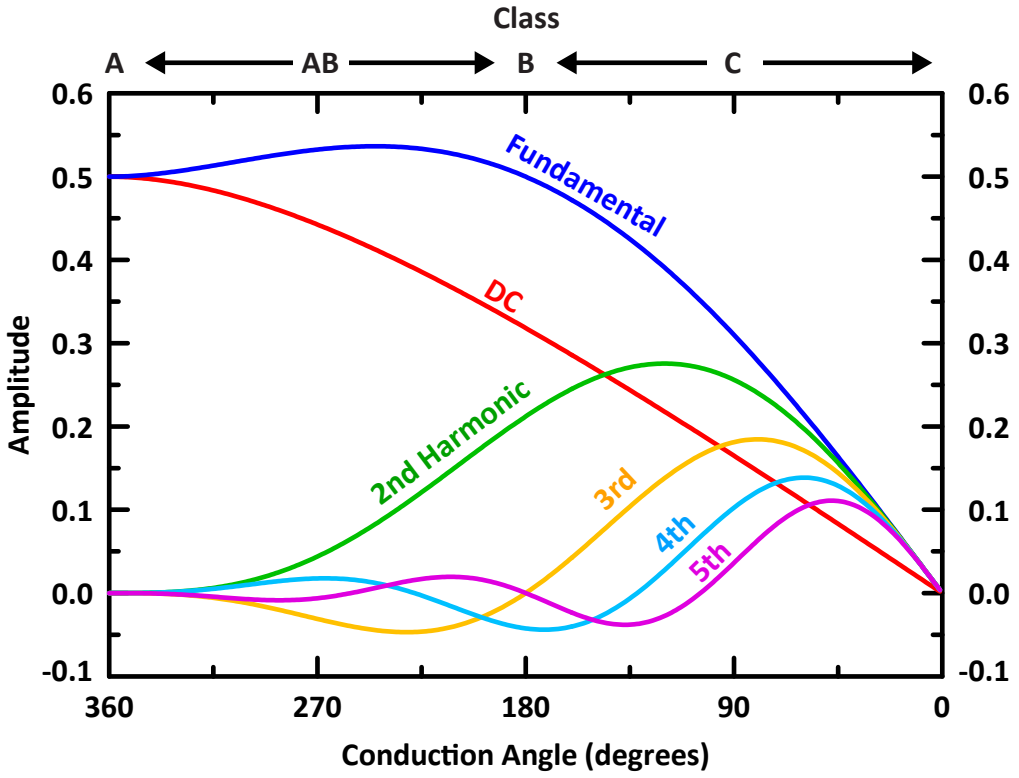


Figure 1.2: Collector current vs. conduction angle ($I_{max} = 1$).

are required compared to purely Class A. Fundamental concepts in linear circuit design and analysis such as equivalent resistance assume sinusoidal, small-signal waveforms. When waveforms begin to clip and when the signal levels change the operating characteristics of the transistor, traditional small-signal approaches and assumptions break down.

The load line of an overdriven Class A PA, where input level are high enough to clip waveforms, differs from a purely Class A PA. Much like reduced conduction-angle Class AB operation, large-signal load-line match to overdriven conditions can result in higher output powers and efficiency [64, 65]

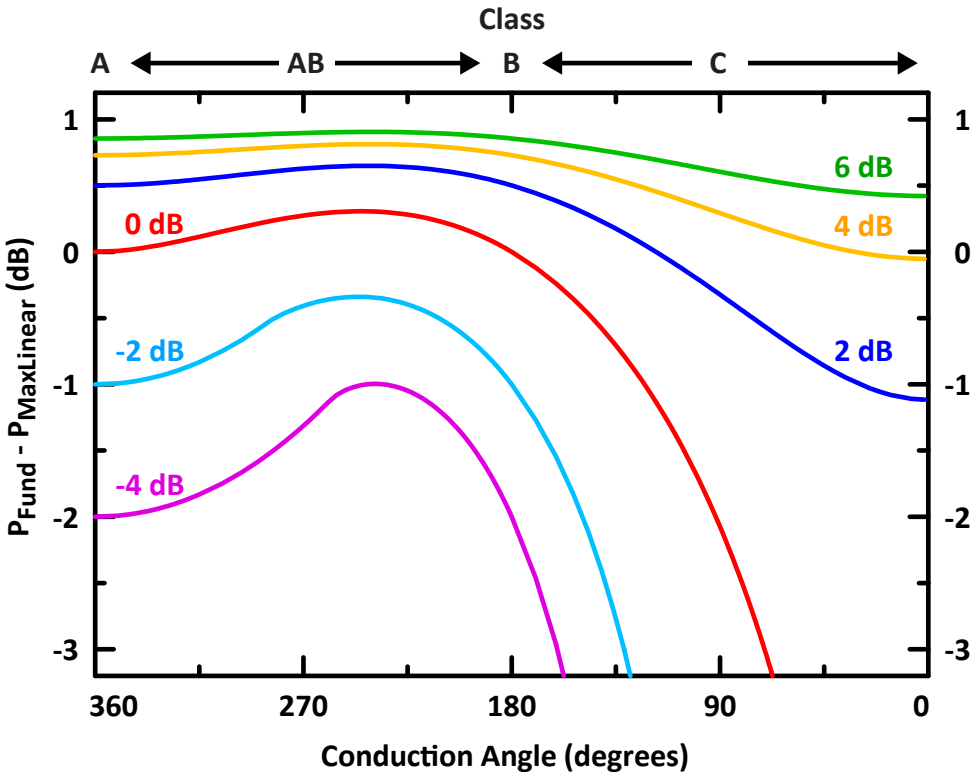


Figure 1.3: Fundamental output power vs conduction angle for input overdrive -2 to 6 dB.

The output power at the fundamental frequency with different input overdrives, relative to the maximum linear output power— no output current waveform clipping—, is plotted in Fig. 1.2.4 [63]. The amount of overdrive is relative to the input drive power for maximum, unsaturated current swing at each conduction angle. From conduction angles ranging approximately 360°–225° in Class A to AB operation, increased input overdrive powers result in increased output powers and efficiencies, as shown in Fig. 1.2.4. Thus, a PA, especially those with multiple gain stages, can be optimized and tuned for additional output power & efficiency by managing overdrive conditions and harmonic content (waveform shaping) appropriately. For Class AB and overdriven Class A/AB operation, harmonic components

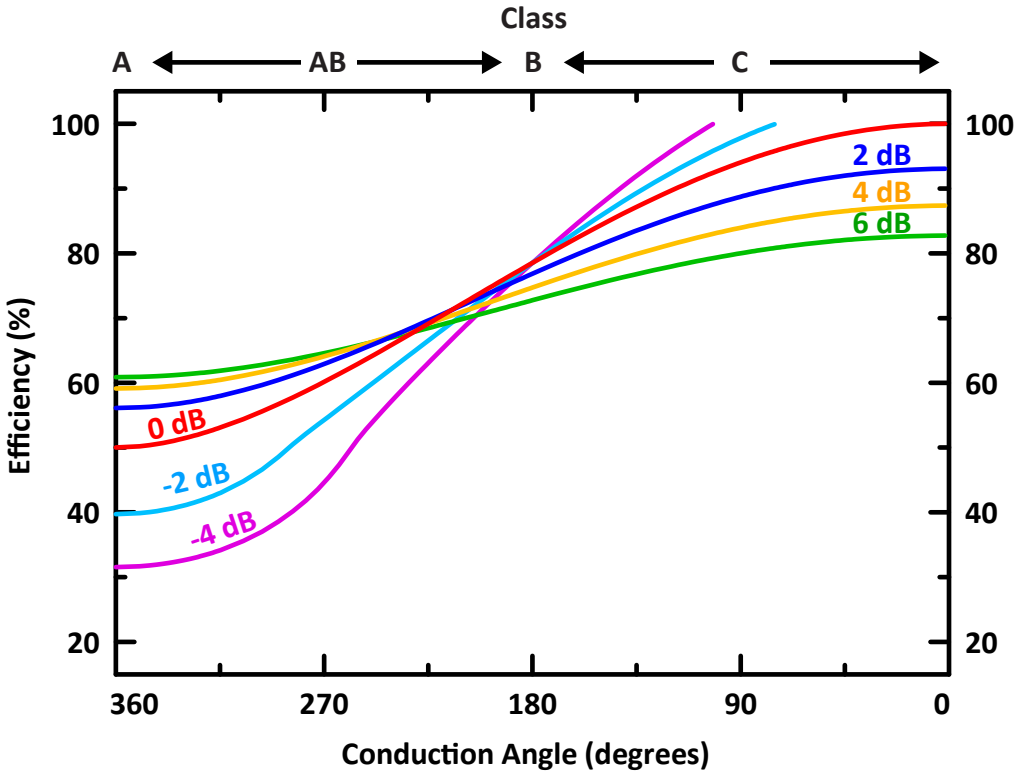


Figure 1.4: Efficiency vs conduction angle for input overdrive -2 to 6 dB.

of collector current waveform should be terminated in shorts to ensure the output current waveform is purely sinusoidal at the fundamental operating frequency .

1.3 GlobalFoundries 8HP Silicon Germanium Technology

The presented design uses the GlobalFoundries (GF) SiGe 8HP BiCMOS technology, formally known as IBM 8HP [66, 67]. GF 8HP is the 4th generation GF SiGe BiCMOS process and readily available as a high-volume commercial foundry offering in a 200 mm (8 inch) fab. Reticle sizes in SiGe 8HP are 18 mm × 20 mm, stepped 60 times per 200 mm wafer. The silicon wafer starts out at 760 µm thick and is polished down to 250 µm from the back side at the end of the fab processing.

The GF 8HP process incorporates a 0.13 µm gate-length CMOS process with a 0.12 µm emitter-width HBT device. The 8HP BiCMOS process follows a base-after-gate (BAG) process with self-aligned emitter where most of the CMOS processing is completed before the formation of the SiGe HBT base. Only the CMOS source/drain implants remain after the bipolar device is completed. The deep-trench and subcollector structures and the HBT base is formed after the CMOS gate.

GF SiGe 8HP has a silicon p^- substrate ($\epsilon_r = 11.9$) with bulk resistivity of 13.5 Ω·cm, which corresponds to a substrate conductivity of 7.41 $\mho$/m. The front-end-of-line (FEOL, in-substrate) process includes oxide isolation structures within the substrate. A shallow-trench isolation (STI) oxide reaches 0.35 µm deep into the silicon substrate. An even higher-isolation structure, the deep-trench (DT) isolation, is formed from etched 6 µm-deep trenches and filled with oxide & polysilicon, which are both electrically and thermally insulating from the substrate.

There are 7 metal interconnect layers in the back-end-of-line (BEOL) process: two aluminum, four 1×-thick copper, one 2×-thick copper. The 4 µm thick aluminum top-level metal (AM layer) sits within silicon dioxide (SiO_2: $\epsilon_r = 4.1$) 13.35 µm above the silicon substrate and allows low-loss microwave transmission lines to be synthesized on-chip. High-density metal-insulator-metal (MIM) capacitors with aluminum plates use a nitride dielectric ($\epsilon_r = 7.0$) to achieve a MIM capacitance density of 1 fF/µm^2. The back-end-of-line also includes precision tantalum nitride (TaN) metal resistors with sheet resistance $R_s = 60.5\ \Omega/\square$ that enables precision resistors to be placed above the SiGe substrate layer.

The PA designs use only the high-performance, high-f_T HBT as the active transistor device— no other active devices are used (such as FETs)—, but additional transistor devices

are available for higher integration of transmitter blocks. 0.13 µm gate-length MOSFET devices (2.2 nm gate oxide) support 1.2 V operation, and triple-well NMOS nFETs are available, as well. 8HP supports 1.2 V and 2.5 V digital 0.13 µm CMOS transistors as well as 2.5 V I/O (input/output) transistors. High breakdown HBT transistors (npn) with f_T = 110 GHz & f_{MAX} = 55 GHz and vertical-pnp bipolar transistors are available, as well.

Table 1.5 summarizes the technology features of the 8HP foundry process[12], and the metal stack-up is presented in Fig. 1.5. Fig. 1.6 charts the f_T and f_{MAX} of a GF 8HP 5 µm × 0.12 µm HBT (npn) device simulated with the original VBIC model.

Key parameters of interest in power amplifier design are the npn HBT transistor's f_T (200 GHz), f_{MAX} (265 GHz), and BV_{CEO} (1.8 V), $BV_{\mathrm{CBO}} = 6.0$ V. These parameters influence the available gain and power for large-signal designs in mmW power amplifier designs and will be discussed in more detail in the coming chapter.

Table 1.5: GlobalFoundries SiGe BiCMOS 8HP Technology Features

High f_T npn: Peak f_T	200 GHz[13]
High f_T npn: Peak f_{MAX}	265 GHz
High f_T npn: BV_{CEO}	1.8 V
High f_T npn: BV_{CBO}	6.0 V
High-Breakdown npn: Peak f_T	57 GHz
High Breakdown npn: BV_{CEO}	3.6 V
nFET Peak f_T @ 1.2 V Supply	f_T = 110 GHz
nFET Peak f_{MAX} @ 1.2 V Supply	f_{MAX} = 55 GHz
MIM Capacitor: 1.0 fF/µm^2	Q = 7–10 @ 90 GHz; Q > 30 @ 30 GHz

[12] At the time of the initial process release under which the presented PAs were designed, f_T = 200 GHz, originally. Process improvements have since increased the 8HP f_T to 210 GHz (a 5% improvement that does not drastically affect existing 8HP designs). 8HP is quoted as both f_T = 200 GHz and 210 GHz, depending on time of publication. The HBT BV_{CEO} is often quoted from 1.7–1.8 V.

[13] All designs have been fabricated under the earlier 8HP process with f_T = 200 GHz.

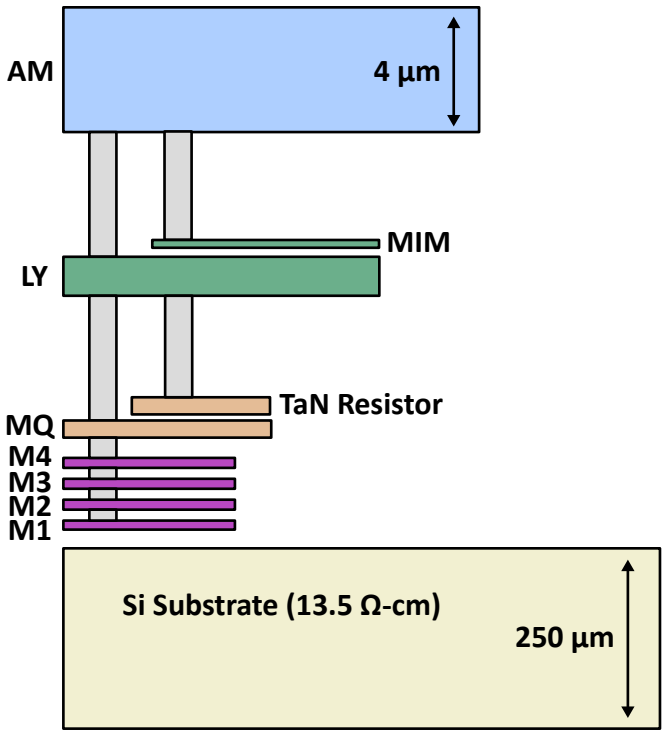


Figure 1.5: GlobalFoundries SiGe BiCMOS 8HP back-end-of-line stack-up. The process offers a 4 µm aluminum top metal, a MIM capacitor, and a high-precision TaN resistor above the active substrate.

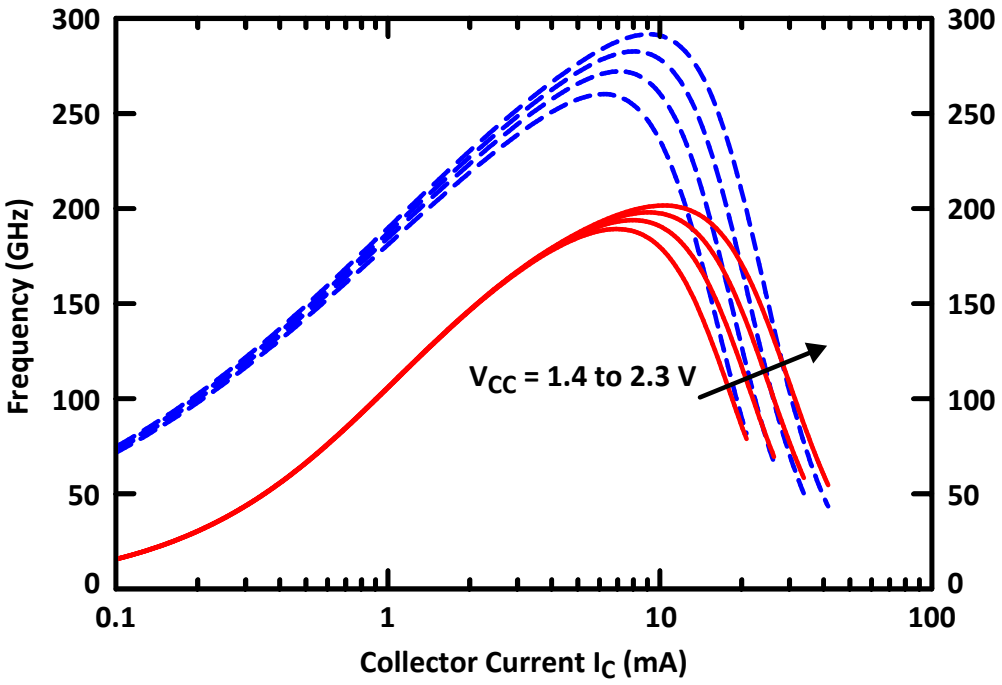


Figure 1.6: f_T & f_{MAX} versus collector current for a 5 µm × 0.12 µm HBT npn transistor within GF SiGe 8HP technology for V_{CC} = 1.4 V to 2.3 V in 0.3 V steps.

Chapter 2

W-Band SiGe HBT Power Amplifiers

2.1 Introduction

The power amplifier is the highest DC power consuming block of a wireless transceiver. Higher efficiency reduces the on-chip wire-size requirements when combining multiple PA elements in a large array fed from common DC-distribution networks and simplifies the overall transmitter system design. More output power from a single, non-power combined PA generally leads to fewer required array elements and smaller arrays. Wider PA bandwidths and higher linearity enable higher maximum data rates in wireless communications. Multi-channel systems require linear amplifiers with gain flatness within the operating bandwidth. Gain flatness simplifies any necessary system linearization. A wide-bandwidth PA with flat group delay response maintains constant group delay & gain throughout the narrower channel bandwidth of the modulated signal.

Early silicon mmW designs at W-band frequencies follow microwave distributed-layout techniques proven in GaAs monolithic microwave integrated circuit (MMIC) designs. In GaAs MMICs, where substrate thermal conductivity only one third that of silicon (0.46 vs. 1.5 W/cm-°C [68]), densely-packed power transistors often have stability issues without proper thermal ballasting through electrical feedback: negative-feedback via emitter and/or base ballasting resistors. Very large transistors, at mmW frequency, also have a distributed effect; so corporate feeds ensure proper phasing, matching, and isolated temperatures between transistor groups within a common amplifier stage, but the cost is an extremely large fan-out area consumed by long transmission lines. Additional combiners, which also consume large areas, are required to form the final combined output, as well. While traditional MMIC approaches to power amplifier design using exclusively high-Q distributed elements and matching networks can be successful in silicon technologies at microwave frequencies, the designs are extremely large in chip area (up to 20–40 mm^2). Their size prohibits the combining of multiple PA elements into very large arrays (e.g., 64-element transmit arrays) at reasonable costs. The tradeoff in SiGe design is to use a combined approach of compact, lower-Q

lumped-elements and distributed approaches in order to drastically shrink the die size and make silicon designs suitable for further circuit integration on the same chip.

The W-band (75–110 GHz) power amplifier (PA) design presented here builds upon the two earlier successful iterations of the Ka-band (30 GHz) power amplifier designs presented later in Chapter 3. The W-band designs improve upon the proven Ka-band design approaches, scale the Ka-band designs to 3× higher frequencies, and address the additional challenges of extremely limited gain and impacts of parasitics at 90 GHz within SiGe IC technology. The primary goal of the W-band work is to push the limits of linear Class A/AB output powers at 90 GHz within 0.12 μm SiGe technology while maintaining acceptable small-signal gain (> 10 dB) as well as power-added efficiency ($> 10\%$) for practical integration into wireless transmitters and integrated phased arrays.

In order to achieve the power and efficiency goals, HBT gain must be preserved at 90 GHz, and operational collector voltages must be pushed beyond $BV_{\mathrm{CEO}} = 1.8$ V.

- Push $BV_{\mathrm{CEO}} = 1.8$ V closer to 4.2 V by using extremely low external DC base resistances (20 Ω) presented to the external HBT base nodes.

- Create custom inverted microstrip lines using parallel-stacked high sheet-resistance, thin lower copper metal layers, to insert DC biasing with low 20 Ω resistances while ensuring that impacts on RF gain are minimal.

- Aggregate 5 μm unit-cell HBT transistors into larger custom power-cell transistors with low inter-cell wiring parasitics by packing them in dense 2D configurations. Emitter ballasting resistors, which reduce precious RF gain at W-band, are not required as long as the aggregated HBT is thermally stable.

- Extract custom thermal models for the densely-packed aggregated HBT power-cells, ensure operational electro-thermal stability with the models, and apply the custom thermal modeling to optimize the DC and RF designs.

- Use resonant structures throughout the PA to create compact low-loss bias-insertion networks (quarter-wavelength bias-insertion networks) and low-impedance references at 90 GHz (self-resonant capacitors).

- Apply a unit-cell approach from device-level to system-block level throughout the PA design so that process variations and modeling deficiencies do not de-tune the design, especially since there is little gain headroom available at 90 GHz.

- Include independent biasing provisions for all gain stages in the PA so that collector voltages can be pushed and the design tuned beyond $V_{CC} = 1.7$ V above which the original HBT transistor models could not converge in large-signal simulations.

The W-band power amplifiers (PAs) are fully integrated into a silicon integrated-circuit (IC) chip, fabricated in a commercially-available, high-volume 200 mm foundry process: GlobalFoundries U.S. (formerly IBM) 0.12 μm silicon germanium (SiGe) BiCMOS 8HP. GlobalFoundries GF 8HP offers 0.13 μm gate-length CMOS FET devices as well as 0.12 μm emitter-width npn transistors with $f_T \approx 200$ GHz and $BV_{\mathrm{CEO}} = 1.8$ V. The W-band power amplifier (PA) designs demonstrate the maximum output power capabilities of linear-mode, Class A/AB, common-emitter power amplifiers in this SiGe technology node without power combining techniques beyond differential-mode (balanced) operation. Wideband small-signal and large-signal performance with smooth responses over frequency (e.g., gain, P_{1dB}) is achieved without using distributed amplifier techniques by applying unit-cell approaches to all matching network elements.

The wideband, linear Class A/AB design achieves a maximum small-signal gain of 14.6 dB with a fractional 3dB-bandwidth of 20% (79–97 GHz). The wideband performance extends to the large signal characteristics: at 90 GHz, the differential/balanced PA achieves an output[1] P_{1dB} of 18.8 dBm (76 mW) and a saturated output power P_{sat} of 19.9 dBm (98 mW) with a peak PAE of 15.4%. The 0.12 μm SiGe chip is 2.4 mm^2 and consumes a maximum DC-quiescent power of 488 mW from a 1.7–2.3 V supply. All matching networks, bias circuits, and wideband capacitor bypass arrays are fully integrated on chip.

The W-band PA design follows the foundry's reliability guidelines for 100,000 hours of operation (11.4 years) at 100°C, which is suitable for commercial electronics applications. To our knowledge, this is one of the highest-power standalone (non-power-combined) Class A/AB (linear-mode) W-band amplifiers to date in 0.12 μm SiGe technology.

2.2 Design Challenges and Approaches

This section reviews key design challenges in the W-band design. More specific technical details are included later in the chapter in dedicated sections to each topic.

The W-band PA builds upon two earlier successful iterations of the Ka-band PA[2] at

[1]Expected differential output powers are reported from single-ended measurements by adding +3 dB because the available source power in the measurement setup is extremely limited. Even without input baluns, which would drop powers around 4 dB (3 dB power splitting + ∼1 dB balun losses) there is not enough input power to fully compress the PA above 93–97 GHz, depending on the PA gain for different biasing conditions.

[2]Each Ka-band iteration and W-band (single iteration) contains a baseline PA design and small variants in multiple design breakouts.

30 GHz while tackling new challenges from 90 GHz operation. The progression of the design approach is discussed earlier in Section 1.1.2.

The W-band design uses the same approach demonstrated at Ka-band, but no circuit structures and layouts can be re-used directly. All circuits and layouts are re-designed around new challenges at the higher frequency and optimized for W-band. Fundamental aspects of the system approach and circuit topologies are carried forward to the W-band work, however.

Integrated power amplifier development at W-band is especially challenging in silicon because the limited transistor gain and low breakdown voltage typically result in low output power and efficiency. The tripling of the center frequency from Ka-band to W-band results in many new design challenges since transistor maximum available gain is 6 dB or less, circuit and layout parasitics at 90 GHz contribute significant reactance and loss, and even short metal interconnects have non-negligible parasitic effects on gain.

Some common PA designs challenges are shared between Ka-band and W-band designs:

- The 8HP technology does not have large transistors (beyond > 18 μm) necessary for high-output power designs. Even the 18 μm transistor is a single emitter-striped device, and there are no standard-cell, aggregated large transistor devices or associated models for dedicated power-HBT devices.

- Limited collector breakdown voltages is a shared, frequency-independent challenge between the Ka-band and W-band designs. The collector breakdown voltages are limited because high-f_T devices trade-off speed for lowered breakdown voltages from increased impact ionization effects, which limits output voltage swings and output power.

- Absolute values of components can vary as much as ±25% or more, but relative values of identical components— matching— is generally better than $< 0.1\%$ for passive components, but still ±25 for certain DC parameters within HBT active devices. The unit-cell approach ensures that the PA does not de-tune from small component variations— that critical components track together with process variation and minor modeling deficiencies. Active bias circuits and independent current and voltage tuning provisions for each gain stage mitigate the differences in the HBTs, built from unit-cell blocks, to achieve better RF-performance matching to expected design values.

There are many new PA design challenges at W-band within 8HP HBT technology, not present at Ka-band, because of the 3× increase in frequency:

- Transistor have lower available gain (< 6 dB) at W-band, even when biased at peak-f_T/f_{MAX} current densities, leaving very little gain headroom large-signal designs and

making efficiency (PAE) especially difficult. In terms of frequency, W-band designs in 8HP push the limits of the HBT capabilities, especially for power amplifiers.

- Large transistors have low input and output impedances and can be difficult to match given that small layout parasitics have impact on losses and matching at 90 GHz.
- Parasitics, especially those around HBT wiring, have much greater impact at 90 GHz. Scaling the frequency does not scale down other aspects of component connections and interconnects, making associated parasitics more problematic.
- RF short-circuits from capacitive bypass (large capacitor) are non-ideal and have residual resistances at 90 GHz. Large capacitances ($>$ 500 fF) have self-resonant frequencies below W-band.

The PA is designed to demonstrate the fundamental frequency & power limits of a standalone, non-power-combined common-emitter PA in commercially-available 0.12 μm SiGe HBTs. This work employs a differential (balanced) amplifier topology for a +3 dB increase in output power, but does not utilize any microwave power-combining techniques. The goal is to maximize output power while maintaining reasonable gain ($>$ 10 dB) and reasonable PAE ($>$ 10%). If more system power is required, this PA can be used as a unit building block in an array with power combining, as shown in a design extension in Chapter 4. The need for reasonable gains and PAEs in PAs becomes even more evident when a PA is used as an element in a large integrated array. High output-powered PAs with gains below 10 dB become difficult to drive with the proper input power, and power efficiency limits how many PA elements can be arrayed on a single chip before power-bus feeds reach their electromigration limits from multi-amp-level currents and watt-level power dissipation.

A major design goal is to push output power and voltages of the common-emitter HBT towards frequency and power limits while while maintaining reasonable performance in gain and efficiency such that the PA can be integrated further into multi-element transmitter arrays. The PA design follows strict foundry reliability guidelines to achieve expected chip lifetimes for reliable commercial applications: 100,000 power-on-hour (POH) at 100°C. Operational frequency is limited by low gains at 90 GHz (low f_T & MAG) as well as parasitic losses and effects of wiring and connections at 90 GHz. Power limits are primarily constrained by device breakdown voltages, design.

The design approach uses existing foundry models for the active transistor devices and custom EM simulations for distributed structures and parasitics. No custom active-device modeling or simulations via technology computer-aided design (TCAD) is involved. A unit-cell approach is used throughout the design enabling standard HBT models and passives to

be adjusted using DC & RF measurements from prior Ka-band PA designs— back-modeled with "in-situ" PA-level measurements and not individual component characterizations. The design approach uses robust design techniques to overcome model limitations to build a successful power amplifier at millimeter-wave frequencies with first-pass silicon success. The design approach is highly scalable to even higher frequencies in more advanced, future SiGe technology nodes.

2.2.1 Design Approach

The PA design approach utilizes the following concepts: build all driver & power transistors (Q1, Q2, Q3) from one $Q_0 = 5 \times 0.12\ \mu m^2$ unit-cell transistor; re-use similar passive structures that offer wide frequency response; and allow for independent amplifier stage bias controls of both quiescent (DC) voltage & current for precise tuning of the silicon parts in the lab since the transistor models are known to have limitations at W-band and at high collector currents & voltages. The tuning work on the lab bench explores the practical limitations of operational collector voltage, in particular, for common-emitter configurations. The W-band PA is designed with additional, practical objectives in mind: reliability, further integration (reasonable performance metrics for inclusion in real system blocks), bias tuning and optimizations when pushing collector currents on the measurement bench, highly linear design, and wideband response (low loaded-Q matching networks).

To attain wideband performance without active feedback or distributed amplification, the design takes advantage of the high-f_T npn HBT and a judicious use of distributed and lumped elements in the RF matching networks and bias insertion networks. Simple, two-element, high-bandwidth ($Q_{loaded} \sim 2$) high-pass matching networks, wideband resonant interstage matching structures, and wideband quarter-wavelength collector & base bias insertion networks are used to achieve both small-signal and large-signal performance across a wide bandwidth.

The general design philosophy adopted is one that favors design predictability over incremental performance gain, especially for the passive components. In the unit-cell approach, a higher tolerance passive structure is more attractive than a higher-Q component with much higher process variation or worse adjacent-component matching. The active transistor HBT models are known to have modeling errors above 60 GHz and near voltage & current limits in large-signal operation. Some tuning of the HBT transistor response is possible through independent quiescent DC voltage and current biasing in each gain stage, but the passive devices have no such post-design tuning provisions on the measurement bench.

Wherever possible, design components and methodologies are reused: transmission lines,

capacitors, matching network topologies, and power cells built with the baseline unit cells and used in similar circuit topologies. Unit building cells and similar circuit topologies are used so that any modeling limitation or fab-processing skews affect portions of the design nearly identically. This approach makes the PA design robust and less sensitive to any single point of modeling or component tolerance failure, and the PA frequency response will always be nearly the same outside of small frequency shifts. The resulting design also becomes easily scalable in frequency and simple to re-center to a different frequency of interest with only a few changes on the metal levels.

To maximize power output, the npn transistor and supporting passive matching networks are optimized for PA operation near P_{1dB} and P_{sat} rather than purely in small signal. A purely small-signal Class A design used far from its gain compression point P_{1dB} can, in theory, rely simply on classic scattering-parameter (S-parameter) design approaches, but to push a Class A PA to its power limits, large-signal load-pull simulation techniques must be used since S-parameters will vary with input/output power levels. Class AB and overdriven Class A (driven into full gain compression past P_{1dB} require further large-signal analysis of harmonic terminations because of the large amount of clipping that occurs in the waveform. Class AB PAs also require study of any gain expansion that might arise in any of its stages—another form of small-signal gain variation with input power.

Architecture

The power-amplifier development work presented here focuses on linear continuous-mode Class A and Class AB modes of operation, including overdriven Class A/AB— Class A & AB driven into gain compression for higher output powers. For the case of purely Class A biasing and operation, the output current and voltage waveforms of the PA are, ideally, linearly amplified version of their inputs with no additional harmonic distortion of the waveforms. This Class A high linearity comes at a cost of power efficiency since the PA is consuming a high amount of DC current regardless of the AC input waveform. The theoretical maximum collector efficiency of a Class A PA is 50%. The current biasing of Class A amplifiers can be lowered into Class AB to reduce the quiescent DC power consumption and thus raise the power efficiency, but the troughs of a sinusoidal output current waveform clip at higher input powers, effectively reducing the conduction angle of the PA and introducing harmonic distortion. The Class AB PA thus achieves higher efficiency at the expense of signal distortion, which can be mitigated by terminating the collector current harmonics properly. Mild Class AB operation involves harmonic management of the 2nd harmonic.

The PA design includes three gain stages with separate bias-tuning provisions for each stage. This architecture is optimal use of the available silicon area. The bias tuning provisions

are included so that the PA can be moved into Class A, Class AB, overdriven Class A, or overdriven Class AB depending on the quiescent DC currents and voltages set in each stage of the PA. By trading off pure Class A linearity, both higher output powers and higher efficiencies can be achieved through the harmonics generated by clipped waveforms and managed by output harmonic tuning networks.

For the large-signal Class AB design, to obtain higher output powers and efficiencies, the W-band PA RF matching networks are optimized for large-signal operation with HBTs into compression near the $P_{1\text{dB}}$ point.

Unit Cell Approach

For millimeter-wave RF integrated circuit (RFIC) designs, a significant challenge is managing and reducing the design variability arising from modeling inaccuracies at high frequencies and high currents & voltages, especially in a new process technology. To tackle this challenge, fundamental PA building blocks are synthesized and characterized via simulation, and these transmission lines, capacitors, inductors, bias feeds, and transistors are re-used throughout the PA design. The unit-cell approach re-uses similar circuit structures and blocks built from unit cells to make designs more robust to modeling errors and process shifts. There is inherent design compromise with unit cells— not necessarily optimal in all use cases—, but RF performance is maintained despite any expected process biases or modeling errors.

The unit-cell approach allows excellent matching between identical or very similar elements. The approach is applied to both active (HBT) and passive structures (MIM capacitors, resistors, transmission-line structures, etc.) throughout the PA. Though a MIM capacitor value may shift $\pm 10\%$ with process from its nominal value, the *relative* matching[3] between identically-dimensioned capacitors on the chip is better than 0.1%.

A unit-cell approach is applied to every active & passive component and circuit block in the chip: resonant MIM RF-shorting resonant capacitors, base quarter-wave biasing lines, collector biasing lines, transmission-line elements, shunt-stub inductors, wideband MIM capacitor bypass networks, HBT unit cells, active-bias-network block unit cells. Most of the reactive design tuning is done through length adjustments of shunt-stub transmission-line unit cells. The synthesized inductance values differ, but they are based on the same transmission-line structure throughout the PA. Small exceptions and changes (e.g., line width or capacitor value) in components are necessary in parts of the design to meet reliability conditions in the output gain stage: the collector feed transmission-line and the interstage MIM capacitor. These exceptions are still based closely on the original unit-cell devices, so

[3]This means that a 100 fF MIM capacitor may vary from 90–100 fF from wafer-to-wafer, but all identically-dimensioned capacitors on the same chip will have capacitor values within $< 0.1\%$ of each other.

they will skew with process together closely.

Back-end-of-line (BEOL) components within the metal stackup are used where possible over similar devices within the substrate. Structures defined by metal etching have higher tolerances compared to devices defined by diffusion processes within the silicon substrate (e.g., substrate resistors). Moreover, structures in the aluminum-metal layers have higher tolerances than those made from structures in the lower copper layer because of the copper damascene planarization and etching processes.

The HBT active devices only have 25% relative DC beta matching from adjacent device to device, which is considerable worse than that of passive devices on the chip. The DC currents within the HBTs need to match target current densities in order to match expected RF performance. The solution is to use separate active-bias reference circuits in each gain stage (rather than purely resistive biasing) to reduce relative matching effects and to have independent biasing provisions to tune the HBT bias points of each gain stage. Once DC currents and voltages are adjusted properly, the RF performance will be recovered despite any process shifts or relative device mismatching.

Choosing unit cells is an inherent compromise– not the absolute optimum choice, but a tradeoff in the design. The benefits are higher predictability in the face of expected process variations and known modeling deficiencies or inaccuracies. This unit-cell approach is not the absolute optimum choice for every situation, but it provides several advantages: the unit cell can easily be reused throughout the design, the approach is robust to modeling and process skew, the design can easily be scaled, and the approach allows for easier back-modeling of measured data (by constraining the degrees of freedom or number of parameters in the optimization process) in order to re-tune designs.

The unit-cell approach makes the design more robust to process variations and modeling errors. As a bonus, re-use of unit-cell components and circuit blocks throughout the design allow for easier back-modeling of measured data. Back-modeling DC & RF measured data into simulations is greatly simplified because of the well-constrained parameter space— fewer degrees of freedom of possible parameter skews. The back-modeling process incorporates the process biases and modeling errors back into model adjustment for more accurate hardware to simulation correlations and for more accurate designs using the adjusted models going forward.

HBT

The npn heterojunction bipolar transistor (HBT) is a key building block and only active device within in the PA designs. HBTs offer the high-frequency signal-gain performance necessary for power gain at millimeter-wave (mmW) frequencies.

Key challenges in designing SiGe mm-Wave PAs involve transistor-model limitations. The lack of dedicated large power cells (large devices designed for power with > 18 μm emitter lengths) necessitates building custom power cells that are not explicitly modeled within the design kit HBT models. The W-band PA must be designed from aggregating design-kit standard cells with modeling errors & deficiencies in mind. Furthermore, available models have limited accuracy and convergence problems at higher collector voltages (> 1.7 V) under gain compression.

The W-band design is the first to use densely-packed cells in an aggregated power transistor, and no prior hardware existed to model the aggregation effects precisely. For custom aggregated arrays of transistors to form larger power cells, mutual thermal coupling is not modeled in the foundry kit, and secondary substrate effects from the relocation of the surrounding substrate ring around the aggregated HBT are not modeled, either. Extensive Ka-band DC measurements on similarly-packed 5 μm unit cells provide the necessary mutual thermal coupling data to center the DC & RF performance of the custom W-band power transistor cells. Since PA biasing is particularly sensitive under high-current and high-voltage power conditions, bias provisions are built-in to allow optimization of PA biasing on the lab bench to explore the true output capabilities of the PA.

In order to achieve the push the SiGe HBT to its limits at W-band, the 8HP HBTs, in common-emitter configuration, have DC collector voltages pushed to 2.3 V. To push output powers above the common-emitter (with open base) $BV_{\mathrm{CEO}} = 1.8$ V limitation closer toward the common-base (with open emitter) $BV_{\mathrm{CBO}} = 6.0$ V, novel DC base-biasing structures are designed to place transmission-line feeds beneath the global RF ground plane, sharing the ground plane with the top-level AM-metal transmission-line structures. The distributed transmission lines in the base biasing feeds present extremely low 20 Ω external DC base resistances to the HBT, which increases the operational collector breakdown voltage well beyond the open-base condition $BV_{\mathrm{CEO}} = 1.8$ V while maintaining high shunt RF impedances with the quarter-wavelength distributed transmission line structure terminated in an RF short circuit and used as a resonator at W-band.

Mutual thermal coupling within the power-transistor cells are estimated from back-modeling of Ka-band DC measurements in order to make proper design adjustments and optimizations

Independent Bias Provisions

Bias-adjustment provisions for both collector currents and voltage in each HBT gain stage are necessary to push power output at higher collector voltages. Independent active-bias reference circuits provide the fine current-tuning provisions while collector voltages have

separate DC bias pads for external DC biasing needles. Fine DC-biasing tuning provisions are especially important at higher collector voltages, $V_{CC} > 1.7$ V, where collector biasing is extremely sensitive. The collector biases must be set accurately and consistently with collector-voltage increases since obtaining the correct DC quiescent points is critically important to obtain the desired RF performance.

The independent biasing provisions also allow pushing the design or selective gain stages deeper into Class AB operation or moving toward a more linear Class A operation— trade off some small-signal gain & linearity for higher output powers and higher efficiencies. Independent biasing also makes the PA more stable at very-low frequencies (MHz range) where the SiGe chip must depend on additional off-chip capacitive bypass components to filter the measurement setup's adjustable precision power supplies. The tradeoff of flexible, independent biasing in each gain stage is increased PA area consumption, especially with quarter-wavelength-resonator bias-insertion networks and their associated wideband MIM capacitor arrays.

Bias tuning provisions are especially useful since the foundry transistor models do not accuracy predict the DC currents of the custom-layout power transistors (due to unmodeled electro-thermal coupling), and the high collector currents & voltages used in this PA design fall on portions of the DC-IV curve where there are significant npn transistor modeling difficulties. These high collector currents & voltages are pursued to push the SiGe HBT transistor to its power limits despite the known modeling limitations.

Individual, external analog biasing provisions allow each of the 3 PA-gain stages' operating points to be adjusted independently and fine-tuned on the bench. Separate, integrated active biasing circuits control each stage's collector current I_C via external DC-biasing pads. Additional pads feed the collector voltage V_{CC} of each amplifier stage. These independent DC tuning controls result in 6 DC pads for each half-PA (single-ended), not including additional ground pads. The biasing flexibility allows the PA design to be tuned and optimized since quiescent bias points are critical in PA operation.

Passive Structures

The design uses both high-Q distributed-element transmission-lines and compact lumped-element MIM capacitors. Distributed inductive elements have much larger area footprints, but less loss ($Q > 25$) than the lumped elements like MIM (Q = 7–10) at 90 GHz. However, lumped MIM elements provide extremely compact series reactances while being DC-blocked, making interconnects between gain stages short, which reduces connection parasitics. In comparison, traditional MMIC design, using only distributed elements, are very large and thus difficult to integrated into larger integrated, on-chip arrays.

There are 16 resonant RF structures used throughout the differential PA (8 in the half-PA). Self-resonant MIM capacitors, resonant quarter-wavelength feeds, and series-resonant transmission-line & capacitor matching network elements are found in the RF design. Resonant structures are generally sensitive to small parameter shifts, and ±20% shifts in different-valued inductors or capacitors can de-tune an amplifier design completely. The unit-cell approach alleviates severe design de-tuning by ensuring that all unit-cell elements track closely together. The PA design does not severely de-tune from expected process shifts or small modeling errors.

True AC-grounding that is decoupled from DC-ground (i.e., very large capacitances) is difficult to achieve at 90 GHz because of associated losses and parasitics of capacitors. A resonant MIM capacitor structure is chosen to provide a known low-impedance reference at 90 GHz— an RF "shorting" capacitor that blocks DC, but shorts RF. This special MIM capacitor is used as the RF-short termination of key quarter-wavelength (QW) bias lines. to create high-impedance bias-insertion resonators. The RF short is transformed to high impedance insertion points by the on-chip quarter-wave transmission lines. This same "RF-shorting" MIM capacitor is used to create a resonant LC structure in the interstage matching networks.

To maintain a compact differential PA layout and to provide the necessary low DC external base-biasing resistance of 20 Ω while maintaining high RF input impedances in a bias-insertion network, a custom inverted microstrip line is created from a stack of lower-level copper metals (M2, M3, and M4) underneath the global RF ground shield (MQ, copper) shared with the top-level AM-metal transmission lines.

Fig. 2.1 highlights all the resonant structures in the half-PA.

2.2.2 Simulation Approach

One of the major design challenges in creating and optimizing a state-of-the-art fully integrated W-band PA centers around practical simulation capabilities & model limitations.

The HBT models are based on foundry models using the VBIC (vertical bipolar inter-company) transistor modeling. VBIC is an advanced Gummel-Poon model that takes weak avalanche multiplication from impact ionization into account, which is important when operating the collector voltage beyond BV_{CEO}. The newer, improved HiCUM models were not available at the time, but they are used in some simulations presented here because of their simulation convergence and improved accuracy.

All passive distributed transmission-line structures and junctions including bends, transitions, and other discontinuities are fully modeled within the EM simulator to the PA's

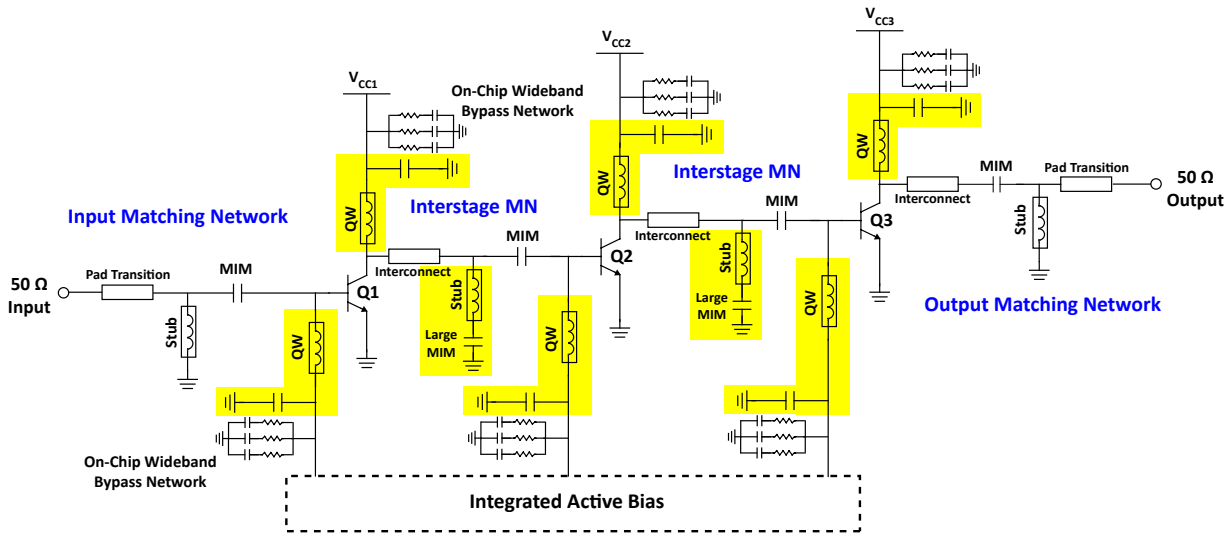


Figure 2.1: Half-PA simplified circuit schematic with all resonant structures highlighted in yellow.

3rd harmonic (330 GHz) for model continuity reasons in large-signal harmonic-balance simulations. Lumped passive devices such as MIM capacitors and TaN metal resistors use the supplied design-kit compact SPICE models. Capacitor leads are modeled in EM simulations. Parasitics associated with the HBT wiring are fully modeled with EM simulations, as well (S-parameter models).

Simulation Limits

Prior measured data on from the Ka-band designs show that > 1.7 V collector-voltage operation is possible, but large-signal simulations are extremely problematic above 1.5 V because of simulation convergence issues with the active-transistor HBT models, and very limited simulation convergence occurs > 1.5 V.

The initial Ka-band design iteration #1 was designed and optimized around collector voltages of 1.4 V because large-signal harmonic balance simulations could converge reliably across all powers and frequencies. Convergence becomes a problem at 1.5 V collector voltages and problems continue to rise until 1.7 V, the limit of the for PA simulations under the original VBIC transistor model. The W-band PA design attempts to center the design at 1.7 V with reduced large-signal simulation accuracy (convergence issues) and provide enough bias tuning provisions and range to tune the DC & RF operating points on the measurement bench as collector voltages are raised toward 2.3 V.

32-Bit Computing and Simulation Complexity

Simulations are done using only 32-bit computing capabilities, which has significant limitations in the addressable random-access memory (RAM). RAM is limited to 4 GB total in a 32-bit system, and only 3 GB of that RAM is available for EM simulations since 1 GB is reserved for the operating system. The computational limitations require simplification of complex layout structures to yield < 3 GB simulation sizes and reasonable compute times. The results are known to have reduced accuracy compared to full-wave 3D FEM analysis on newer 64-bit systems with hundreds of GB of RAM. These EM simulation errors can lead to an estimated 5–10% shift in transmission-line characteristics, which is still within the SiGe process variation. P_{sat} and large-signal convergence issues are not being limited by quantity of RAM, but by the circuit complexity and the limitations of the VBIC model at high collector currents and high collector voltages.

The Agilent (now Keysight) Advanced Design System (ADS) CAD/design environment and its harmonic balance engine, which was only available outside of the Cadence design environment at the time, was the most useful and powerful microwave design tool for mmW power amplifiers. The harmonic balance simulations are more robust to the HBT model convergence issues since accuracy can easily be scarified by reducing the order of harmonic used. In combination with other relaxed DC and RF convergence parameters, the reduced harmonic balanced simulations can obtain some limited simulation data at high output powers which are not extremely accurate, but enough to find some suitable matching points using the original VBIC transistor models.

To use the microwave design tools and harmonic balance engines within Keysight ADS, custom ADS models were manually ported to use the a standalone ADS environment— not tied to the Cadence Virtuoso IC design environment. Since standalone Agilent (now Keysight) ADS kits were not developed yet during the time of the W-band design, all required Cadence Spectre models (HBT, MIM capacitor, TaN resistor, substrate resistor, and other lumped passive elements) are ported manually into custom ADS components using Matlab and custom scripting wrappers. The custom model-kit library is used to simulate the PA in the native ADS design environment. At the time, harmonic balance simulation was not available in the Cadence framework, and the ADS RFDE (RF Dynamic Link/Engine) was a new product that still had many difficulties with ADS-based engines within the Cadence design environment.

HBT Simulations

The PA is designed using the early 8HP design kits with foundry-supplied VBIC-based npn HBT models. ADS harmonic balance is the choice for the large-signal simulations rather than time-domain simulations. Harmonic balance was not available in the Cadence Spectre simulations environment at the time, so the HBT VBIC models are manually ported into Keysight ADS via Matlab and manual scripting.

Harmonic balance solves the steady state response to a periodic input excitation while time domain analysis finds the transient response to an arbitrary input excitation. Harmonic balance is preferred for microwave design since microwave designers are more interested in steady-state response than the transient response. Further, most EM models and passives are described in frequency domain, e.g., S-parameters, and SPICE is not well suited for distributed circuits. In addition, non-ideal transmission lines and microstrip discontinuities are easier to describe in the frequency domain. However, many non-linear HBT transistor models are tailored for SPICE and not harmonic balance, making harmonic balance more difficult. Harmonic balance simulators often rely on nonlinear expressions being continuous through their second derivatives. Discontinuities in the nonlinear expressions can result in convergence difficulties [62].

In common-emitter configuration, below a collector voltage of 1.5 V, simulation convergence is achieved most of the time. With collector voltages above $V_{CE} > 1.5$ V, VBIC models have simulation convergence, especially at large-signal levels when the transistor amplifier is in compression. At 1.7 V collector voltage, there are very limited simulation points, even with reduced accuracy and relaxed convergence settings in the simulator. At the high collector voltages from 1.9–2.3 V, the PA simulation with VBIC transistors provides no convergence in large-signal conditions within harmonic balance.

Reducing the number of nodes in the simulations, lumping parasitics into more simplified models (e.g., converting eight parallel HBT ground inductances of 8 parallel HBT cells into a single, large HBT cell with a single combined inductance) incurs some minor modeling errors, but helps greatly with large-signal simulation convergence, even at modest collector voltages from 1.5–1.7 V.

Passive-Structure Simulations

The majority of the mmW PA chip area is consumed by passive structures— pads, grounding, transmission lines, capacitors, and distributed inductors— rather than active transistor area ($< 1\%$ of total area). Accurate modeling of these passive structures, created in back-end metals above the substrate, is critical. A 2.5-D planar method-of-moments simulator,

Keysight Momentum (formerly Agilent Momentum) is used as the primary electromagnetic (EM) simulator. Although more accurate 3D finite-element method (FEM) simulators are available, the speed-accuracy tradeoff of a 2.5-D planar method-of-moments simulator is preferred since the passive-structure design is a highly iterative process, and planar simulators offer great advantages in simulation time. Material properties and layer thickness in GF 8HP have fairly wide tolerances up to ±20% in key inputs like layer heights, sheet conductivities, and metal line widths.

Custom EM models are created that cover higher-order harmonics ($2f_0$, $3f_0$) up to at least 330 GHz to capture effects of the Class AB PA harmonic terminations. To simplify EM simulations, high modeling accuracy is used up to the 2nd harmonic, 220 GHz, then reduced accuracy speeds up simulations significantly from 220–330 GHz. $2f_0$ harmonic modeling is the most important harmonic above W-band for this Class AB design. Although the S-parameters do not need to be accurate for these higher-order harmonics, they need to well behaved over frequency, e.g., without discontinuities, for proper simulation convergence [62, 69].

2.3 Schematic and Layout

The implemented W-band PA schematic and layout overview are presented here first, and more detailed discussions of design considerations are presented later in the chapter separately for each design aspect. For simplicity, the differential PA circuit schematics are presented for a half-PA only. The two halves of the PA are symmetric, identical, well-isolated PAs, designed to use differential inputs and drive truly differential 100 Ω outputs.

2.3.1 Schematics

Fig. 2.2 is a high-level half-PA block schematic of the implemented 3-stage Class A common-emitter PA. (The other half of the differential PA is identical and symmetric.) The W-band PA consists of three gain stages made from aggregated HBT unit-cells to form custom power cells: Q1, Q2, and Q3. Each gain stage has associated fully-integrated RF matching networks to provide the necessary impedance transformation for target load and source impedances. The RF input impedance to the half-PA is 50 Ω at W-band (or 100 Ω for the differential PA), and the output RF impedance is 50 Ω, as well.

Each gain stage has its own independent collector-biasing and base-biasing provisions, allowing the voltages and currents in each stage to be set and tuned independently from other gain stages. The fully-integrated active-bias network sets the DC quiescent collector currents for each gain stage independently. Voltage is applied to the integrated collector-bias

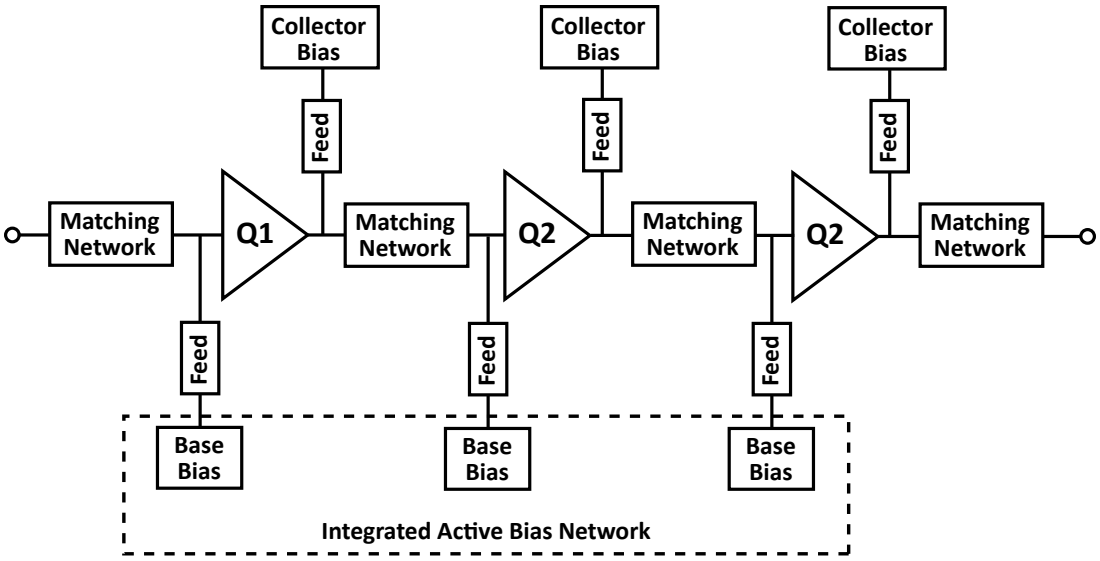


Figure 2.2: W-Band half-PA block diagram.

feeds to set the operational collector voltage of each gain stage (1.7–2.3 V).

Fig. 2.3 presents a simplified schematic of one half of the differential/balance PA. The other half is identical, but is driven 180° out-of-phase.

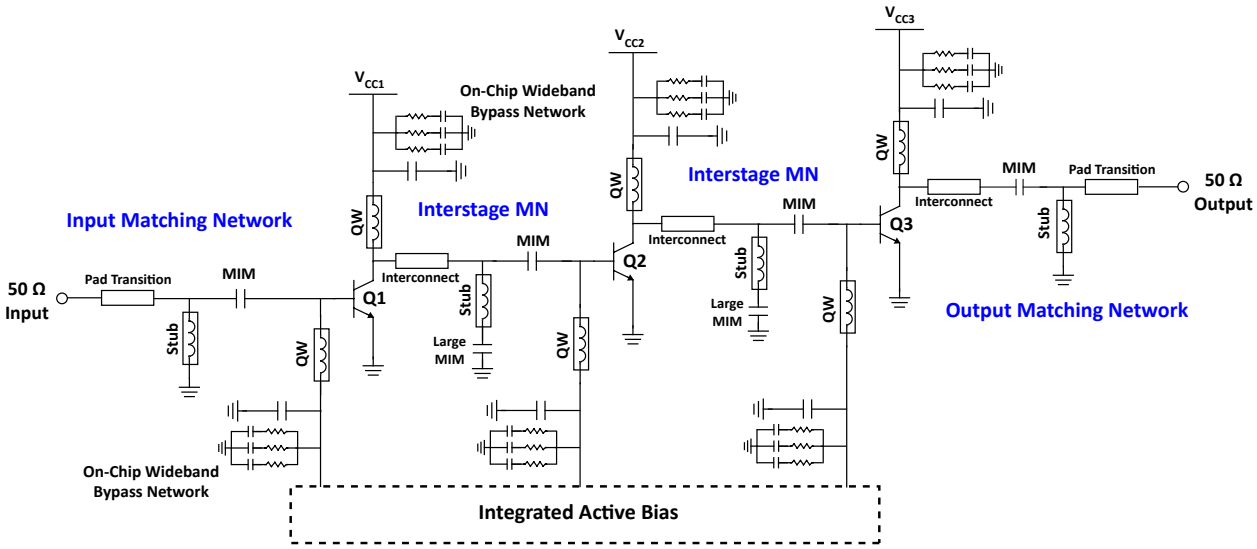


Figure 2.3: Half-PA simplified circuit schematic. Elements marked QW are quarter-wavelength transmission lines at 90 GHz. Elements marked stub are short transmission-line stubs that appear inductive at 90 GHz.

Figure 2.57 shows a corresponding chip micrograph. The chip is X × Y = 1.90 mm × 1.25 mm = 2.36 mm^2, which includes on-chip bypass capacitors, pad areas, and empty space available

for interfacing structures or circuits at the input and output. The PA is symmetric about the center symmetry plane.

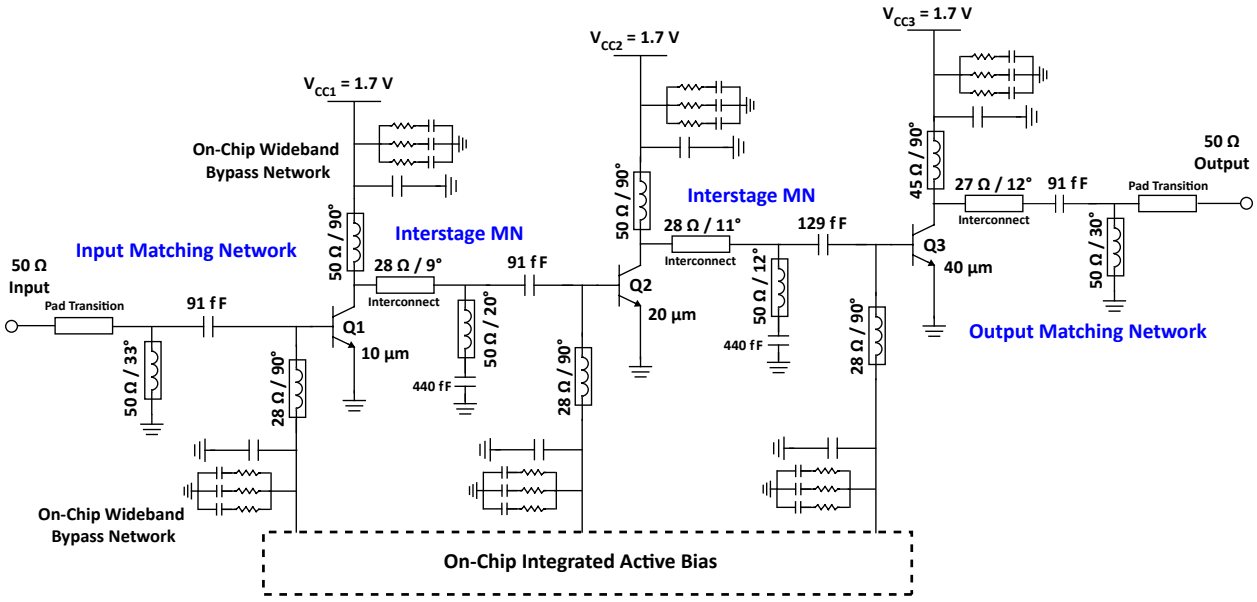


Figure 2.4: Annotated half-PA schematic. The differential PA schematic contains two identical halves driven differentially with differential outputs (0°/180°).

2.3.2 Chip Layout Floorplan

Careful layout considerations are made in order to make the PA a symmetric, differential amplifier with independent current & voltage biasing provisions for each of the three gain stages. Fig. 2.5 annotates the W-band PA layout floorplan. Multiple large capacitors arrays and active-bias circuits are fully integrated into the SiGe chip.

The 12 independent DC bias pads, not including DC ground pads, (with east-west orientation) provide fine-tuning controls over all three gain stages of the differential PA. Each half-PA has 3 DC pads for collector voltages and 3 additional DC pads for reference voltages to bias on-chip active bias networks that set the DC collector currents of the power transistors. The differential PA is perfectly symmetric across the illustrated north-south line down the center, and each half may be operated separately as a single-ended PA for testing & characterization ($P_{\text{out,DIFF}} = P_{\text{out,SE}} + 3$ dB).

Design tradeoffs in the PA half-circuit physical IC layouts are made in order to place all biasing pads near the perimeter of the die of the balanced PA. The full chip, including pads and integrated capacitor arrays, consumes an area of 2.4 mm^2. The layout is designed for on-chip RF micro-probing (north-south) and biasing through DC wedge probes (east-west).

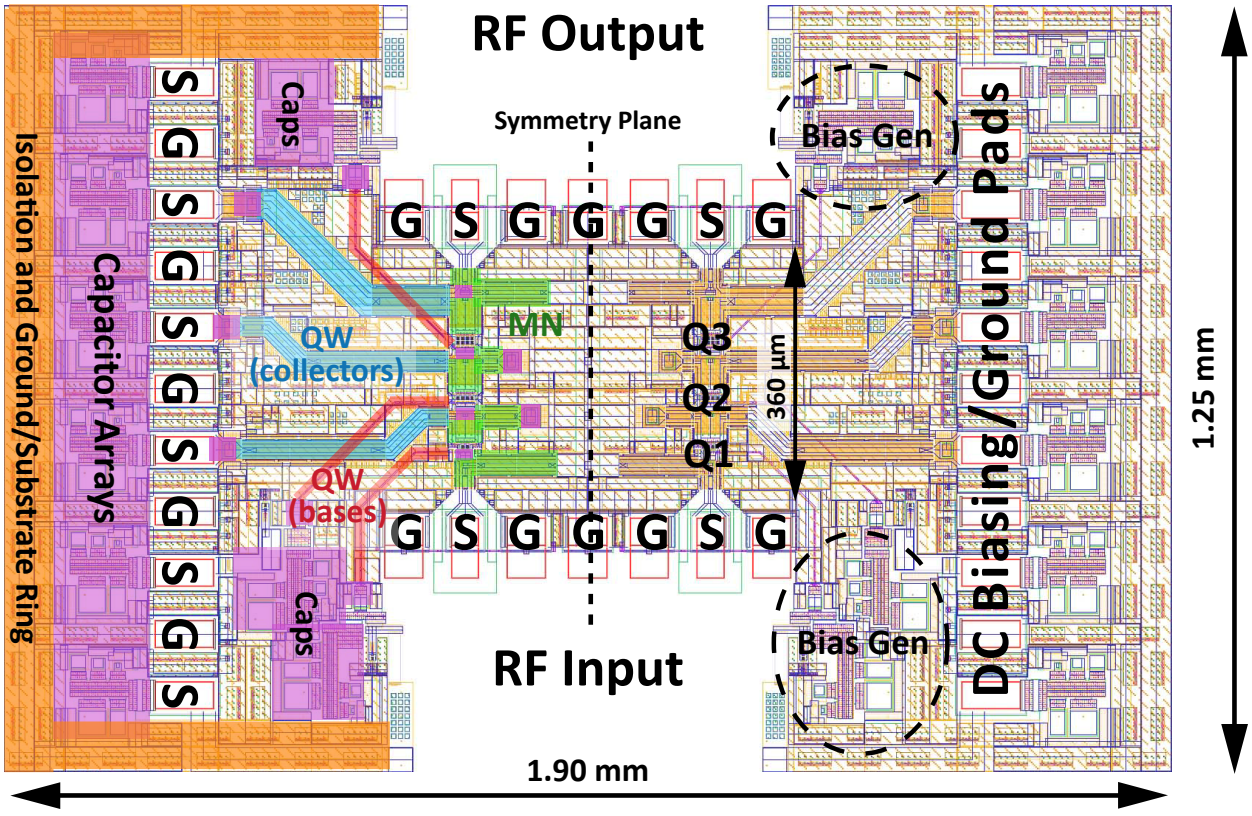


Figure 2.5: Differential power amplifier layout. Each half (left/right) is identical— the layout is mirrored about the center symmetry line. To reduce labeling clutter, both halves of the layout are used for the mark-up labeling of different design components.

The RF input-to-output path is very short (400 μm), which along with compact 2-element matching networks, makes the group delay very short. G–S–G–G–S–G input/output RF pads have extra separation between the half-PA G–S–G halves for micro-probing configuration flexibility— two separate G–S–G probes). This extra ground pad and its associated ground-plane area can be removed later without any performance impact— the two sides of the PA halves can be brought together more closely until probe clearance limits them.

All transistor biasing lines are quarter-wave-length transmission lines above or below the global RF ground plane and provide $2f_0$ harmonic terminations at each gain-stage output.

The MQ layer is 0.55 μm thick and sits below the top two 1.25 μm-thick LY and 4.0 μm-thick AM aluminum layers. MQ is the uppermost copper layer in the metal stack and chosen as the global RF ground shield that isolates top-level AM-metal transmission-line structures from inducing currents in the lossy 13.5 Ω-cm p-type silicon substrate. Global DC ground is routed throughout the chip using wide layers (> 50 μm) of stacked AM, LY, and MQ tied together for extremely low resistances and inductances.

Deep-trench isolation moats and structures with substrate grounding rings are incorporated to prevent unwanted signal coupling between gain stages. Deep-trench lattice networks separate the two halves of the PA, de-coupling the substrates electrically and decreasing the amount of lateral heat spreading from the power-cell HBT devices. The entire PA is then isolated with isolation rings and moats, using alternating regions of high-resistance deep-trench isolation regions and well-grounded substrate-contacted conductive-substrate regions, so that PA can be integrated into a large array of similar elements or with other sensitive circuits (mixed-signal and digital) without causing excessive noise injected into adjacent circuits

2.4 PA Topology

The balanced/differential PA consists of two identical PA half-circuits driven differentially (0°/180°). The layout allows for simple testing with G–S–G–S–G differential probes and for differential integration into a higher-level, differential transceiver module. The input and output of each PA-half are matched to 50 Ω, resulting in a 100 Ω differential impedance referenced to the balanced G–S–G–G–S–G RF input/output pads (with north-south orientation).

Although the PA is differential (balanced), the analysis presented from this point forward will be based upon the single-ended PA (half-circuit), for simplicity. The design analyses apply to each half of the symmetric PA identically. Measurements are taken single-ended due to limited available source power. The differential output power is +3 dB more than the single-ended output power of the half-PA. Total gain and PAE of the half-PA, however, are

identical to the differentially-driven full PA ($G_{\mathrm{DIFF}} = G_{\mathrm{SE}}$; $\mathrm{PAE}_{\mathrm{DIFF}} = \mathrm{PAE}_{\mathrm{SE}}$).

Each PA half-circuit consists of three common-emitter gain stages, associated bias networks, and fully-integrated input, output, and interstage matching networks. These sub-components are described in detail in the following sections.

2.4.1 Common-Emitter Topology

Fig. 2.6 depicts a basic common-emitter (CE) single-transistor amplifier configuration with ideal inductive (RF-choke) biasing. RF matching networks and biasing networks are not included.

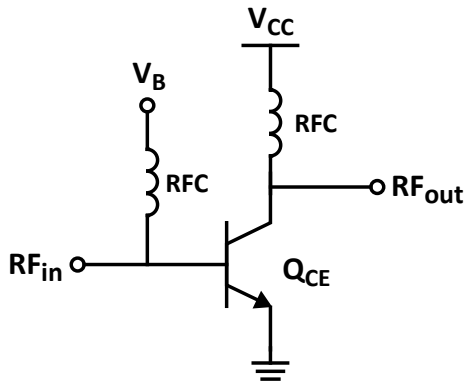


Figure 2.6: Common emitter (CE) topology with ideal inductive, RF-choke (RFC) biasing.

The common-emitter transistor is simple configuration to aggregate into larger power-cell transistors with manageable wiring parasitics. In a common-emitter topology, a key RF feature is that the AC emitter ground is also a DC ground. Thus, no additional AC-bypass capacitors are required, and the emitter is well-grounded with low inductance and low resistance, resulting in little emitter degeneration, which preserves RF gain & HBT performance. From a device layout standpoint, grounding the emitter in CE configuration is straightforward. DC and RF ground can be distributed on the same ground bus that directly contacts the emitters of each transistor unit cell without additional wiring parasitics.

The CE design is easily scalable to additional gain stages. Additional driver stages can be easily added using similar biasing, input matching, and interstage configurations, as is done in the PA Array in Chapter 4. The Ka-band PAs use two CE gain stages, the W-band PA scales to three CE gain stages, and the PA array scales to four CE gain stages. All gain stages use similar base & collector bias insertion, interstage matching networks, and layouts in all three major designs.

Fig. 2.7 compares the available gains (maximum stable gain MSG and maximum available

gain MAG) for a 5 μm × 0.12 μm transistor at $V_{CE} = 1.7$ V. The simulated high-performance transistor, in high-f_T C–B–E–B–C finger configuration, is biased near its peak f_T current density.

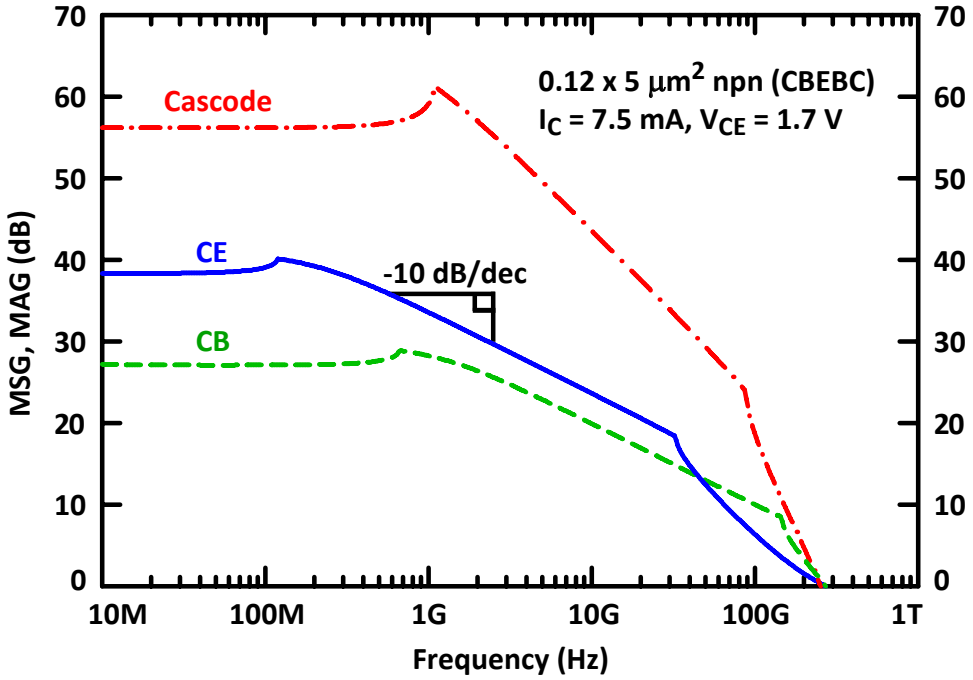


Figure 2.7: MSG and MAG for 5 μm transistor with collector current $I_C = 7.5$ mA and $V_{CE} = 1.7$ V, simulated with latest, updated HiCUM HBT models.

Although the cascode topology, which is a two-stage common-emitter and common-base configuration, is a completely valid design approach, simulations showed that the benefits over low-external base resistance-biasing of the common-emitter are small after parasitics are considered. The base node of the common-base stage needs to be AC-grounded, which is extremely difficult at 90 GHz, compared to emitter grounding in the common-emitter stage, which is both a DC & AC ground terminal. The cascode topology can offer 0.5 dB more output power, but the common-emitter PA is more easily tuned and optimized to push collector voltages higher in the lab since the cascode transistors have coupled voltages and currents.

See Appendix D for a more detailed discussion of the common-emitter versus cascode amplifier topologies and their design trade-offs.

2.4.2 Differential Amplifiers

Fig. 2.8 shows a simplified schematic of the implemented differential (balanced) W-band power amplifier. The physical layout on the SiGe chip mirrors the schematic. The two halves of the high-level schematic are symmetric about the ground nodes. The ground at the symmetry plane is part of both the global DC ground and the global RF ground and shared between the two differential PA halves. Each gain stage on each side of the differential PA has its own active-biasing control for fine tuning all quiescent currents to target values. A 0°/180° differential RF signal is applied at the input and results in a differential 0°/180° output. DC and RF ground is shared between the two PA halves of the differential PA. Each half of the PA can be tested independently.

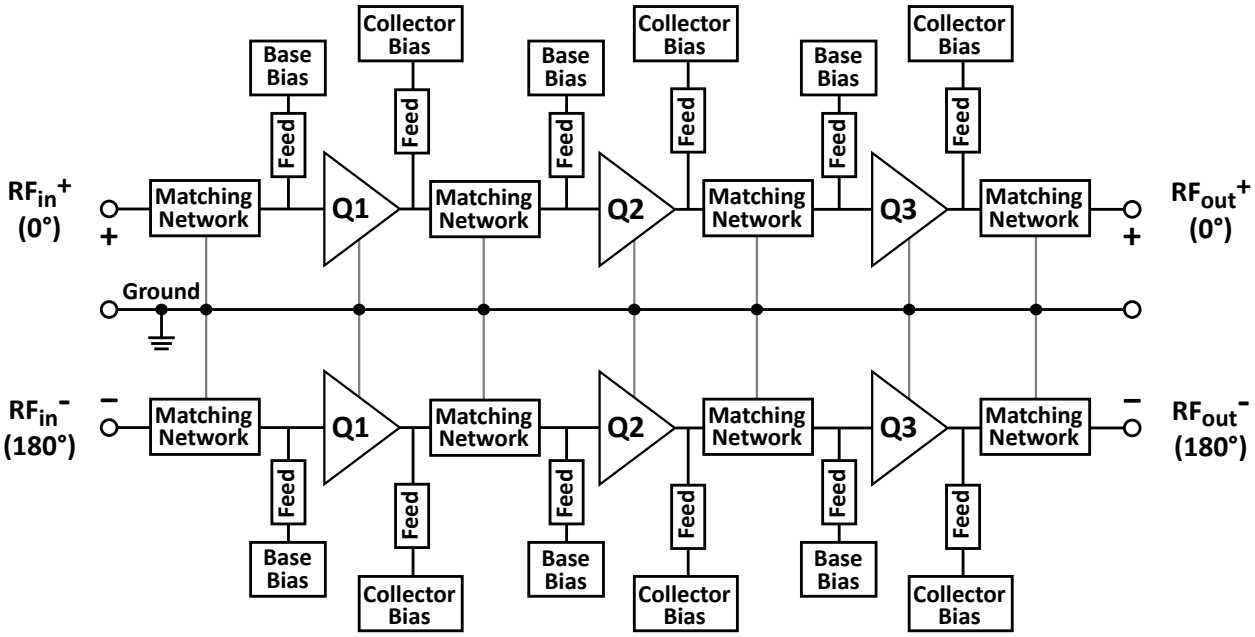


Figure 2.8: Differential PA block schematic as implemented in the W-band work. The signal ground around the point of symmetry is both the RF and DC global ground, so no common-mode rejection is present in the overall amplifier.

The power transistors of each half-PA are separate by 350 μm. The dedicated differential layout allows the power devices to be close together for better device matching. The two sides are separated enough such that no thermal coupling between halves is observable in measurements. They can be moved closer together as much as the input measurement probes can practically allow. The closer devices are together in distance, the better the device matching, which is especially important for the power cells in each half of the differential PA.

Even though there is no true common-mode rejection (see next Section 2.4.2 built into the

current differential implementation, careful layout symmetry and device matching (proximity of key HBT devices to each other) can lead to improved suppression of differential-to-common-mode signal conversion. The HBT power transistors from each half of the differential PA are separated by 350 μm limited by the RF input pad sizes and practical chip-probing limitations. (The differential 0° and 180° RF inputs and outputs are separate by 350 μm as well.) This ensures that HBT devices and passive components are well-matched components, locally. In an integrated transmission design— without RF probe pads—, the special differential floorplan allows the transistors from each half to be brought together as close as 150 μm, only limited by the resonant interstage matching networks, for maximum transistor matching characteristics not possible by simply placing two copies of a single-ended PA side-by-side. A side-by-side PA instance placement would result in the power HBT transistors being at least 700 μm apart with some high-current DC bias pads in the middle of the die, which is not desirable for wirebonding dies (or for on-chip microprobe testing).

To bring the two halves of the PA together so that the HBT power transistors main excellent matching between halves while keeping all RF and DC I/O at the periphery, base-biasing transmission-line feeds must be done underneath the RF global ground plane in custom inverted microstrip structures rather than on the low-loss, top-level AM metal Having the DC I/O at the outer edges simplifies PA measurements, allowing on-chip RF & DC micro-probing of the wirebond pads— the SiGe die does not need to be mounted to other substrate for measurements. The DC I/O placement also allows for additional integrated capacitors at the periphery, which is important for both high-frequency and low-frequency stability.

Common-Mode Rejection

Strictly speaking, most mmW differential PAs are only pseudo-differential not truly differential because the differential PAs provide no common-mode rejection (CMR) that is typical of lower-frequency, truly differential amplifiers. The common-mode ground is DC ground and not high RF impedance (as is the case for a differential pair with current-source tail).

To have truly differential common-mode rejection, common-node points need to be AC-grounded to differential-mode signals but high-impedance to common-mode signals. In lower frequency designs with enough voltage headroom (small-signal conditions or high-voltage technology), a high-impedance common-mode current-biasing tail can be placed at the common-mode of a differential transistor pair to achieve common-mode rejection [70]. The common-mode current tails consume too much voltage headroom that is that is essential for achieving high output powers in mmW silicon PAs, but other passive-only strategies can

achieve some common-mode rejection.

In pseudo-differential, balanced[4] operation, each of the two differential signals are referenced to ground independently, and the differential inputs have opposite phase (0°/180°). Each half is single-ended, effectively, or un-balanced, but combined to form a pseudo-differential, pseudo-balanced signal to drive a differential load (e.g, dipole antenna or another differential amplifier). (A truly differential signal would be independent to an absolute ground reference.)

In circuit with truly differential operation, the differential signals are referenced to each other and are independent of absolute ground. Moreover, there must be proper common-mode rejection (CMR) present to realize the full benefits of truly differential operation. True fully-differential amplifiers have common-mode rejection (CMR) with low common-mode gain and high differential-mode gain.

Fig. 2.9 illustrates the difference between pseudo-differential and truly-differential amplifiers with respect to common-mode rejection and AC grounding.

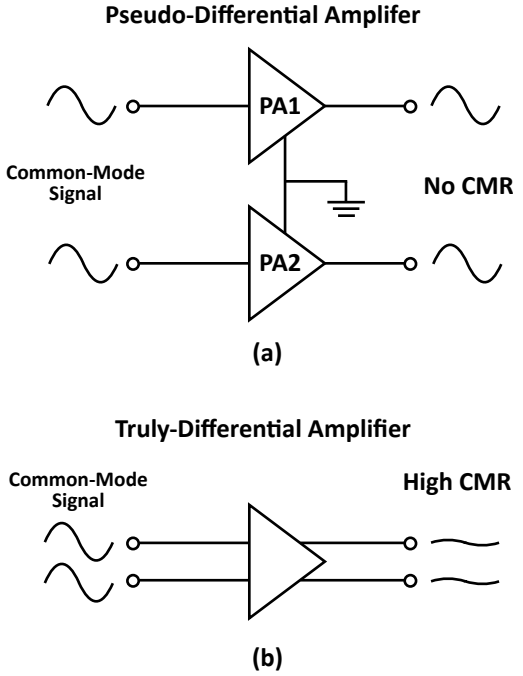


Figure 2.9: (a) Pseudo-differential amplifier with no common-mode rejection. (b) Truly differential amplifier with high common-mode rejection and internal virtual AC grounds at common-mode nodes.

The common-mode rejection ratio (CMRR) of a pseudo-differential amplifier is 0 dB— the

[4]Differential amplifier and balanced amplifier are differing terminologies for the same topology. Balanced amplifier has its usage roots from the microwave engineering community while differential amplifier has its roots from the more recent IC and RFIC circuit-design community. The two expressions may be used interchangeably.

differential-mode and common-mode have equal gains. By eliminating the shared common-mode current source, pseudo-differential amplifiers can achieve higher output swings at the expense of no common-mode rejection. The common-mode current source limits the output swing and thus output power of a truly differential amplifier.

If the differential output of a PA are terminated in 100 Ω, the differential-mode load impedance is 100 Ω (2× single-ended impedance). The odd-mode (single-ended PA or half the differential PA) load impedance is 50 Ω. The common-mode impedance is 25 Ω (half the single-ended impedance).

Conversions for Common-Mode

Even at sub-mmW frequencies, to maximize signal swing and output power in PAs, pseudo-differential/balanced circuit are still often preferred over truly differential amplifiers, especially when increasing output power is a primary design goal. The pseudo-differential amplifiers eliminate the common-mode current-biasing tail shared at the common-mode point between the amplifier halves. The addition of common-mode current sources reduces the output voltage swing considerably. At mmW frequencies, inductors may be physically small enough such that they can provide significant common-mode inductive degeneration and tail impedances in place of a current-biasing tail.

A truly symmetric, differential layout allows for possible DC bias injection at the symmetry point that is also AC-grounded. This is particularly beneficial for AC-grounding the base terminal in the common-base stage of a cascode pair, as demonstrated in [17].

The pseudo-differential W-band layout floorplan allows for truly-differential mode conversion in the future. A truly differential layout, with proper common-mode rejection, is not executed in the current design due to limited available input power for differential inputs. The full differential PA is powered on, but only one half of the PA is measured at a time in order to be able to saturate, properly, the HBTs in large-signal, high output-power conditions.

The pseudo-differential W-band PA can be converted to a truly differential PA with common-mode degeneration by altering the global grounding structure that sits between the two halves of the PA. Selective sections of the PA can be made into virtual AC grounds, resulting in some limited common-mode rejection for the PA. Currently, the global grounding between the two halves from input pad to output pad, connects all grounds on both sides of the PA together. Selective choices can be made as to which portions are fully differential and which parts are left pseudo-differential. For instance, there is option of placing common-mode rejection on the input drivers (Q1 & Q2) only while the Q3 out is still left pseudo-differential mode.

Fig. 2.10 shows, at the block level, how the pseudo-differential PA can be easily modified

to a truly differential PA with common-mode rejection. DC grounds and AC grounds can be separated such that certain common AC nodes can apply degeneration inductors without affecting the differential-mode RF signal. The common-mode gains of the active transistors are reduced via common-mode emitter degeneration while the differential-mode gain remains the same, as a virtual AC ground exists. Even-ordered distortions, even those created within the PA, are rejected as well in differential operation, which is normally beneficial for Class AB harmonic terminations. The amount of common-mode rejection can set by which blocks in the receive virtual AC-grounding only or common-mode inductance treatments.

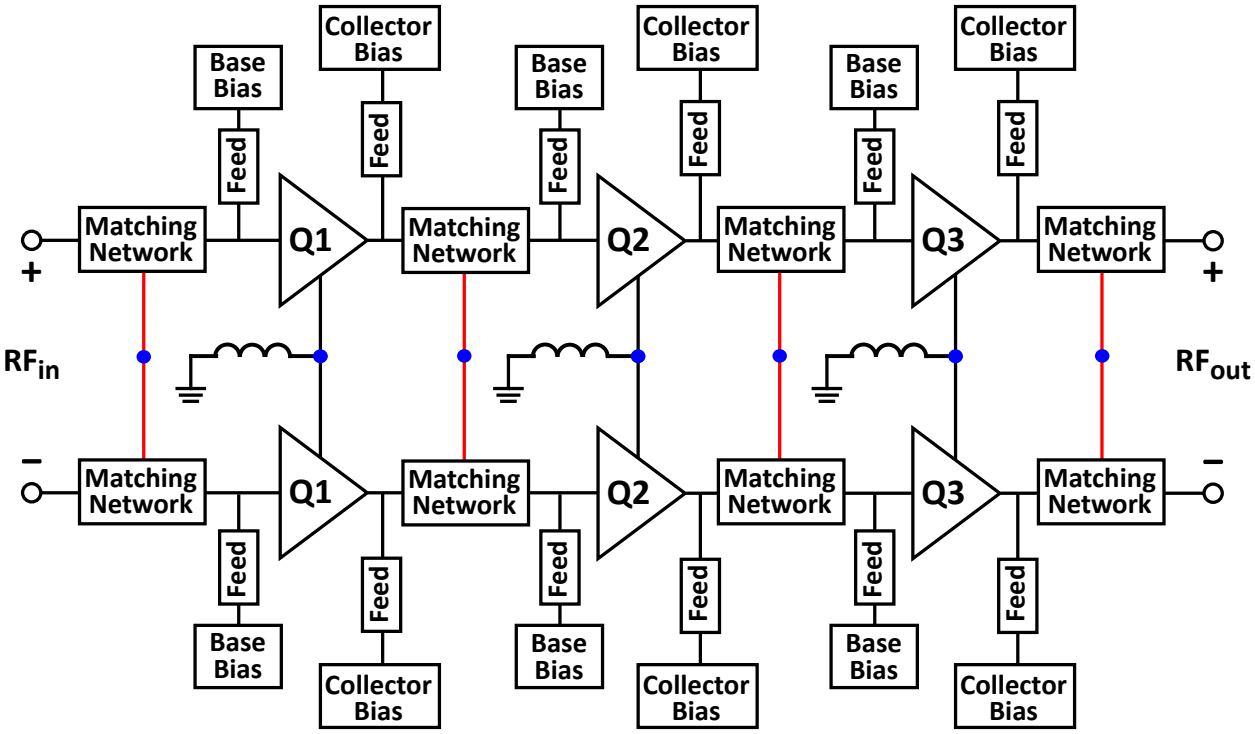


Figure 2.10: Block diagram of truly differential power amplifier with common-mode rejection. HBT gain stages have common-mode inductors to DC ground and virtual AC grounding. Red lines between the matching networks indicate connections that have no explicit DC ground— virtual AC grounding only. Blue dots indicate the nodes with virtual AC grounds from differential operation.

A common-mode inductance can be added to the tails of the each differential pair ($Q1^+/Q1^-$, $Q2^+/Q2^-$, $Q3^+/Q3^-$) to act as a shared, common-mode emitter-degeneration inductance. Common-mode inductors at the tail node in a differential pair can successfully act as common-mode degeneration without impacting the differential-mode signal. With truly differential operation and common-mode rejection *internally-generated* even-ordered harmonic distortions are cancelled since they are in-phase/common-mode for both PA halves.

This improves the overall PA linearity, especially operating into gain compression or deep Class AB operation.

Fig. 2.11 illustrates, at the transistor level, how a differential amplifier with true common-mode rejection can be implemented with a common-mode tail inductance as well as matching networks with virtual AC-grounding only.

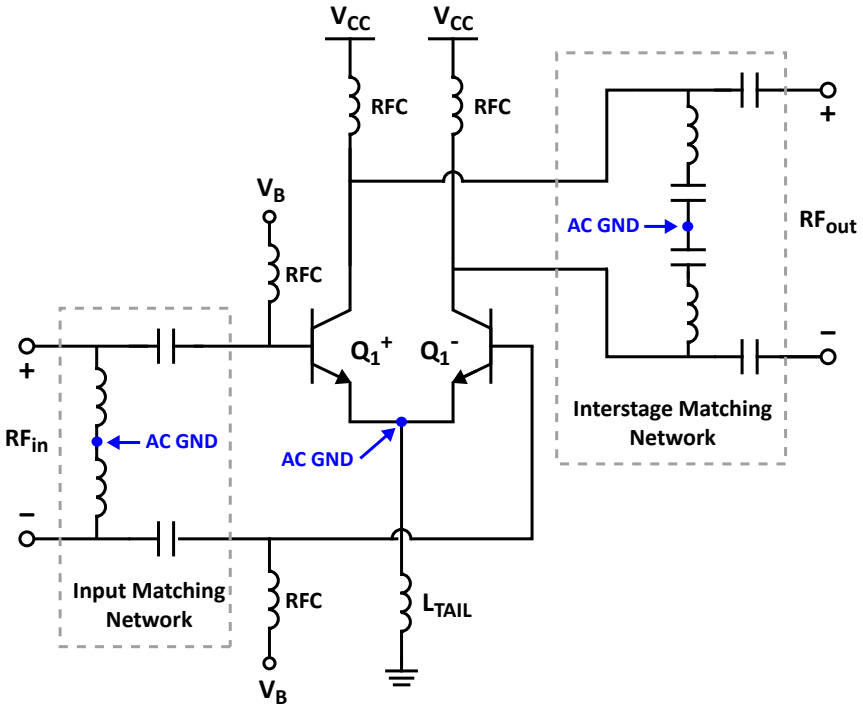


Figure 2.11: Circuit schematic for truly differential amplifier with common-mode rejection and virtual AC grounding points. A common-mode inductor presents emitter degeneration at W-band frequencies to a common-mode signal, only. Inductors at the common nodes provide common-mode degeneration.

2.4.3 Measurements: Single-Ended

The PA has been designed and laid out as a symmetric, differential amplifier. The W-band PA is designed to be driven with a differential input and to drive a truly differential 100 Ω load (e.g., antenna dipole element). Because of limited available source power, the output power measurement is done on just half of the differential PA— a single-ended PA.

A W-band balun, as a measurement provision, have not been integrated on-chip or used externally during measurements. The available source power at W-band is very limited, even with low-loss WR-10 waveguide components. There is not enough available source power from the measurement equipment to drive the differential power amplifier into compression,

so losses in the measurement input-chain must be minimized as much as possible. Eliminating the input balun increases the available source power into a single-ended PA measurement by 3 dB + balun losses (which can be 1 dB or more) for $>$ 4 dB additional input power. To utilize the additional input power, the differential PA needs to be measured in a single-ended manner.

The measurements for the amplifier are only done in single-ended mode with RF input and output only on one half/side of the differential PA. The other half of the differential PA is powered on (DC) to check that thermal effects between the two PA halves are isolated. Each PA half is measured separately single-ended, and results are within 0.2 dB measurement error.

Because of the measurement limitations and challenges, when available measurement input power is limited, common practice is to measure differential mmW PAs single-ended without internal or external baluns, and add +3 dB to the output power results to obtain the expected differential performance, as done in [14, 15, 17–19, 71–73] This methodology can be applied to differential integrated-circuit PAs where the differential layout provides symmetric circuits well-isolated from each other— no electromagnetic coupling between the two sides— resulting in negligible common-mode to differential-mode conversion. In [14], Pfeiffer, Reynolds, et al. validated this practice by taking fully-differential measurements with external input and output waveguide baluns and comparing the results to single-ended ones.

The expected, extrapolated differential output power is +3 dB more than the single-ended output power of the half-PA. Total gain and PAE of the half-PA, however, are identical to the differentially-driven full PA ($G_{\mathrm{DIFF}} = G_{\mathrm{SE}}$; $\mathrm{PAE}_{\mathrm{DIFF}} = \mathrm{PAE}_{\mathrm{SE}}$).

2.4.4 Multi-Stage Design

A multi-stage amplifier is required to achieve sufficient PA gain because the power gain of the HBT npn transistor is severely limited W-band frequencies (75–110 GHz). The maximum available gain (MAG) per common-emitter transistor stage is approximately 5 to 6 dB. The attainable gain per amplifier stage will be even lower than the single-transistor MAG value since wiring parasitics (e.g. emitter inductance & resistance), mutual thermal coupling effects within aggregated transistor arrays, matching-network losses, bias insertion feed losses, and output mismatch for large-signal load-line conditions account for further degradation in achievable gain. The losses of passive components and transmission lines are lower at 90 GHz than 30 GHz, characterized by a lower quality factor Q = 7–30. The low gain-per-stage limits the possible power-added efficiency since more amplifier stages are required to attain a

reasonable total PA gain and larger driver transistors (Q2) are required to provide sufficient input power into the final output stage power transistors (Q3).

Each PA stage must be able to provide enough output power to drive the subsequent PA stage to the compression point (P_{1dB}). Implementing such a scheme ensures that the last stage, which is the largest and most power-consuming stage, becomes the limiting factor for compression and not earlier PA stages. The driving stages, Q1 and Q2, are over-designed in order to ensure that the limiting portion of the circuit is the Q3 output power transistor. Additional PAE can be recovered from the design in future iterations by decreasing transistor sizes (and thus quiescent currents in the transistors) and bias-reference currents. PAE can also be improved by reducing quiescent current to bias the PA more into Class AB operation at the expense of linearity, gain expansion from small-signal conditions, and maximum output power P_{sat}. More details of the transistor sizing can be found in Section 2.9.2.

The common-emitter (CE) amplifier topology is chosen because of its high linearity and independent-bias tunability to adjust for HBT modeling errors as collector voltages are pushed higher in the lab. The common-emitter transistor is also very compact and simple to aggregate into larger power-cell transistor with manageable wiring parasitics. The emitter is both DC and AC grounded with the global ground plane, and additional AC/RF grounding via bypass capacitors are not required. Providing a low-DC external base resistance in the bias insertion network of the base terminal on the npn HBT allowed the CE topology to reach similar breakdown voltages as the common-base configuration. Proper bias design allows the CE configuration to boost the collector-emitter breakdown voltage from $V_{CE} = BV_{\mathrm{CEO}}$ = 1.8 V to near $BV_{\mathrm{CBO}} = 6.0$ V.

Power amplifier design is highly iterative, especially when dealing with interstage matching networks at high frequencies (relative to transistor f_T) where parasitics are significant. Key components and design discussions are covered in the sections to follow. The mechanics of the design process include choosing a bias point (both current and voltage), studying the simulated load-pull and source-pull responses of the design-kit HBT model for optimum source and load impedance for maximum output powers, designing the preliminary output matching network which affects the optimum source match of the output transistor, designing the preliminary interstage matching network between output transistor Q3 and Q2 to match optimum source load into Q3 to optimum output load presented to Q2, repeating this interstage matching process for Q2 and Q1, and designing the input matching network for Q1.

In the large-signal design and gain distribution, the output stage transistor, Q3, limits the compression and overall saturated output power. In other words, Q3 saturates (gain compression) before Q2 or Q1 in the prior gain stages. This ensures maximum utilization of

Q3 and improves peak PAE.

2.4.5 Linear Class A/AB Operation

The W-band PA is designed to operate in linear-mode Class A/AB (as opposed to other linear modes Class B/C or switched modes Class E/F & beyond). Class A/AB power amplifiers can deliver high linearity and high-gain in both small-signal and large-signal amplification, across a wide range of output powers. Class A/AB amplifiers allow the operational frequency to be pushed to the limits of the transistor gain better than other high-order classes [65]. At mmW frequencies, especially in W-band around 90 GHz, transistors' maximum available gain is relatively low ($<$ 6 dB in 8HP and other processes with $f_T \sim$ 200 GHz), and the Class A/AB topology becomes attractive to push the transistor technology to its frequency limits.

For purely Class A operation, PA gain drops gradually and monotonically from small-signal levels to the 1 dB gain-compression point. For Class AB biasing & operation, DC quiescent biasing currents are reduced from Class A values, and the gain may expand and increase from its small-signal case before the 1 dB gain-compression point. The gain expansion arises from self-biasing of the current at higher input powers, resulting in slightly more gain. The gain expansion with higher input powers leads to a degradation in signal linearity (IP3) compared to the monotonically decreasing gain compression characteristics of purely Class A amplifiers. Driver stages can be dynamically biased closer to Class A for better linearity while the output transistor can be biased deeper into Class AB operation . High output powers and high PAE can be achieved since the output transistor dominates the overall PA power and efficiency characteristics.

2.5 HBT Transistors

The PA design overcomes the technology's 1.8 V BV_{CEO} by presenting low external base impedances to the power transistor cells [74]. The external base impedance seen from Q1, Q2, and Q3 is 20 Ω at DC and ∼120 Ω (high impedance RF) within W-band frequencies. The resulting effective collector-emitter breakdown voltage allows the output voltage at the collectors to peak above 4.2 V, a factor of 2.5× improvement over BV_{CEO}.

In this work, the term "power transistor" refers to the HBT npn transistors Q1, Q2, and Q3— one power transistor per amplifier gain stage. Each power transistor is an aggregated transistor array comprised of multiple 5 μm × 0.12 μm unit cells (Q_0) arranged in a 1D (for Q1, Q2) or 2D (for Q3) array [5].

[5]The convention used in this text is to list the emitter length only when describing HBT device sizes since the

Only the VBIC (vertical bipolar inter-company) transistor models were available at the time of the W-band design, but newer HiCUM models (high-current model) provide valuable insights in subsequent discussions, especially at higher collector voltages where VBIC could not converge successfully even in simplified simulation schematics with relaxed convergence criteria. Large-signal harmonic balance simulations toward amplifier compression, in particular, are the most difficult to simulate with VBIC models at collector voltages > 1.5 V. VBIC models and HiCUM models correspond well at lower collector voltages < 1.3 V, but rapidly diverge with higher collector voltages and high currents.

The HBT transistor simulations presented here, unless otherwise noted, are using HiCUM models since they are better correlated to foundry-measured hardware data. More detailed discussion around the VBIC and HiCUM transistor-model limitations is covered in Appendix C.

The large-signal model around the aggregated power-cell transistor is covered in more detail in Section 2.9.3.

2.5.1 f_T and f_{MAX}

In this section, the GF 8HP npn HBT is examined for its high-frequency performance to determine the size of the unit cell and its biasing for peak mmW performance. The choice of the DC operating point greatly affects the resulting RF characteristics. One of the most well-known metrics for HBT high-frequency performance is transistor f_T. f_T is the transition frequency or cutoff frequency where the current-gain (AC short-circuited output) from the HBT base to collector drops below unity or in terms of hybrid-parameters (h-parameters), $h_{21} = 1$. The total transit time from emitter terminal to collector terminal is usually dominated by the base transit time, which is a strong function of base width and germanium grading that sets the built-in electric field.

The transistor f_T can be described in relation to the total transit times of carriers within the SiGe HBT, as shown in Eq. (2.1).

$$f_T = \frac{1}{2\pi(\tau_E + \tau_B + \tau_C + \tau_{CSCL})} \tag{2.1}$$

$$= \frac{1}{2\pi}\left[\frac{kT}{qI_C}C_{EB} + \left(\frac{kT}{qI_C} + R_C + R_E\right)C_{CB} + \frac{W_B^2}{\gamma D_n} + \frac{W_{CSCL}}{2\nu_{sat}}\right]^{-1} \tag{2.2}$$

τ_F is the forward transit time (emitter-to-collector transit time)

f_T is a useful metric for direct HBT technology comparisons because it is less sensitive

emitter width is fixed at 0.12 μm. A "5 μm HBT" implies a 5×0.12 μm^2 sized device.

to layout choices around the intrinsic transistor. In particular, base resistance R_B and C_{CB} capacitance do not affect the f_T metric. Transistor f_T is highly dependent on vertical process scaling while the parasitics R_{base} and C_{CB} are dependent on lateral scaling effects.

At a given collector voltage, f_T initially increases with increasing collector current I_C since base-transit times are reduced. After a certain collector current density, the HBT enters a high-current injection regime where the minority-carrier density is on the order of the collector doping concentration. The high current effectively widens the transistor's internal base width, causing an increase in the overall carrier transit times, reducing f_T. The decrease of f_T with high I_C is known as the Kirk effect.

The Kirk effect can be reduced by increasing the collector doping concentration, which helps to decrease carrier transit times, thus allowing for faster f_T transistors at higher currents, but the tradeoff is the collector breakdown voltage, BV_{CEO}.

Transistor f_{MAX}, the maximum frequency of oscillation, is also often used as a transistor metric and is closely related to f_T according to Eq. (2.3).

$$f_{MAX} \approx \sqrt{\frac{f_T}{8\pi R_B C_{CB}}} \tag{2.3}$$

f_{max} is sometimes a better metric than f_T for transistors in mmW PAs since it gives the maximum frequency that power gain is available. f_{MAX} takes into account the base resistance R_B and C_{CB} collector-base capacitance at the HBT base terminal. Modern HBTs generally have $f_{\text{MAX}} > f_T$, and devices are able to provide power gain through voltage gain even when current gain drops below 1. Vertical device process scaling reduces transit time, increases f_T, extends Kirk effect to higher current densities while lateral process scaling shrinks emitter widths and reduces intrinsic base parasitics R_B and C_{CB} to increase f_{MAX}.

Maximum Available Gain

In this section, the effects of device size, collector current, and collector voltage on f_T, f_{MAX}, and MAG are examined in detail for the GF 8HP SiGe HBT. The simulated data guides the choice of the HBT unit-cell size as well as target biasing points.

Unless specified otherwise, all transistor-characterization plots are for the high-performance npn HBT in C–B–E–B–C finger configuration with standard 0.12 µm emitter length. The simulation results are for common-emitter configuration driven by an ideal current source (presents an infinite external base DC resistance) at room temperature of 25°C. f_T & f_{MAX} are normally quoted for V_{CB} = 0–0.5 V bias, which corresponds to V_{CE} = 0.85–1.40 V, depending on the biasing current.

Fig. 2.12 shows the simulated peak f_T versus collector currents for different emitter

lengths (using the most recent HiCUM foundry models). All transistor widths are 0.12 µm, high-performance (as opposed to high-breakdown) devices in order to obtain necessary RF gains at 90 GHz. As emitter lengths become longer, the absolute currents for peak f_T also increase following a collector-current density of 1.35–1.50 mA/µm length emitter (or 11.25–12.5 mA/µm^2 per emitter area). Peak f_T drops with longer emitter lengths because of increasing intrinsic junction capacitances and parasitic capacitances. External wiring inductances and capacitances are not included here, and they will reduce peak-f_T even further for larger devices.

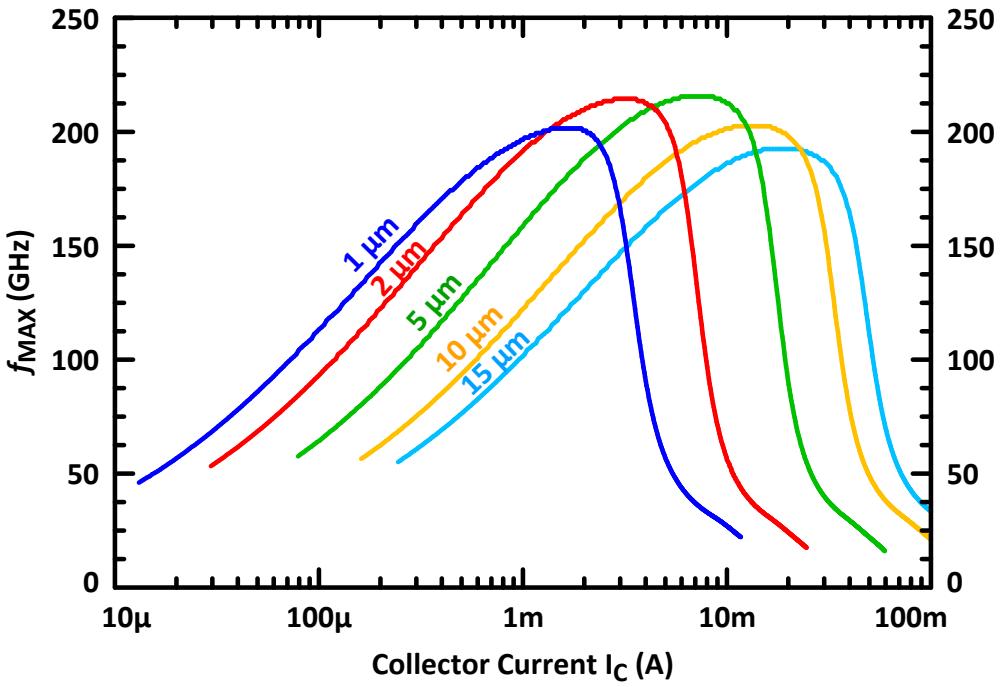


Figure 2.12: f_{MAX} vs. DC collector current I_C at $V_{CE} = 1.5$ V for various emitter lengths of the high-performance npn HBT. Emitter length = 1 µm, 2 µm, 5 µm, 10 µm, and 15 µm. All emitter widths are 0.12 µm.

Maximum available gain (MAG) is based on f_{MAX}, which gives the maximum frequency where the transistor can still generate oscillations (i.e., gain). MAG assumes ideal input and output matching conditions, so practical small-signal gains under output mismatches optimized for large-signal output will drop the maximum gain further, as will emitter degeneration, layout parasitics, and loss matching networks and bias feeds. The maximum available gain, MAG, can be calculated from the HBT's small-signal two-port S-parameters in Eq. (2.4).

$$\mathrm{MAG}\ (\mathrm{dB}) = 20\log_{10}\left|\frac{S_{21}}{S_{12}}\left(K - \sqrt{K^2-1}\right)\right| \tag{2.4}$$

where the stability factor K is given by

$$K = \frac{1 + |S_{11}S_{22} - S_{12}S_{21}|^2 - |S_{11}|^2 - |S_{22}|^2}{2|S_{12}S_{21}|} \tag{2.5}$$

MAG is valid when the transistors is unconditionally stable with Rollett's stability factor $K > 1$. MAG has a -20 dB/dec slope (-6 dB/oct) above frequencies where the transistor is unconditionally stable. At frequencies below which the HBT is not unconditionally stable ($K < 1$), the maximum stable gain (MSG[6]) is quoted, instead. MSG has a -10 dB/dec slope (-3 dB/oct) below frequencies where $K < 1$ and represents the maximum stable gain that can be achieved.

Fig. 2.13 is a plot of MAG at 90 GHz versus collector currents for different HBT emitter lengths at $V_{CE} = 1.5$ V. The peak MAG device is the 5 µm HBT if no wiring parasitics are taken into account. The peak MAG at 90 GHz for the 5 µm HBT is 7.0 dB at 7.5 mA current (at 1.5 mA/µm current density).

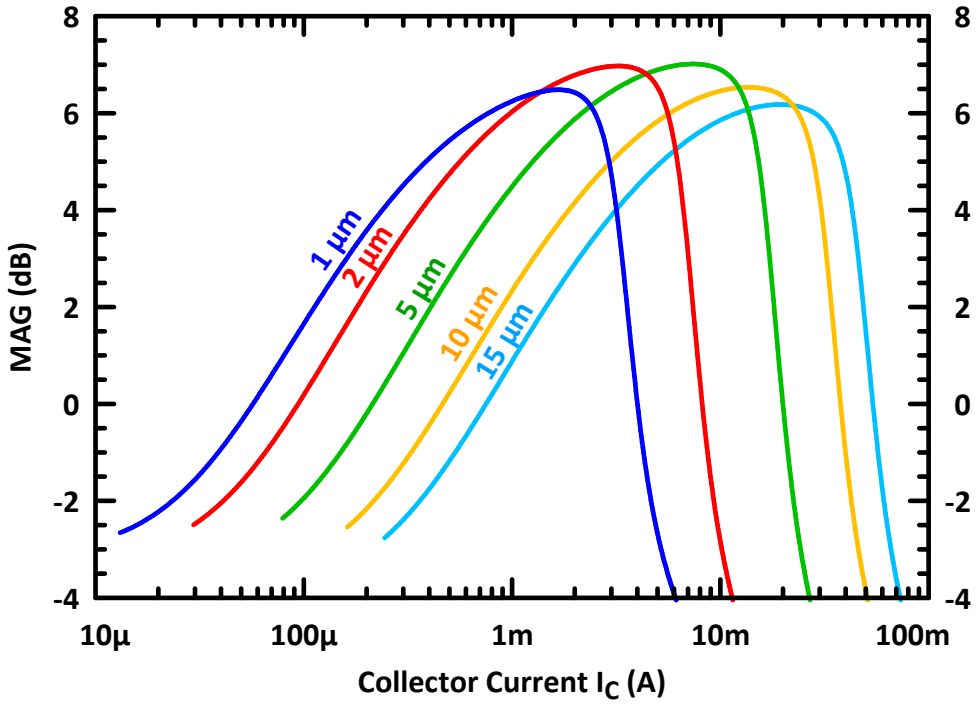


Figure 2.13: MAG vs. collector current I_C at 90 GHz for HBT emitter lengths 1, 2, 5, 10, and 15 µm. Wiring parasitics are not included.

Fig. 2.14 shows the corresponding f_{MAX} values across emitter lengths for the high-performance 0.12 µm-width npn HBT at $V_{CE} = 1.5$ V. Discontinuities are present the in

[6]See Appendix A for maximum stable gain equations.

graph because of discrete layout changes (wide via contacts) in the parameterized layout cell (p-cell). Peak f_T and peak f_{MAX} drop with larger emitter-length devices because of increased device parasitics. At $V_{CE} = 1.5$ V, the peak simulated f_{MAX} is 215 GHz for a 5 μm transistor. Putting multiple unit-cell transistors in parallel results in higher output powers—higher total-current operation—, but the maximum available gain drops because of additional inter-cell parasitics and mutual thermal coupling effects.

At high current densities, the Kirk effect reduces the f_{MAX} and MAG in the same manner that it impacts f_T. Increasing the DC collector current I_C above a certain current density will result in degradation of the maximum available gain of the device at mmW frequencies.

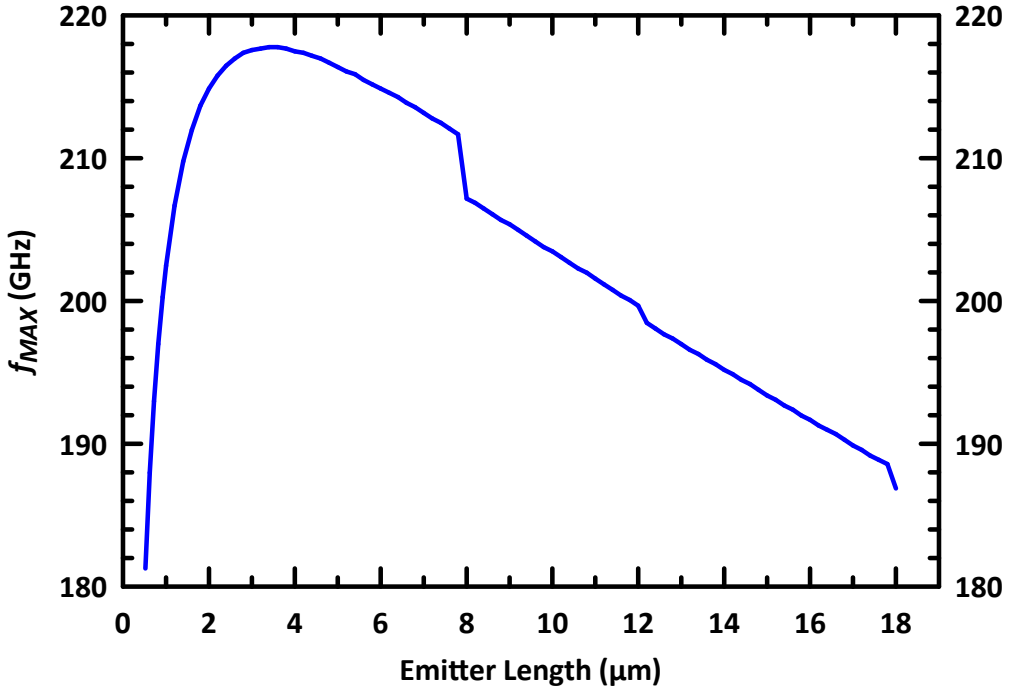


Figure 2.14: f_{MAX} vs. emitter length at $V_{CE} = 1.5$ V. The discontinuities (steps) in the chart arise from discrete-stepped geometry and collector-contact layout changes in the parameterized cell (p-cell) for larger emitter-length devices.

For a fixed HBT emitter length, increasing collector voltage V_{CE} slightly decreases the peak f_T, f_{MAX}, and MAG , as shown in Fig. 2.15 for peak MAG at 90 GHz.

Fig. 2.16 plots maximum available gain at different RF frequencies versus emitter lengths. At emitter lengths < 1 μm, MAG increases, but beyond 2 μm lengths, MAG tends to decrease with increasing lengths. At 90 GHz, MAG is < 7 dB for any emitter length before accounting for additional external wiring parasitics. Practical PA signal circuit implementations will move gain away from MAG because of additional external wiring parasitics and deliberate impedance mismatches for large-signal Class AB power matching— optimal large-signal

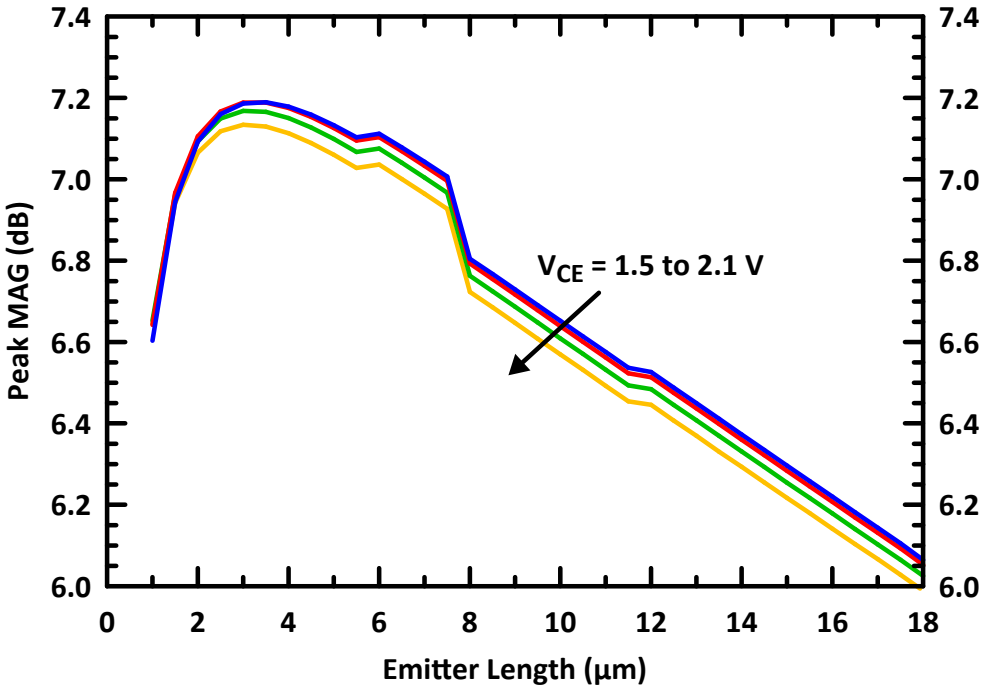


Figure 2.15: Peak MAG vs. HBT emitter length for V_{CE} = 1.5 V to 2.1 V in 0.2 V steps. The discontinuities (steps) in the chart arise from discrete-stepped geometry and collector-contact layout changes in the parameterized cell (p-cell) for larger emitter-length devices.

output-power matching is not small-signal complex-conjugate matching the transistor inputs or outputs for maximum power transfer.

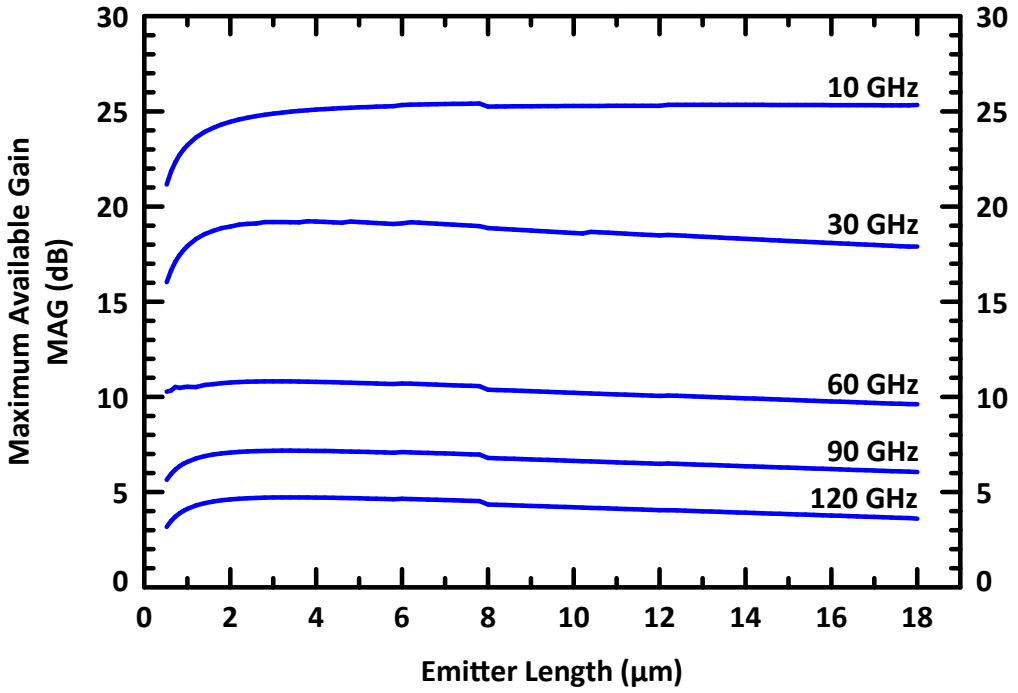


Figure 2.16: Maximum available gain (MAG) vs. emitter length at $V_{CE} = 1.5$ V for different GHz-range operating frequencies.

2.5.2 5 μm npn Unit Cell

In this section, the GF 8HP npn HBT transistors' DC and RF characteristics are discussed to provide insight into choosing an appropriate unit-cell emitter length used throughout the entire W-band PA design. All device characteristics are quoted for room-temperature operation (25°C) unless specified otherwise. All HBT devices used in the PA are of the npn type— no pnp transistors. For notational simplicity, since all 8HP HBT transistors have an emitter width of 0.12 μm, only emitter lengths are quoted– the unquoted emitter width is always 0.12 μm.

At the heart of power amplifiers, large-emitter-area HBT active devices are necessary to deliver the necessary current and power. In GF SiGe 8HP, these large-emitter-area devices must be created by aggregating multiple smaller cells. Very large devices (> 18μm) are not available as standard cells, custom HBT cells are not allowed, and the standard HBTs < 18 μm are not necessarily optimized for power operation. Large, multi-emitter-fingered

HBT devices (with multiple emitter stripes) are not available or allowed in GF 8HP because of current hogging concerns and thermal runaway within any multi-fingered HBT device.

A unit-cell approach is used for all the HBT devices. All transistors across the entire PA have similar collector current densities, thermal profiles, and parasitics that response to modeling errors, process shifts, and other non-idealities in the same manner such that the HBT characteristics track across the entire chip. Process skews and modeling non-idealities track across the entire PA chip.

All 8HP HBT p-cells are surrounded in a mandatory 1 μm wide and 6 μm deep-trench polysilicon & oxide (DT) rings for proper device operation. At the time of the W-band design, the foundry permanently changed and lifted key layout restrictions involving local-density restrictions of dense deep-trench (DT) isolation networks in the silicon substrate. The relaxation of deep-trench maximum-local-density rules allows tight packing of HBT transistors into dense 2D arrays. The DT region of one HBT can now be shared with an adjacent HBT transistor, and arrays of maximally, densely-packed transistors with all DT edges shared are permissible. Prior to this critical rule change, the maximum packing density of HBT transistors was two 1D arrays separated by > 32 μm, as shown for a 60 μm HBT power cell in Fig. 2.17 for the Ka-band design iterations. Having the DT restrictions for large transistors (> 40 μm) required vertical orientation of the two 1D arrays in the Ka-band design. More complicated, higher-parasitic interconnect wiring is necessary to feed these two arrays, and such a transistor layout does not work as well at W-band frequencies due to increased losses and phasing issues from larger structures (relative to signal wavelength).

At W-band, packing the transistors together more tightly is a means of reducing significant wiring losses and parasitics when using multiple HBT transistors in an array. Prior to the DT rule change, large output transistors at W-band were best synthesized by using large standard cell HBTs (e.g., 18 μm) and combining the outputs of multiple cells with transmission-line power splitters & combiners, which can have losses on the order of 1 dB at 90 GHz. The PA design is the first to explore aggressive 2D HBT packing in maximally-dense (abutted) arrays. Additional HBT modeling challenges arise from unmodeled, considerably more mutual heating among transistor unit cells, but wiring inductances/resistances are drastically reduced, and no power-splitting or combining structures are required within the power cells, which helps to preserve gain and maximum output power at these high mmW frequencies.

Choosing the HBT Unit Cell

The section reviews some of the considerations in choosing the unit-cell size for the HBT that is part of every transistor instance in the PA. For a given target aggregated transistor size (e.g., 40 μm), smaller unit-cells require more wiring overhead to aggregate together (more

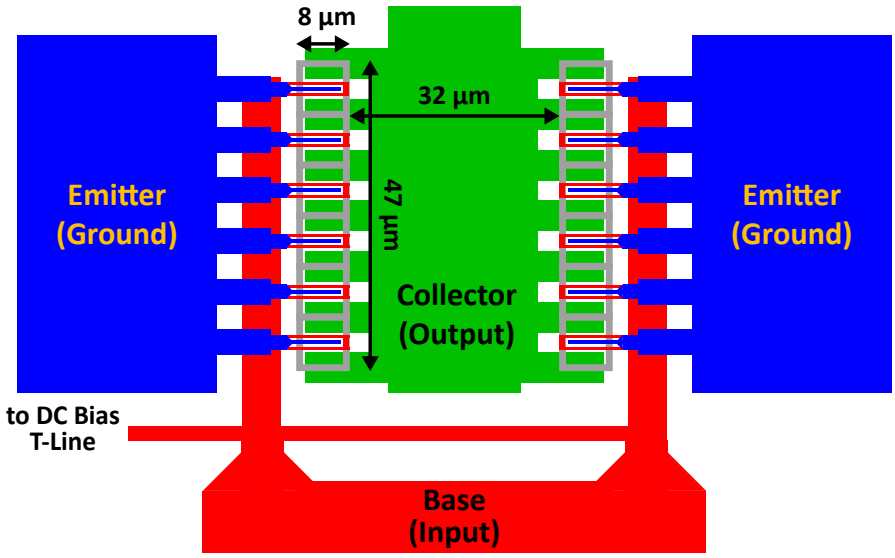


Figure 2.17: Ka-Band 60 µm power transistor layout. The original, former deep-trench local-density rules allowed abutted HBT cells in 1D arrays, but not 2D arrays. Two 1 × 6 1D arrays of 5 µm cells are separated by 32 µm, which is the required deep-trench separation under the old 8HP local-density rules. For clarity, the figure only illustrates connectivity, and metal layering is not detailed. The emitters and collectors are wired up from the substrate to the top-level (AM) thick metal. The base is wired up to the top-level metal, as well, away from the power-cell core, to connect to RF transmission lines.

resistance, capacitance, and inductance in the wiring), but less mutual-coupling thermal effects exist. From a simulation perspective, modeling the EM effects of the wiring overhead is much simpler than determining the electro-thermal effects of mutual thermal coupling among the active devices, but from a performance perspective, reducing wiring parasitics and need for power splitting/combining structures benefits the PA at W-band tremendously.

There are two high-performance npn HBT transistor layouts available: C–B–E and C–B–E–B–C contact fingers. While the C–B–E transistor is more compact and requires less inter-finger routing, the symmetric C–B–E–B–C HBT offers lower external base and collector resistances and lower self-heating thermal resistances. The symmetry around the emitter finger helps to ensure more even thermal profiles, which aids in stability in large-signal, high-powered operation. All transistors used in the PA are the C–B–E–B–C type.

Fig. 2.18 shows a comparison of the 40 μm power-transistor layouts from aggregating different unit-cell transistors: 10 μm, 5 μm, and 2.5 μm emitter lengths. The layout using 10 μm unit cells is 35 μm × 13 μm in area. The layout using 5 μm unit cells is 39 μm × 15 μm, which is 11% wider, 15% taller than 10 μm unit-cell layout. The layout using 2.5 μm unit cells is 39 μm × 20 μm, which is 33% taller than 5 μm unit-cell layout.

The 10 μm emitter-length cell is too large to use as a unit-cell current mirror because it wastes too much standing DC current in each PA stage, making high efficiency and PAE operation difficult. The self-heating and mutual thermal coupling among neighboring transistors is stronger in larger unit cells because there is less DT isolating the active regions. Larger cells have more issues with thermal stability and current hogging in densely packed arrays, as thermal variation on long emitter fingers and long collectors can be problematic for uniform current distributions with the HBT than shorter fingers. Finally, larger transistor models (< 3 μm) have more and more scaled modeling errors compared to foundry-measured hardware both at DC and at RF beyond 60 GHz.

The 2.5 μm HBT device, in contrast, is modeled more accurately, but associated wiring and deep-trench overhead increases the overall size of the aggregated transistor, adding additional inter-cell parasitics.

An additional consideration is that the 2.5 μm emitter-length unit cells also result in large 16:1 current-mirroring ratios when using a 2.5 μm device to control the bias of the 40 μm power cell. Lower current-mirror ratios (< 10) in PAs help keep the bias point more stable with larger standing currents in the bias circuits, especially under high RF-input drives when HBTs go into compression with resulting high base currents that the active bias must support

The 5 μm unit-cell size is a tradeoff between large devices that are more compact & have fewer wiring parasitics and small devices that are modeled more accurately and have better thermal coupling characteristics. The 5 μm unit cell is small enough that it can be

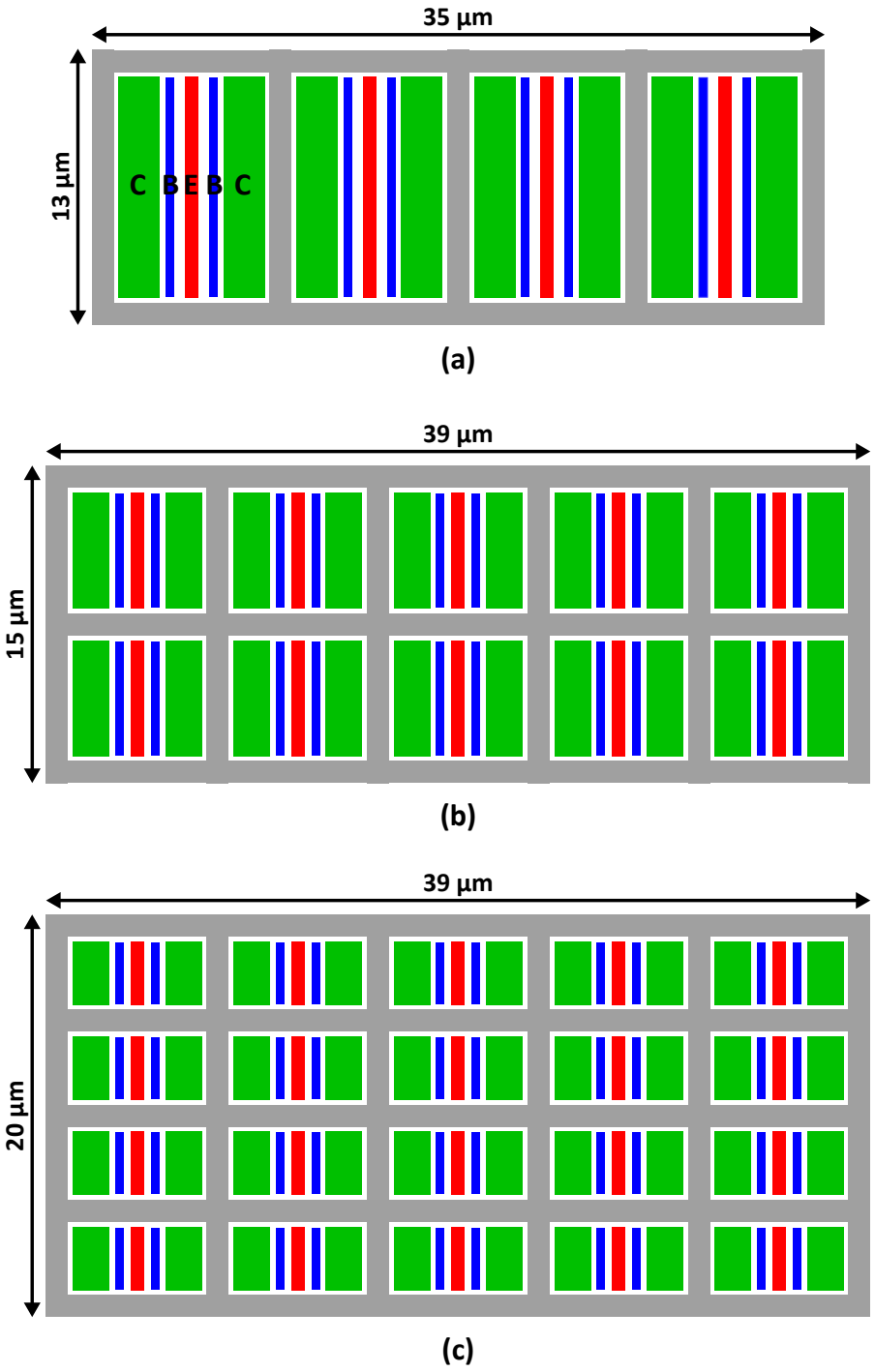


Figure 2.18: Aggregated 40 µm power-transistor layouts using (a) 10 µm unit cells, (b) 5 µm unit cells (implemented configuration), and (c) 2.5 µm unit cells. The required deep-trench (DT) isolation rings are in gray, and inter-cell wiring connections are not shown.

used within the active biasing as the reference transistor operating at the same high current density (1.5 mA/μm^2) with smaller impacts on overall power dissipation & efficiency. (For example, to bias at the same current density, a larger 10 μm unit cell in the biasing network needs to dissipate 2$\times$ the power over a 5 μm cell.)

The same considerations are made for choosing the earlier Ka-band design's unit cell, and using the same 5 μm unit cell at W-band has the advantage of having characterized data from the measured Ka-band designs.

5 μm Device

The 5 μm $\times$ 0.12 μm transistor standard parameterized cell (p-cell) is chosen as the unit-cell building block, Q_0, for all active transistors in the PA— from active bias-reference transistors (Q_{B1}, Q_{B2}, Q_{B3}) to power transistors (Q1, Q2, Q3). 17 instances of the 5 μm HBT unit-cell appear in the half-PA (34 for the total differential PA). The Q_0 = 5 μm HBT achieves a balanced tradeoff between size and accuracy of the available standard-cell npn models as well as a tradeoff between size and interconnect-wiring overhead. In the design kit, the 3–5 μm devices have the best hardware correlation for both DC and RF parameters, as shown by the foundry's measured characteristics. Above 5 μm, the model scaling for larger devices, up to 18 μm emitter lengths, does not produce as accurate of results— the model fitting is optimized around 3–5 μm. Fig. 2.19 illustrates the expected f_{MAX} variation over collector current I_C and collector voltage V_{CE}. In the common-emitter configuration, f_T increases with higher V_{CE} collector biasing. The forward transit time decreases with the widening of the base-collector space-charge region until high-level injection effects in the collector dominate.

The 5 μm unit cell transistor is also small enough such that excessive quiescent, standing bias current is not wasted in the biasing circuits for PAE concerns. The biasing transistors, altogether, dissipate 20% of the DC quiescent power consumption. This is the maximum fractional contribution to power dissipation when no input signal is present. At higher RF input drives and PA output powers, the bias circuit contributes $<$ 5% of the total DC power consumption.

Fig. 2.20 illustrates the chosen 5 μm unit cell used throughout the power amplifier design. The unit cell is a standard 8HP high-performance transistor cell (as opposed to the special high-breakdown, lower f_T device).

The HBT unit cell has a single emitter stripe 0.12 μm wide and 5 μm long. The entire 5 μm unit cell measures 8.7 μm $\times$ 8.1 μm (DT-bounded) and is configured in a C–B–E–B–C finger arrangement. Dual base stripes flank the emitter and dual extra-wide collectors, for increased reliability, contact the HBT subcollector region in the substrate. A required

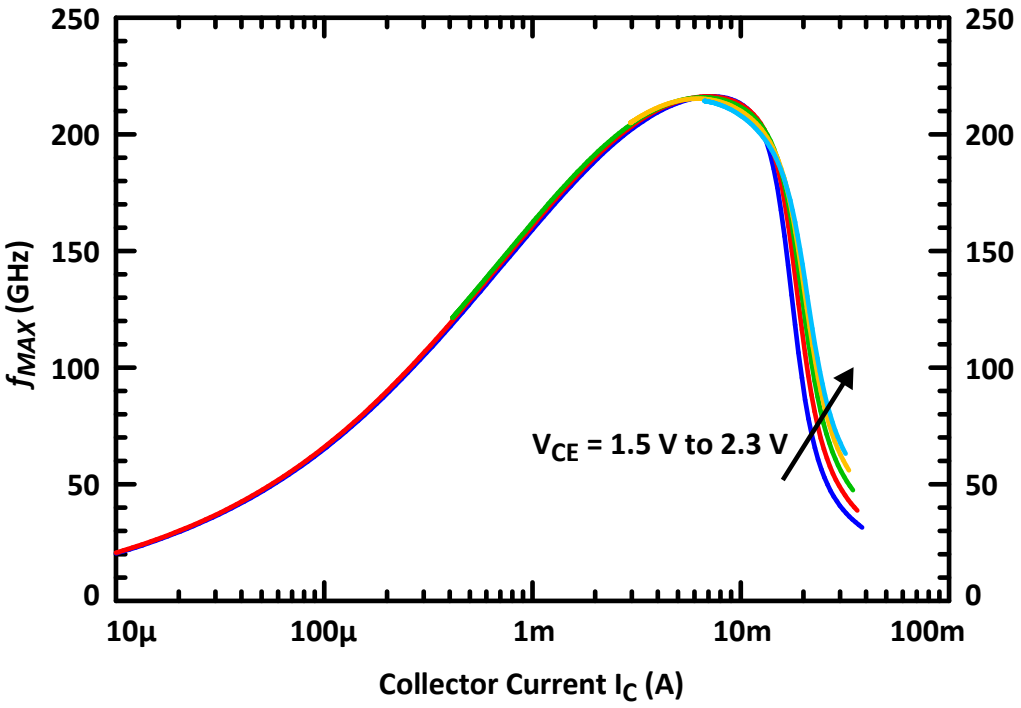


Figure 2.19: f_{MAX} vs. collector current I_C for 5 µm emitter length HBT device. V_{CE} = 1.5 V to 2.3 V in 0.2 V steps.

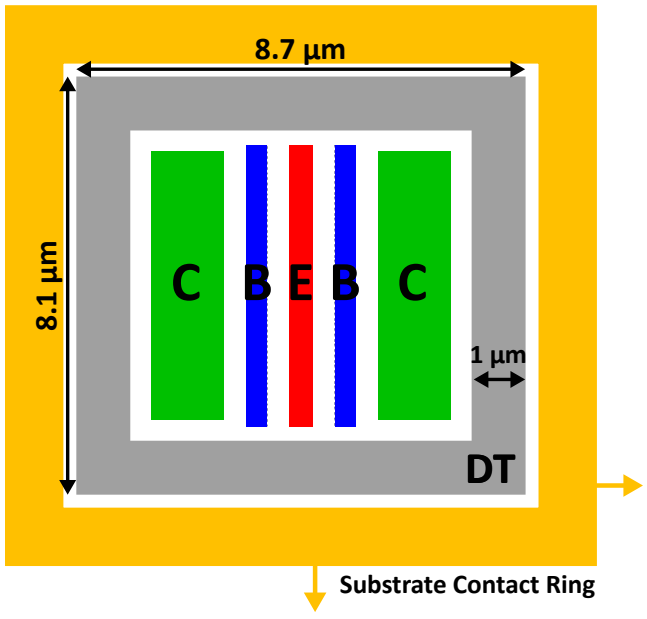


Figure 2.20: Chosen 5 µm unit-cell transistor layout with required deep-trench (DT) isolation ring. The outer boundary of the surrounding substrate-contact ring (in yellow) is not shown. The HBT fingers are arranged C–B–E–B–C with a single emitter strip, dual base stripes, and dual extra-wide collector stripes for reliability.)

1 μm-wide deep-trench (DT) ring surrounds the entire HBT structure with a 6 μm-deep isolation oxide & polysilicon. The deepest active layer within the HBT extends only 2.6 μm into the substrate, and the designer must also place wide substrate-contact rings around the HBT cell for proper substrate grounding and device isolation.

DC Characteristics

Fig. 2.21 is the Gummel plot (non-log scale) of the chosen 5 μm unit cell in common-emitter configuration. The plot shows the transfer function of the HBT base-emitter voltage, V_{BE} to collector current I_C with $V_{CC} = 1.5$ V. The shape of the input-output transfer function (V_{BE} to I_C) changes with collector voltage, $V_{CE} = V_{CC}$. $V_{BE,\mathrm{ON}}$ for the 5 μm device is approximately 890 mV at peak-f_T/peak-gain I_C current densities at the nominal $V_{CC} = 1.7$ V operating point. As collector voltages approach 2.3 V, biasing becomes very sensitive to V_{BE} voltages. Small changes in the V_{BE} DC bias point result in large changes in I_C. Integrated on-chip active bias circuits address this bias sensitivity issue in Section 2.6. Separate external biasing controls for each active bias allows precise current control over each gain stage of the PA (on each half of the differential PA). Since RF performance is tightly coupled to the DC bias point, achieving the target bias conditions is the first step in obtaining expected PA performance.

Self-heating effects within the standalone 5 μm unit cell HBT cause the DC-IV curve of the transistor to sag, at a given base biasing, with increasing collector voltages, V_{CE}. Fig 2.22b illustrates this effect with thermal effects simulated in isolation of impact ionization— no impact ionization modeled.

Fig. 2.22a shows the effects of transistor self-heating and impact ionization together. Fig. 2.22c shows the effects of impact ionization in isolation— no thermal self-heating effects. Lastly, Fig. 2.22a is an idealized DC-IV curve of the 8HP HBT with no self heating and no thermal self-heating effects. Impact ionization causes the collector current I_C to increase rapidly beyond a critical operational collector voltage V_{CER} while self-heating thermal effects cause the DC-IV curves to sag at higher power dissipations.

RF Characteristics

For PA designs in SiGe at W-band frequencies, power gain is one of the most important RF characteristics of the HBT, especially since available gain is low in GF 8HP technology (< 7 dB) around 90 GHz. The HBT transistor RF characteristics are highly dependent on f_T & f_{MAX}, which are themselves strongly dependent on the DC bias currents and voltages. In the W-band PA design, all transistors, power transistors and associated transistors in

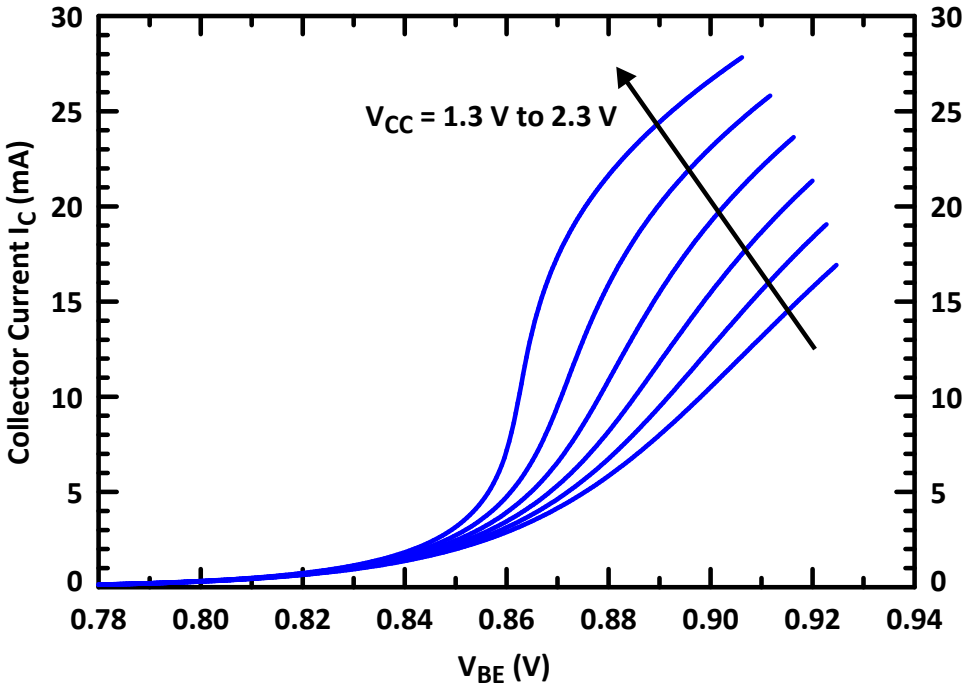


Figure 2.21: Gummel plot (I_C vs. V_{BE}) for 5 μm HBT unit cell. V_{CC} = 1.3 V to 2.3 V in 0.2 V steps. As higher collector voltages are used, the collector current is more and more sensitive to changes in V_{BE}.

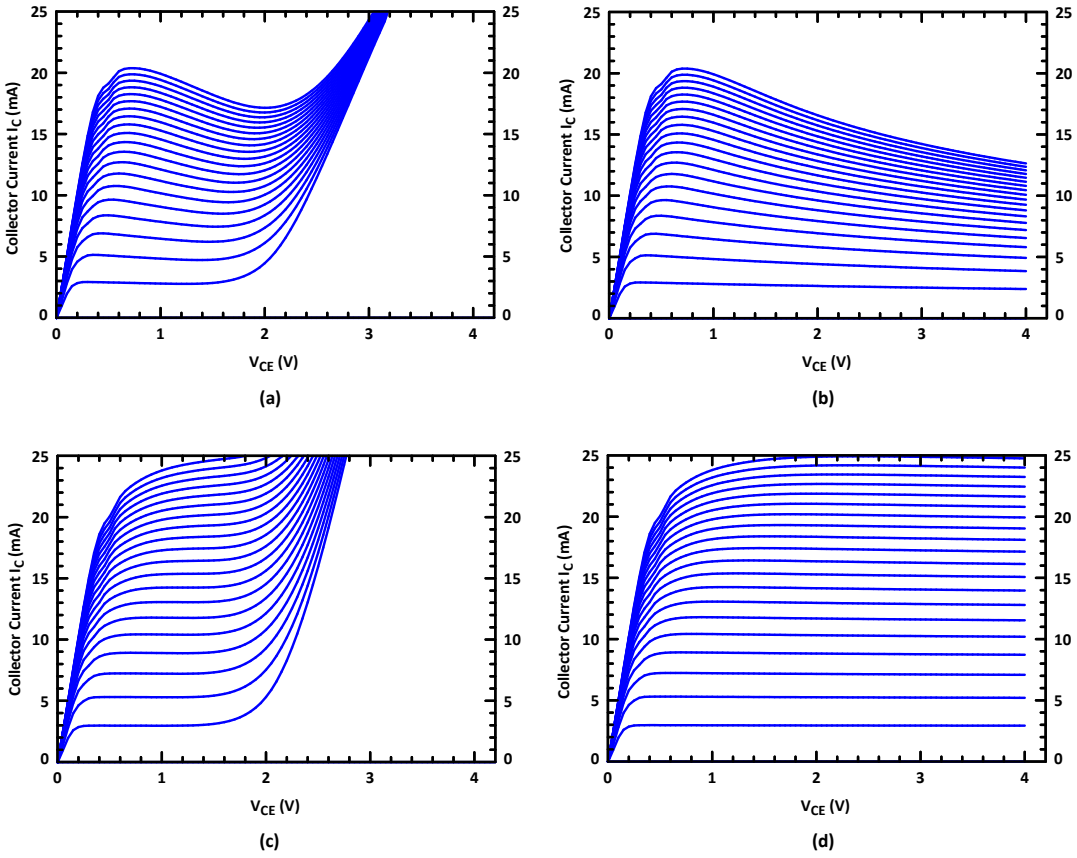


Figure 2.22: (a) Nominal simulated DC-IV curves for 5 μm npn HBT using HiCUM models under open-base biasing condition (base input is current-driven in 20 μA steps), includes impact-ionization and self-heating effects, (b) without impact ionization and with self heating, (c) with impact ionization and without self heating, and (d) without impact ionization and without self heating (ideal curves). The plots show the isolated effects of impact ionization and self-heating on the DC-IV curves.

the active-biasing circuits, operate near their peak f_T/f_{MAX}/MAG current density[7]. For the 5 μm unit-cell HBT, peak f_T biasing occurs at 1.5 mA/μm bias current density is per emitter length (with fixed 0.12 μm width), which is equivalent to 12.5 mA/μm^2 current density per emitter-area[8]. The foundry-measured peak f_T and f_{MAX} for the 5 μm device is approximately 200 GHz and 285 GHz, respectively, at $V_{CB} = 0.5$ V.

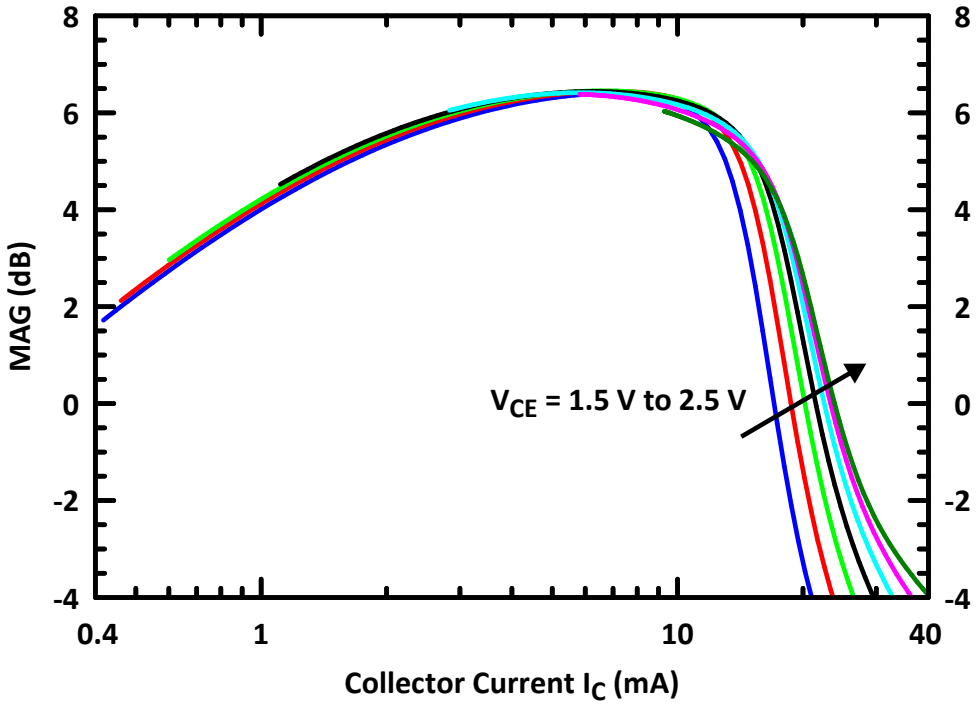


Figure 2.23: MAG vs. collector current I_C for 5 μm unit-cell with $V_{CE} = 1.3$ V to 2.5 V in 0.2 V steps.

Finally, Fig. 2.23 show the MAG of the 5 μm unit cell versus different collector biasing for a family of V_{CE} collector voltages.

At high current densities, the high mobile-charge density alters the electric-field profile within the collector-base space-charge region and causes an effective widening of the base region, known as the Kirk effect. At high HBT current densities, the Kirk effect degrades both transistor β_F (base-to-collector current gain) and f_T/f_{MAX}/MAG, dropping the high-frequency gain. The Kirk effect results in an f_T falloff at high current densities because of spreading & widening of the neutral-base layer into the collector region [75]. The effect on

[7]For HBTs in PA Class A/AB operation, the biasing current density for peak f_T, f_{MAX}, and maximum available gain (MAG) are the same— biasing for peak f_T is equivalent to biasing for peak f_{MAX} or MAG.
[8]The newer HiCUM transistor model now predicts peak f_T biasing at 6.8 mA for 5 μm HBT, which corresponds to 1.36 mA/μm per emitter length or 11.3 mA/μm^2 current density per emitter area, slightly lower than the 12 mA/μm used and predicted by the early VBIC HBT model.

f_T can be delayed to higher current densities through heavier collector doping, but this leads to a tradeoff of reduced collector breakdown voltages.

The simulated AC β (also called β_{AC} or h_{21}) for the 5 μm bias-reference transistor at peak-f_T currents is shown in Fig. 2.24 and Fig. 2.25. The DC β_F at currents near peak-f_T is 170 (and even higher, > 400, at very small currents). β @ 90 GHz is about 2.3 for 1.5 mA/ μm for 5 μm npn @ $V_{CE} = 1.5$ V

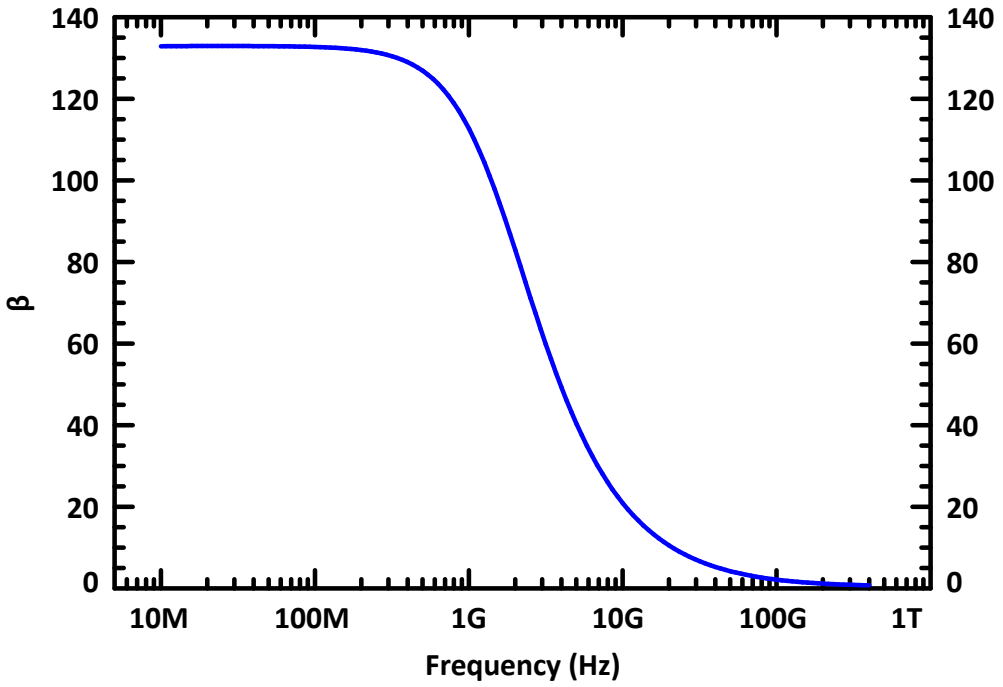


Figure 2.24: β_{AC} at peak f_T biasing. Emitter length = 5 μm, $I_C = 7.5$ mA @ $V_{CE} = 1.5$ V.

Emitter Inductance

Fig. 2.26 shows the effects of emitter inductance on MAG over frequency. For a 5 μm device, EM simulations indicate that emitter ground inductance for an individual 5 μm unit cell is approximately 12 pH to wire all the way up to top-level AM-metal DC/RF ground. This parasitic ground inductance results in a drop in MAG by more than 1 dB due to inductive degeneration. Larger transistors, when used in aggregated arrays, have additional inter-cell wiring inductances, and these additional parasitics must be taken into account via EM simulation when connecting multiple HBT unit-cells together to form larger, aggregated power-cells.

Low-inductance emitter-ground wiring is necessary to preserve the little gain available at W-band. A wide, thick top-metal ground bus, pulled across the top of the transistor array to

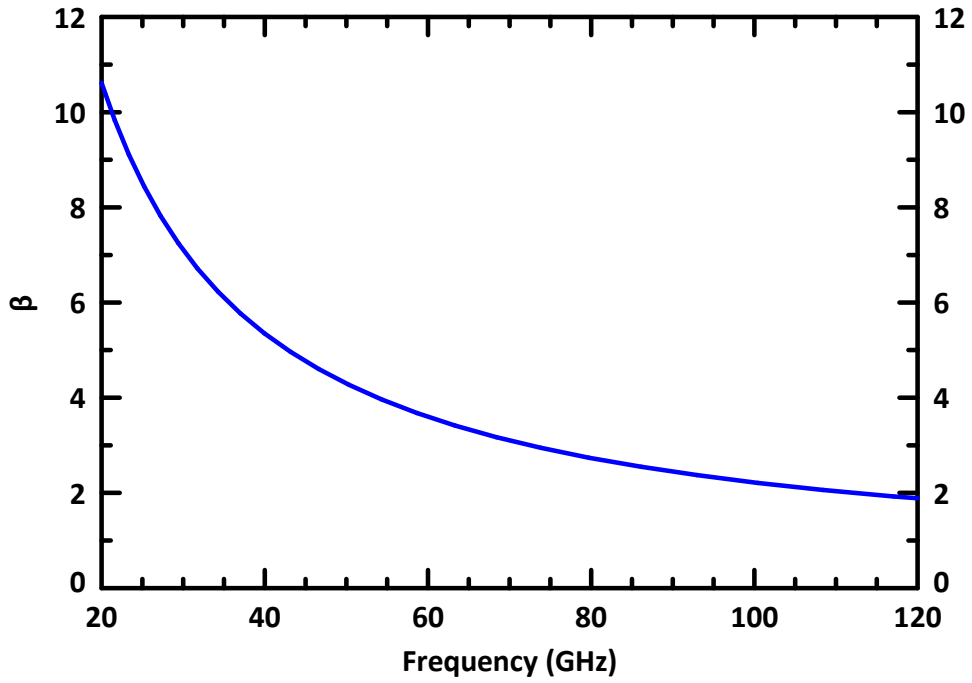


Figure 2.25: β_{AC} at peak f_T biasing currents. Emitter length = 5 μm; I_C = 7.5 mA @ V_{CE} = 1.5 V.

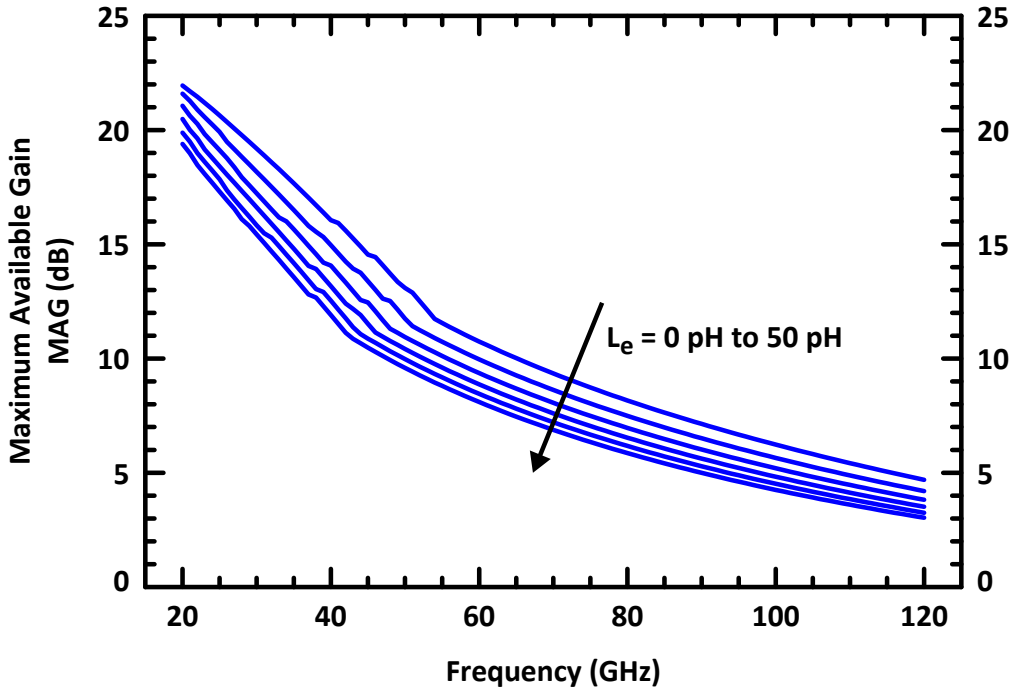


Figure 2.26: Maximum available gain (MAG) vs. frequency for emitter inductance = 0 pH to 50 pH in 10 pH steps. Emitter length = 5 μm, V_{CE} = 1.5 V.

contact the emitters, is necessary to minimize the ground inductance.

2.5.3 Breakdown Voltage Limitations

The section discusses the breakdown characteristics and limitations of the chosen 5 μm unit-cell HBT transistor. Since power amplifiers' primary function is to convert DC into large RF/AC waveforms, the HBT's DC current and voltage operating points have great influence on the overall performance of the PA. There are two primary operational boundaries in HBTs within power amplifiers: voltage limitations and current limitations. Voltage limits are determined by collector breakdown voltages while operational current limits of the HBT are primarily determined by electromigration and thermal-dissipation limits of connecting metal lines and vias.

In addition to limited gain at high frequencies, the breakdown voltage is a significant limitation for high-powered mmW PA design in advanced SiGe technology. The collector is very heavily doped in order to shorten the charge transit times and increase the speed and gain of the device at high frequencies (f_T), and but at the expense of decreased collector-base junction breakdown voltages. Heavier collector doping pushes out the peak-f_T to higher current densities, which is helpful for high-current PA operation, but the collector-emitter breakdown voltage, BV_{CEO}, decreases.

The Johnson limit describes the fundamental tradeoffs between an npn transistor's f_T and its breakdown voltage BV_{CEO} arising from the collector doping [76]. The limit describes the first-order relationship between BV_{CEO} and f_T based on the semiconductor material properties of the bipolar transistor.

$$BV_{\mathrm{CEO}} \times f_T \propto \frac{E_{max}\nu_{sat}}{2\pi} \sim 200 \text{ GHz-volts} \tag{2.6}$$

where E_{max} is the dielectric breakdown strength of a semiconductor material and ν_{sat} is the saturated drift velocity of carriers.

The $f_T \times BV_{\mathrm{CEO}}$ product is approximately constant for a given collector material. The limit assumes uniform collector doping and is conservative for predicting breakdown voltages of modern HBTs. For silicon homojunction bipolar junction transistors (BJTs), the Johnson limit is approximately 200 GHz-V, based on simple one-dimensional drift-diffusion models.

Modern HBTs have collectors with more sophisticated, non-constant doping profiles and strained & graded bases, and heterojunctions. HBT devices have successfully continued to increase the $BV \times f_T$ product with each successive generation of new devices [1, 67]. A modified Johnson limit for current high-performance SiGe 8HP HBT devices is on the order of $\sim$ 300–350 GHz-V.

The Johnson limit illustrates the trends of fundamental device limits and reveals that there is a gain-power limit for each technology. Fortunately, a circuit-design work-around exists to 8HP's low $BV_{\text{CEO}} = 1.8\ V$ collector-breakdown voltages because the assumed npn open-base condition in the Johnson limit and BV_{CEO} collector breakdown voltage can be modified with a proper low-resistance external bias circuit to push collector-emitter breakdown voltages closer to the 8HP collector-base breakdown voltage $BV_{\text{CBO}} = 6.0$ V. Details of the bias circuits to achieve higher collector breakdown voltages are found in Section 2.6.

Operational Collector Voltage BV_{CER}

This section discusses the high-level underlying physics of carrier events in the HBT that determine the collector breakdown voltage. Attaining higher operation collector breakdown voltages is possible through circuit techniques in the external base-biasing configuration.

Beyond a critical V_{CE} value, impact ionization events and avalanche multiplication in the collector-base space-charge region (SCR) cause collector currents to increase rapidly until avalanche breakdown. The operational collector-emitter breakdown voltage, BV_{CER}, is highly dependent on the on the external[9] base-biasing resistance at the base of each power transistor terminal, which ballasts currents created by impact ionization at higher collector voltages.

Electrons in the base-collector (BC) space-charge region (SCR) usually experience some scattering before reaching the collector region. The amount of scattering is dependent on the collector-to-base bias voltage V_{CB}, which is directly related to V_{CE} in common-emitter configurations biased near peak-f_T operation. With enough energy (1.1 eV for a silicon npn collector), electron scattering within the base-collector space-charge region can generate an electron-hole pair from the silicon lattice in an impact ionization event. The generated electrons drift toward the neutral collector region while the generated holes drift into the base region. The electrons are eventually swept through the collector, causing a rise in collector currents, while the holes recombine or exit depending on biasing configurations. Under ideal open-base current-source biasing and high external resistance (kΩ) base-biasing conditions, the excess holes cannot exit the base terminal and instead, recombine with electrons being injected from the emitter. The excess recombination results in a generation of even higher emitter currents, which results in more electrons flowing into the base-collection region and more electron-hole pairs being generated by impact-ionization scattering events in the collector space-charge region. The now-multiplied number of excess holes then recombine to causes a rise of even more emitter current– a positive feedback process known as avalanche multiplication. This collector current is known as the avalanche current, and the rapid rise of

[9] "External" resistance refers to resistance outside of (external to) the HBT standard p-cell, beyond the M1 first-metal layer.

current is known as avalanche breakdown. In isolation– without other stress effects such as thermal instabilities–, avalanche breakdown itself is a non-destructive, reversible breakdown process. Avalanche breakdown is an operational boundary and not necessarily a destructive limit.

With an infinite external base resistance— HBT biased and forced with a base current set by an ideal current source— the DC-IV curves experience avalanche breakdown effects from impact ionization processes in the collector near V_{CE} voltages of $BV_{\mathrm{CEO}} = 1.8$ V. The precise voltage at which avalanche breakdown effects dominate depend on the DC bias-current point, and higher DC currents result in lower breakdown voltages ($< BV_{\mathrm{CEO}}$). In addition, higher current-gain devices (higher β_F) result in higher current feedback, which leads to reduced collector-emitter breakdown voltages. The feedback effects can be eliminated or significantly mitigated by allowing the excess holes generated by impact ionization to exit the HBT base terminal via a low-resistance external DC biasing resistance. The generated holes ballast through the base terminal rather than recombine with electrons injected from the emitter, breaking the feedback loop.

By using biasing closer to voltage-mode (low external base resistance) rather than current-mode (high external base resistance), the operation breakdown voltage BV_{CER} increases from $BV_{\mathrm{CEO}} = 1.8$ V toward the ultimate $BV_{\mathrm{CBO}} = 6.0$ V limit (open emitter). With proper low-resistance base biasing, the common-emitter topology can approach the breakdown voltage of the common-base transistor

8HP also has optional high-breakdown $BV_{\mathrm{CEO}} = 3.6$ V HBT devices created from selectively altering the collector doping profiles with lighter doping implants. These high-breakdown devices have only peak f_T below 57 GHz, and do not provide any RF gain for W-band operation. Thus, only the "high-performance" high-f_T HBTs with $BV_{\mathrm{CEO}} = 1.8$ V are considered in this W-band design.

In the implemented PA design, a very low external base impedance of 20 Ω is presented to each HBT power cell while maintaining high RF impedances with a custom inverted microstrip transmission line and resonant RF-shorting MIM capacitor. The active bias network maintains stable operation and pushes the resulting $BV_{\mathrm{CER}} > 5.0$ V, well beyond the 1.8 V BV_{CEO}. 20 Ω is the lowest external base resistance known to be used in a W-band SiGe HBT design and allow the common-emitter configuration to be pushed to its device limits. More details of the implementation are found in Section 2.6.

Fig. 2.27 shows the simulated data for breakdown voltages with different $R_{\mathrm{B,ext}}$ base-biasing resistances. If the external biasing circuit is an ideal current source ($R_{\mathrm{B,ext}} \rightarrow \infty$ Ω), setting the DC base current to a fixed point, then the operational collector-emitter breakdown voltage is simply BV_{CEO} =1.8 V. If the external bias circuit presents is an ideal voltage

source ($R_{external}$ = 0 Ω), setting the DC base voltage, then $BV_{\mathrm{CER}} = BV_{\mathrm{CBO}}$ = 6.0 V as holes generated by the impact ionization process are free to ballast out of the HBT base terminal. The implemented external $R_{\mathrm{B,ext}}$ = 20 Ω in the W-band PA leads to operational breakdown voltages close to BV_{CER} = 4.2 V. (The dynamic RF load lines follow different curves than depicted on the DC-IV curves since the non-idealities in the curves at high currents and voltages are DC effects.) Measured hardware data from the foundry shows that the newer HiCUM HBT models still have difficultly predicting breakdown behavior at the device's voltage and current extremes [77], but the HiCUM model still correlates to measured hardware better than the older VBIC HBT models.

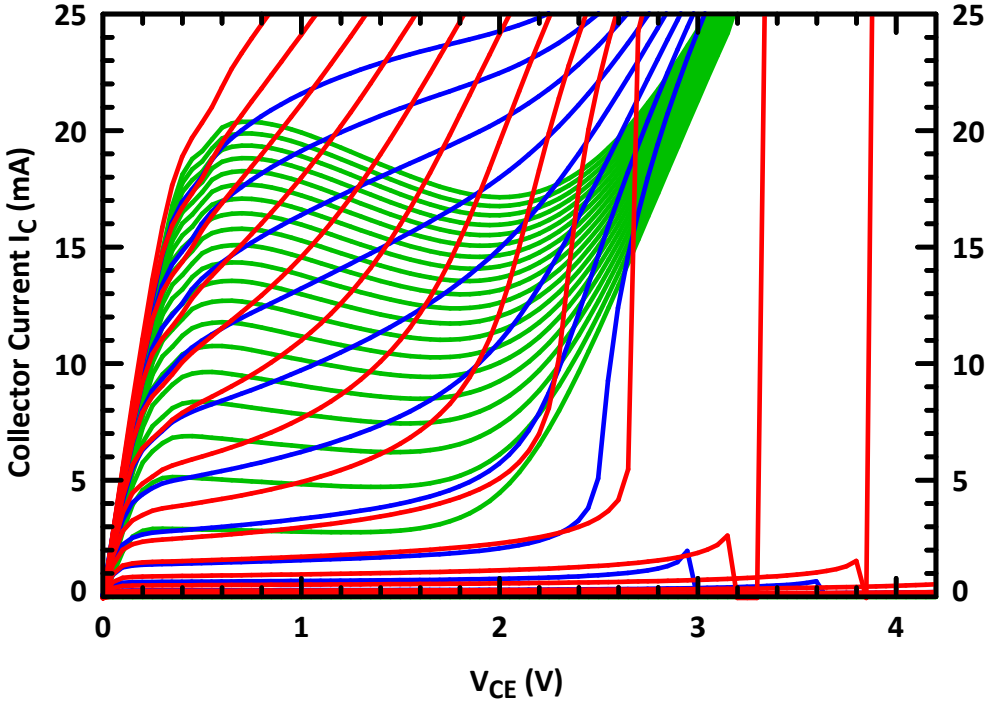


Figure 2.27: Simulated DC-IV curves (HiCUM model) for 5 μm HBT with DC external base resistance of 20 Ω in red, 200 Ω in blue, and ∞ Ω (open-base) green.

At high collector voltages, the electric fields cause impact ionization events in collector-base junction, causing electron-hole pairs to be generated. The holes drift back into the base region (from the electric field) and recombine with electrons injection from the emitter, leading to an overall decrease in forward base current. In the case of a high-impedance current source at the base terminal, the BC breakdown current feeds back into the BE junction, causes a rise in the BE voltage, which in turn injects more electrons from the emitter into the base & collector. At a critical base current, this impact-ionization process becomes a positive feedback mechanism and causes the collector current to rise exponentially.

For a constant V_{BE} biasing condition, the critical base current in Eq. (2.7) is the current at which the impact ionization causes a reversal in the DC base current from flowing *into* the base terminal to flowing *out* of the base terminal [78]. The base-current reversal has to potential to induce electrical-feedback instabilities within the HBT. Ballasting the base terminal via a low-resistance external base impedance delays the onset of negative effects arising from impact ionization.

$$I_B^0 = -\frac{V_T - r_e I_E^0}{r_{Bx} + r_{bi}(I_B^0)} \tag{2.7}$$

where I_B^0 is the critical base current, V_T is the thermal voltage ($\approx$ 26 mV at 25°C), I_E^0 is the corresponding emitter current at the onset of the critical base current, r_e is the small-signal emitter resistance, r_{bi} is the small-signal value of the internal base resistance (dependent on I_B & I_B^0), and r_{Bx} is the external base resistance.

Reliability and Safe Operation

Efforts are made to make the PA compliant to all foundry-recommended reliability guidelines for long-term, continual operation in commercial electronics at 100°C with > 100,000 power-on-hours (100K POH = 11.4 years) expected lifetimes. Metal lines, vias, MIM capacitors, TaN resistors, and HBTs are all chosen to operate within thermal, electromigration, and device-stress limitations for long-term reliability.

For HBT devices, the primary guidelines pertinent power amplifiers are (1) the DC forward-bias emitter current density should be < 15 mA/μm^2 or equivalently, < 9.0 mA/μm per emitter length; and (2) the collector-base DC voltage should be < 1.5 V to avoid reliability issues from impact ionization and avalanche multiplication effects. For the common-emitter PA, this limitation equates to an equivalent V_{CE} collector-emitter voltage of < 2.4 V since $V_{BE,\text{on}}$ = 0.87–0.9 V for the active device near peak-f_T/f_{MAX}/MAG DC biasing in common-emitter configuration.

Fig. 2.28 shows the conceptual DC operational limits of a general npn transistor [79]. DC and RF stresses outside HBT safe-operating areas result in performance degradation, circuit instabilities, or expected lifetime shortened, or even immediate catastrophic device failure.

Additional voltage and current headroom exists, beyond those set by the DC-IV curves, when the RF/AC dynamic load line is considered. The safe operating voltages have been measured to be $\sim$ 10% higher than the load-line condition for DC currents & voltages [80, 81]. The RMS (AC/RF) swings have less effective energy than static DC currents & voltages, and RF load lines may cross out of DC safe operating areas [82, 83]. The failure voltage for the RF dynamic swing at the collector is approximately 10% above $BV_{\text{CBO}} + V_{BE\text{on}}$ or

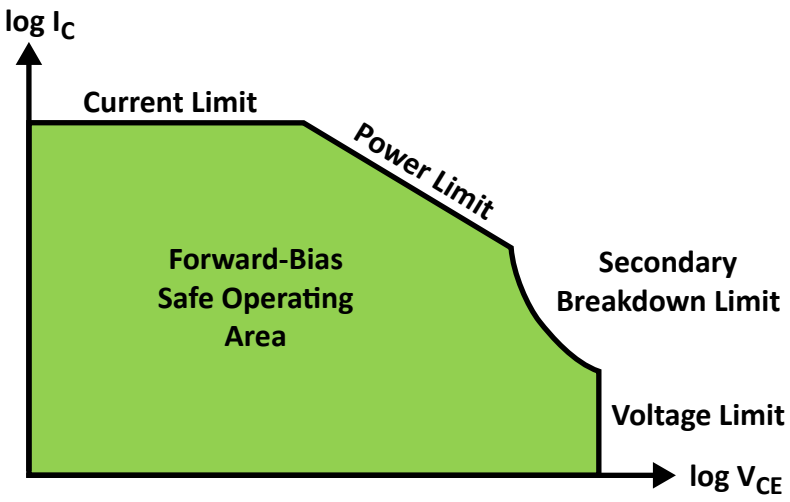


Figure 2.28: Forward-bias HBT safe operating area and general conceptualization of voltage, current, and power operating limits.

~7.6 V, which is well beyond peak voltages expected in the designed Class A/AB W-band PAs operating at a maximum DC $V_{CE} = 2.3$ V. The combined effects of both electrothermal and impact-ionization are examined in [84] and provide additional insight into the tradeoffs for stable operating regions.

The foundry estimates, under mixed-mode stresses (DC & large-signal RF), that base current is expected to increase up to a maximum of 25% over the lifetime, which results in an increased 5% collector-current difference. Shifts in DC β_F are observed, but RF/AC performance generally stays the same, even after extreme DC and RF stress into breakdown [85, 86]. For instance, measured f_T and f_{MAX} metrics remain the same after stresses.

To offset long-term stress effects, designs need be able to adjust the DC biasing to compensate for the forward-biasing degradation effects. Having proper provisions to re-tune DC bias currents aids in ensuring PA operation remains nominal as HBTs stress and age over the years.

Operating with $V_{CB} < 1.5$ V ($V_{CE} < 2.4$ V) leads to little β_F degradation: a 10% worst-case shift in the long term. The small shifts in DC characteristics caused by the mixed-mode RF & DC stresses can be compensated in DC bias adjustments. Once DC bias adjustments are made to account for any stress-related biasing shifts, the RF performance remains essentially the same. RF performance is not degraded with stresses as long as proper DC bias adjustments are made. More device degradation and base leakage current is observed in low-current operating regimes after large DC & RF stresses with DC β_F shifting as much as 20%, but this impact is minimal for PAs biased well above the base-emitter turn-on voltage

in typical high-current Class A and even reduced-current Class AB operation.

For reliability considerations in the W-band PA, there are two predominant operational limits for the SiGe HBT in common-emitter configuration: maximum safe current and maximum collector voltage. The maximum current specification is set by the electromigration of metals and contacts feeding the HBT, and the maximum voltage is set by the operational collector-emitter breakdown voltage BV_{CER}, as discussed previously in Section 2.5.3.

No significant aging has been observed during the measurement of the designed W-band PAs after over 500 hours of mixed testing on a single chip. More data is detailed in measurement reports in Section 2.11.

2.5.4 Aggregated Power Transistors

A key layout limitation involving deep-trench maximum local density was lifted after the Ka-band design work just before the W-band design phase. After major improvements of the STI (shallow-trench isolation) planarization process in the fab, the foundry improved and qualified the 8HP process to the point where the strict deep-trench (DT) local pattern density requirements could be lifted. The new process design rules allow the HBT unit cells to be densely packed together in 2D arrays with adjoining, shared deep-trench structures—butted DT structures. The prior DT local-density rules disallowed any 2D arrays of npn transistors, forcing custom large, aggregated power cells to split into different long 1D arrays separated by > 38 μm. The restrictions created significant area overhead, and total transistor footprints became large with significant parasitics to wire 1D arrays together. The additional inductances and capacitances of the wiring overhead were particularly problematic at W-band, robbing the transistors of precious gain at mmW frequencies. The new rules allow densely-packed HBTs and greatly reduce the inter-cell connection parasitics and preserve the low available transistor gains at W-band frequencies.

Aggregating and packing multiple 5 μm unit-cells into dense, overlapping-DT arrays has not been attempted in SiGe 8HP until this work. Prior to this W-band design, large transistors in W-band had to be broken into larger unit cells and separate 1D arrays had to be power combined at the collector outputs. Packing many cells tightly together reduces inter-cell wiring parasitics as well as the need for power-splitting and power-combining structures to synthesize large transistors.

Since large (> 18 μm) and multi-emitter finger HBT cells are disallowed in 8HP because of internal thermal runaway and current-hogging issues, larger area emitters must be created by aggregating multiple single-emitter-striped HBT cells. Following the general unit-cell design approach of the PA, the HBT power cells are made from aggregating identical HBT

(npn) unit cells with 5 μm emitter lengths. These same unit cells are also used in the active bias circuits, detailed more in Section 2.6.

The HBT unit cell is Q_0 = 5 μm emitter length and 0.12 μm in width. The aggregated power transistor in the first gain stage (Q1) uses the unit cell in a 1 × 2 array: Q1 = 2 × Q_0 = 10 μm emitter length, Fig. 2.29a. The power transistor in the second gain stage uses the unit cell in a 1 × 4 array: Q2 = 4 × Q_0 = 20 μm emitter length, Fig. 2.29b. The third and final gain stage has a power transistor aggregated from the unit cell HBT in a 2×4 array: Q3 = 8 × Q_0 = 40 μm, Fig. 2.29c. The aggregated Q3 output transistor measures 15.1 μm × 31.5 μm.

For each gain stage, a unit cell element Q_0 is also used as an active-bias reference transistor, Q_{B1} = Q_{B2} = Q_{B3} = 1 × Q_0 = 5 μm. (Q_0 is also the same 5 μm unit-cell transistor used and studied throughout the two earlier Ka-band PA design iterations.)

A dense lattice of DT, made of 1 μm-wide and 6 μm-deep oxide & polysilicon trenches, isolates each unit cell and the aggregated array both electrically and thermally. Finally, the entire aggregated power-transistor structure is surrounding by a large moat of connected substrate contacts for proper substrate grounding and device isolation.

The electrical effects of shared DT structures and a shared perimeter substrate-contact ring is not modeled for the aggregated power-cell HBT. The HBT model assumes that a substrate ring surrounds the npn cell. When transistors are aggregated and abutted, the default substrate modeling does not reflect the physical layout, and some minor modeling errors are expected. from aggregating the unit-cells densely, but the net electrical effects are expected to be small relative to the process variation. The unmodeled thermal effects of dense packing at shared DT, however, are not insignificant and must be addressed in a custom thermal model.

Additional isolation-moat structures are placed around the aggregated HBT power-cells. These reduce the substrate coupling between HBT gain stages and are formed using regions of high-resistance substrate (STI and DT) surrounded by shunted low-resistance substrate-contact rings that shunt noise to the large global RF & DC ground planes. The isolation-moat structures provide high-resistance paths between critical HBT elements and low-resistance paths to substrate RF & DC grounding.

2.5.5 Q3 Power Transistor Cell

The output power-transistor layout cell is a key element in the PA design. It delivers the necessary power and has the largest RF voltage and current swing of all the PA stages. Unlike other III-V foundry processes that are tailored to power amplifiers, the GF 8HP SiGe

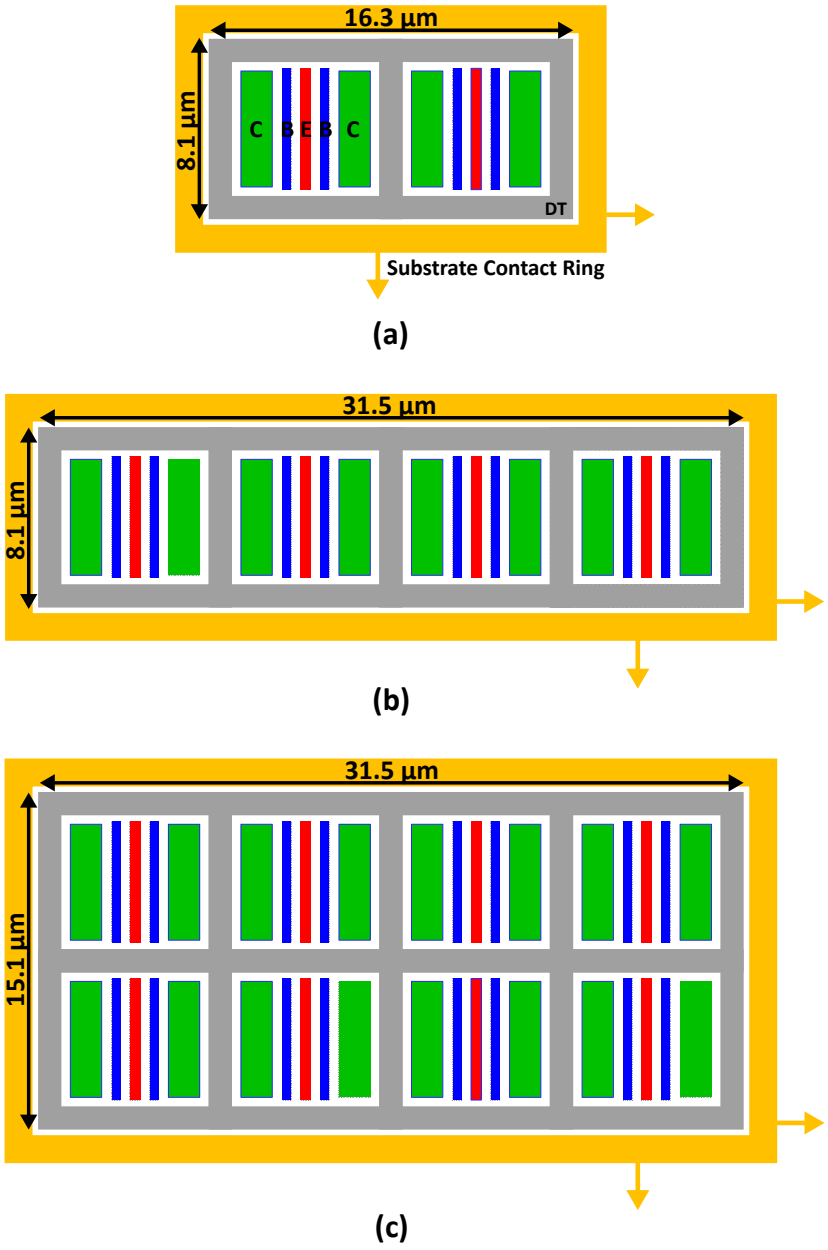


Figure 2.29: Implemented W-band PA layouts of densely-packed 5 µm unit cells aggregated into power transistors with overlapping deep-trench isolation (DT) areas (a) Q1 (10 µm), (b) Q2 (20 µm), and (c) Q3 (40 µm). The substrate ring (in yellow) surrounds the aggregated transistor cells in a large, low-resistance, grounded ring. The outer boundary of the substrate ring is not shown. Inter-cell wiring of the unit cells (dual-base stripe, dual-collector stripe, single-emitter stripe) is not shown.

foundry process does not offer special pre-designed power-transistor cells nor does it allow custom emitter layouts beyond the single-striped npn device. The task of creating large power cells is left to the foundry user, and large devices beyond 18 μm × 0.12 μm must be created by placing smaller standard cells in parallel. These custom cells must then have their wiring parasitics extracted separately, and mutual-heating effects between cells is not modeled.

Electro-thermal coupling within a custom power cell is not easily modeled, especially for dense power-cell layouts, and this must be taken as a known model limitation during simulations. To obtain the higher output powers, larger device sizes are desirable, but as discussed in Section 2.5.2, larger devices result in lower RF gain at 90 GHz which in turn also reduces PAE and the PA's usefulness in system applications like multi-element SiGe transmit phased arrays. For the PA to be easily integrated as a element such integrated phased arrays, target PA performance is set to gain $>$ 10 dB and PAE $>$ 10% while considering attainable matching network values using compact matching networks. After simulating the power, gain, PAE, and ease of proper matching network design tradeoffs, the 40 μm emitter length is chosen for the 90 GHz output transistor Q3.

The transistor size of the Q3 output HBT, aggregated from 8 unit-cell transistors, is 40 μm × 0.12 μm. The Q3 HBT layout and associated terminal wiring is illustrated in Fig. 2.30. Power transistors Q1 and Q2 are wired in a similar configuration with emitter grounding on the top-level AM metal pulled over the entire power-cell HBT structure. The HBT emitter is wired "top-down" from the AM top metal all the way down to M1 and the emitter contacts within the substrate.

The collector emitter fingers are pulled out from a stack of lower M1–M4 copper metals underneath the AM emitter for an extremely short 2.5 μm run before joining together and contacting back up a metal stack to the low-loss, AM top-level metal interconnects and transmission lines (illustrated in blue). Only metals up to M4 are used above the active collector, rather than rather than higher metals closer to the AM emitter ground, to reduce collector-to-emitter capacitances. The base fingers are also pulled away from the emitter grounding bar, but connect up only to LY (2nd aluminum layer from the top) in order to connect directly to the bottom plate of the series MIM capacitor, which is both a reactive matching element as well as a necessary DC block.

For additional reliability and lower parasitics inductances, the emitter is wired with a metal stack beyond the minimum required for DC and RMS high-current electromigration compliance. The DC and RF grounds around the HBT transistors are stacked wide metals on AM, LY, and MQ, providing a very low-impedance ground termination.

The wide emitter ground bus, pulled over the entire power-cell structure, also acts as a heat spreader to evenly distributed the dissipated heat and as a heat sink to wick away heat

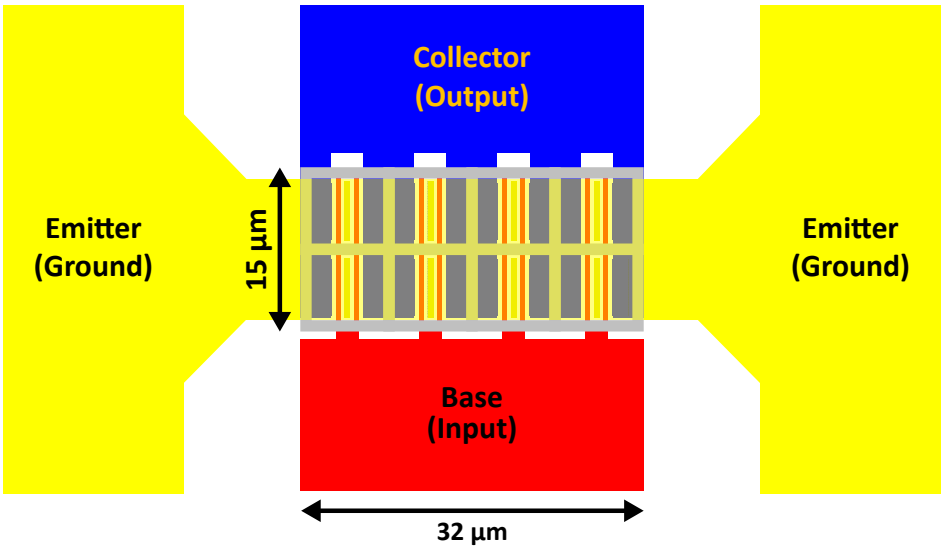


Figure 2.30: Q3 = 40 µm power transistor cell layout made from 2 × 4× 5 µm units cells. The emitter ground is pulled across the entire power transistor and contacts the 8 emitter from above— top-level AM metal all the way down to the substrate contacts. The collector and base fingers are connected from the north and south, respectively, and connect back up to higher metals once clear of the emitter grounding bar.

into the wide, stacked (thick) global ground planes surrounding the HBT.

More small-signal and large-signal aspects of the aggregated power transistors Q1, Q2, and Q3 are covered later in detail in Section 2.9.3.

2.5.6 Custom Thermal Modeling

In purely Class A and mild Class AB (closer to Class A than Class B) amplifiers with high standing DC quiescent currents, significant heat is dissipated in the power transistors, even if there is no RF input drive present. Thermal effects from high power dissipations (either quiescent or under full RF drive) in multi-fingered transistor designs can cause stability issues in HBTs. Compared to traditional bipolar BJTs, however, SiGe HBT β_F DC current gains have less positive temperature dependence than their silicon homojunction counterparts. Compared to GaAs, silicon-based substrates have 3× the thermal conductivity, aiding the thermal stability of HBT devices. Compared to CMOS FETs, which have stabilizing negative temperature coefficients with respect to current draw, SiGe HBTs exhibit positive temperatures with current– more current causes HBT emitter fingers to heat up, and more heat in the emitter causes the emitter current to rise, a positive feedback process.

This section examines the circuit-design impacts of thermal conditions arising from the

tightly-packed unit cells aggregated together as custom power-transistor cells. Custom mutual thermal-coupling models are extracted from measured data, and the thermal stability of the aggregated HBTs are characterized.

The foundry lifted process design rules for the maximum allowable deep-trench local density, allowing the 5 μm transistors to be packed with shared DT structures. Densely-packed unit cells in the HBT power transistor reduce, significantly, the intra-cell wiring parasitics and help to preserve the available gain in the transistors. However, the maximally-compact, minimum-spacing layout leads to increased mutual thermal coupling and HBT heating. These electro-thermal effects cause shifts in DC & RF characteristics that need to be modeled for thermally-stable, optimal designs at W-band. Efforts are made to use prior Ka-band PA measured data to model the mutual thermal coupling of the 2D array of 5 μm unit cells in order to adjust the PA design accordingly.

The foundry-provided HBT models include internal self-heating thermal effects for single HBTs, but no electro-thermal models or coupling coefficients are available for different transistor-packing arrangements of multiple HBTs. Designers must model any additional mutual thermal effects through the thermal impedance multiplier, m_{th}. Proper estimation of the thermal effects of densely-packed transistor cells helps to predict, more accurately, proper DC operating points, optimal RF matching values, and safe thermal operating conditions (emitter ballasting requirements).

Fig. 2.33 shows the effects of HBT self heating on the DC-IV curves. Without self-heating modeled, the collector currents, for a given base-current bias, do not drop with increasing collector voltages. A unit-cell within the aggregated power-transistor array experiences mutual-heating effects from adjacent transistors that have a similar effect as increased self-heating on the DC-IV behavior.

Thermal Resistance R_{th}

The source of the heat generation within an HBT is located just below the emitter finger stripe. The temperature of the internal emitter junction can be described by a thermal resistance and the DC power dissipation [87]:

$$T_{junction} = T_{\text{ambient}} + R_{th}P_{DC} \tag{2.8}$$

where P_{DC} is the DC power dissipation within the HBT, R_{th} is the thermal resistance of the device (in K/W), and T_{ambient} is the ambient chip temperature (in K). The HBT DC power dissipation is $P_{DC} = V_{CE} \times I_C$.

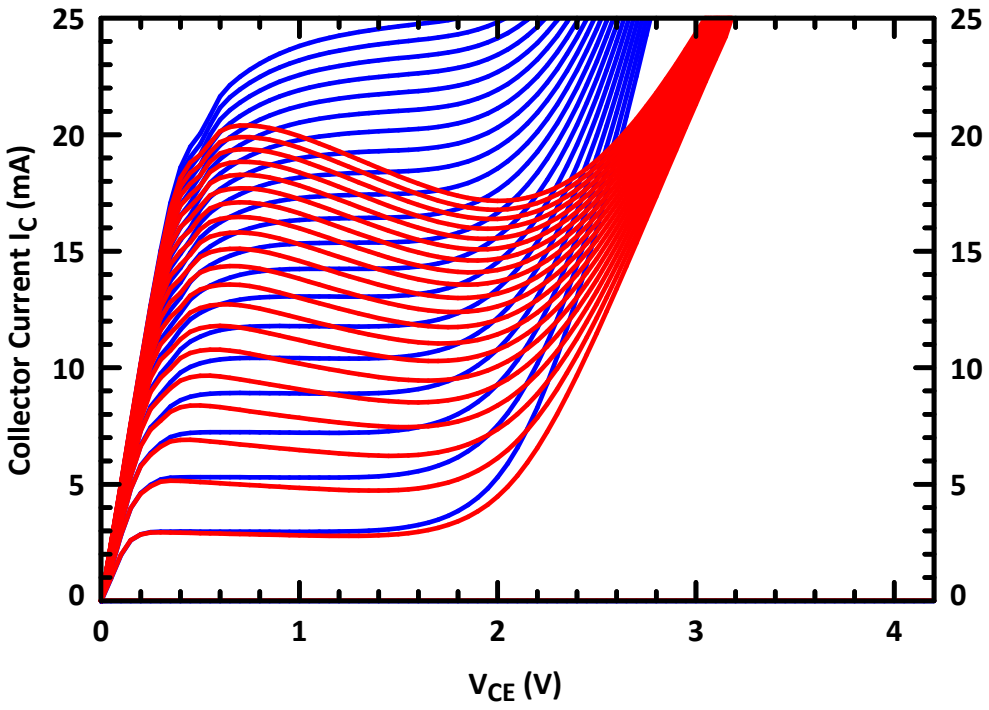


Figure 2.31: DC-IV curves for 5 μm npn HBT with self heating (red) and without self heating (blue). The HBT is biased with an ideal current source at the base in common-emitter configuration at 25°C.

The rise in device temperature due to HBT self heating is then described by Eq. (2.9):

$$\Delta T = R_{th} P_{DC} \tag{2.9}$$

The characteristic HBT thermal resistance is a function of the transistor device size and effective thermal conductivity. R_{th} can be estimated from the classic Eq. (2.10) [88]:

$$R_{th} \ \propto \ \frac{1}{4\kappa\sqrt{LW}} \tag{2.10}$$

where κ is the thermal conductivity constant in W/(m·K), L is the emitter length, and W is the emitter width.

A more recent, more accurate estimation in [89] of device thermal resistances in modern HBTs describes R_{th} as a series combination of thin slabs along the direction of heat propagation z with thermal resistances κ_i, areas A_i, and thicknesses d_i:

$$R_{th} = \sum_i R_{th,i} = \sum_i \frac{d_i}{\kappa_i A_i} \approx \int \frac{1}{\kappa(z) A(z)} dz \tag{2.11}$$

The general expression, when applied to substrates that are much thicker than the HBT finger width and length, simplified into Eq. (2.12):

$$R_{th} = \frac{1}{2\kappa(L-W)} \ln\left[\frac{(W+2t_{\text{sub}})}{(L+2t_{\text{sub}})}\frac{L}{W}\right] \cong \frac{1}{2\kappa(L-W)} \ln\left(\frac{L}{W}\right) \tag{2.12}$$

where t_{sub} is the substrate thickness, assumed to be much larger than the emitter L or W.

The deep-trench (DT) ring around the HBT isolates, both thermally and electrically, each device from nearby adjacent ones [90]. The thermal conductivity of the 1 μm-wide, 6 μm-deep DT oxide isolating ring around each 5μm unit cell is 0.014 W/(cm-K), >100 × lower than that of the silicon substrate at 1.56 W/(cm-K) [68]. (For reference, the deepest HBT conductive layer, the subcollector, is only 2.8 μm deep into the substrate.) The DT low thermal conductivity increases the HBT self-heating, but also restricts the lateral thermal dissipation (heat flux), reducing the mutual thermal coupling between adjacent transistor cells.

The work in [89] applies Eq 2.12 for modeling cases where deep-trench oxide walls confine the heat flux, raising the effective thermal impedance of an HBT.

In the presence of DT isolation structures, the thermal resistance of the HBTs increase with decreasing emitter areas– R_{th} increases for smaller emitter lengths in a given technology [91]. For the same current density, however, smaller HBTs, despite their higher R_{th}, experience less temperature rise from self-heating than larger transistors because the ratio of DT-enclosed area to emitter area is larger, thus providing more cross-sectional heat dissipation per emitter

area and better overall thermal stability. Mutual thermal coupling, as studied in [92], is also decreased when densely-packed arrays use smaller unit HBTs.

The nominal thermal resistance for the standalone 5 μm C–B–E–B–C unit-cell HBT, including effects of deep-trench enclosure, in 8HP technology is $R_{th} = 2590$ K/W at 25°C[10]. Heat is also dissipated through the large areas of thermally-conductive metal wiring and contacts that sit above the HBT.

Effects of Mutual Thermal Coupling on R_{th}

The thermal resistance R_{th} models an HBT's temperature rise due to self-heating effects, but mutual thermal coupling from neighboring cells must be included in order to predict proper DC operating points and to ensure that the arrayed power-cell is thermally stable— that there are no electro-thermal positive-feedback run-away conditions. To account for mutual heating effects from neighboring transistors within a aggregated array, the effective thermal resistance $R_{th,\text{eff}}$ must be adjusted for the packing density and arrangement for each power-cell layout.

Since thermal coupling from neighboring cells result in additional temperature rise (over self-heating) within a unit-cell transistor, the total rise in temperature can be modeled as an effective increase in the HBT thermal resistance R_{th}. Thermal simulations of SiGe HBTs in [90, 92] show the effects of mutual thermal coupling on the effective thermal resistance of transistors packed together.

A mutual coupling factor between transistors, m_{th}, is necessary to described the effective increase in thermal resistance for each of the aggregated power cells Q1, Q2, and Q3. The effective thermal resistance caused by increased mutual thermal coupling is then given by 2.13 and the emitter junction temperature rise over that of self-heating alone is given by 2.14.

$$R_{th,\text{eff}} = \ m_{th} R_{th0} \tag{2.13}$$

where R_{th0} is the thermal resistance of a unit-cell HBT with only self-heating effects— no mutual thermal coupling.

$$\Delta\ T = P_{DC} R_{th0}\,(m_{th} - 1) \tag{2.14}$$

[10]Densely-stacked emitter wiring, from copper all the way to the aluminum top metal acts, as a heat sink from the top side of the HBT and can result in small changes in the effective thermal resistance, depending on layout specifics.

Thermal Resistance Modeling

Custom thermal models are made for each W-band aggregated power-cell array (Q1, Q2, and Q3) using measured data collected from prior Ka-band PA designs in Chapter 3 . Since the W-band HBTs employ packing configurations that were not supported in earlier Ka-band design, initial W-band thermal-coupling estimates are used based on the measured data and simulated scaling factors from Rieh et al. [92]. After the W-band measurements, the estimated thermal models are further refined to reflect the latest measurement data and presented here.

Differences in m_{th} manifest themselves as differences in quiescent currents within the power transistors. m_{th} accounts for the majority of operating differences between the reference unit-cell transistor in the bias circuit and the aggregated power transistor. The active-bias reference transistor, operating at the same peak-f_T current density as the power-cell HBTs, is physically separated from the power cells by at least 300 μm, and the difference in local transistor temperatures leads to a fixed shift in mirrored collector current, which can be measured and characterized through a mutual thermal coupling coefficient.

Although HBT β_F can vary by ±20% even in adjacent transistors, the active-biasing circuit compensates for β_F differences, leaving mutual thermal heating, which is unmodeled by the HBT model, the dominant factor in DC current shifts. The unmodeled current dependence on mutual thermal heating can be used to extract first-order estimates of m_{th}. Since each gain stage of the PA has its own active biasing circuits and voltage-reference pins, the current dependence shift from m_{th} can be extracted in-situ— within the PA directly. More advanced methods of extracting HBT thermal parameters exist [93–95], but for this work, the necessary simplified models can be created simply by using the large data set of DC current and voltage for each gain stage and each active-biasing circuit.

The differences in expected DC collector currents with no mutual thermal coupling effects that of measured collector currents with mutual thermal coupling can be used to extract the thermal resistance multiplier m_{th} to model mutual thermal coupling. The increase of total collector current from mutual thermal coupling is primarily attributed to an effective increase of $R_{th,\text{eff}}$, the effective thermal resistance

Mutual-thermal coupling in the power transistors makes $m_{th} > 1.0$ and causes the temperature of the power transistor to rise accordingly. In contrast, the 5 μm bias-reference transistor does not experience the any rise in temperature from adjacent unit cells— only self-heating effects. The differential increase in temperature between the power transistor and the biasing-reference HBT unit cell causes the power HBT to have more quiescent current than expected under no mutual thermal coupling (self-heating only) $m_{th} = 1.00$. Adjusting the custom HBT model m_{th} to match currents observed in the power HBT cells

under active-bias mirroring gives a useful estimate of the mutual thermal effects of each array packing configuration: 1 × 2, 1 × 4, and 2 × 4 arrays.

$R_{th,\text{eff}}$ and m_{th} are estimated by matching DC quiescent current values measured in each PA stage. Multiple dies are measured to average out readings of increased collector currents from mutual thermal coupling. A range of biasing currents and voltages are used to capture multiple data points on each Q1, Q2, and Q3 aggregated power cell.

Measurements for thermal models are done at room temperature and at a lower collector voltage $V_{CC} = 1.3$ V in order to eliminate any high-voltage impact-ionization effects on measured collector DC currents. This separates the R_{th} modeling effects from the impact ionization modeling effects, which are known to have model-to-hardware correlation issues at high voltages. Operating at lower collector voltages allows capturing an effective thermal resistance that is primarily influenced by the mutual thermal coupling of the aggregated unit cells.

$R_{th,\text{eff}}$ values in back-simulations are increased until the simulated power-transistor collector currents match that of the observed measurements. m_{th} is simply the effective thermal resistance multiplier or scaling factor, given by Eq. (2.15).

$$m_{th} = \frac{R_{th,\text{eff}}}{R_{th0}} \tag{2.15}$$

Although HBT β_F can vary by ±20% even for adjacent transistors, the effects on the power-cell collector currents are also scaled down by $\beta_F = 170$ because of the active-bias tracking. Small skews and variations in Early voltages are also scaled by a factor of ∼5. Measurements of several separate PA chips (dies) averages out transistor variations further, leaving only the fixed, unmodeled mutual thermal coupling as the predominant factor altering expected collector currents in the aggregated power HBTs. This results in a first-order model of the mutual thermal effects. In the future, to obtain even more accurate results, more parts from multiple wafer lots should be measured to refine the custom thermal modeling even further.

Fig. 2.32a and Fig. 2.32b illustrate the implemented Q2 20 µm aggregated power-cell HBT as well as the Q3 40 µm HBT, respectively. In the 20 µm device, only 2 of the 4 unit cells share 2 edges with adjacent HBTs, alleviating the overall effective mutual thermal coupling compared to devices packed in long 1D arrays in Fig. 2.32c. The 2D-packed 40 µm HBT device has no unit cell completely surrounded by adjacent HBTs on all edges and corners, which reduces the mutual thermal coupling of unit cells compared to those modeled in Fig. 2.32d. Avoiding maximum long 1D and 2D arrays decreases the effective mutual thermal coupling m_{th} compared to those modeled in [92].

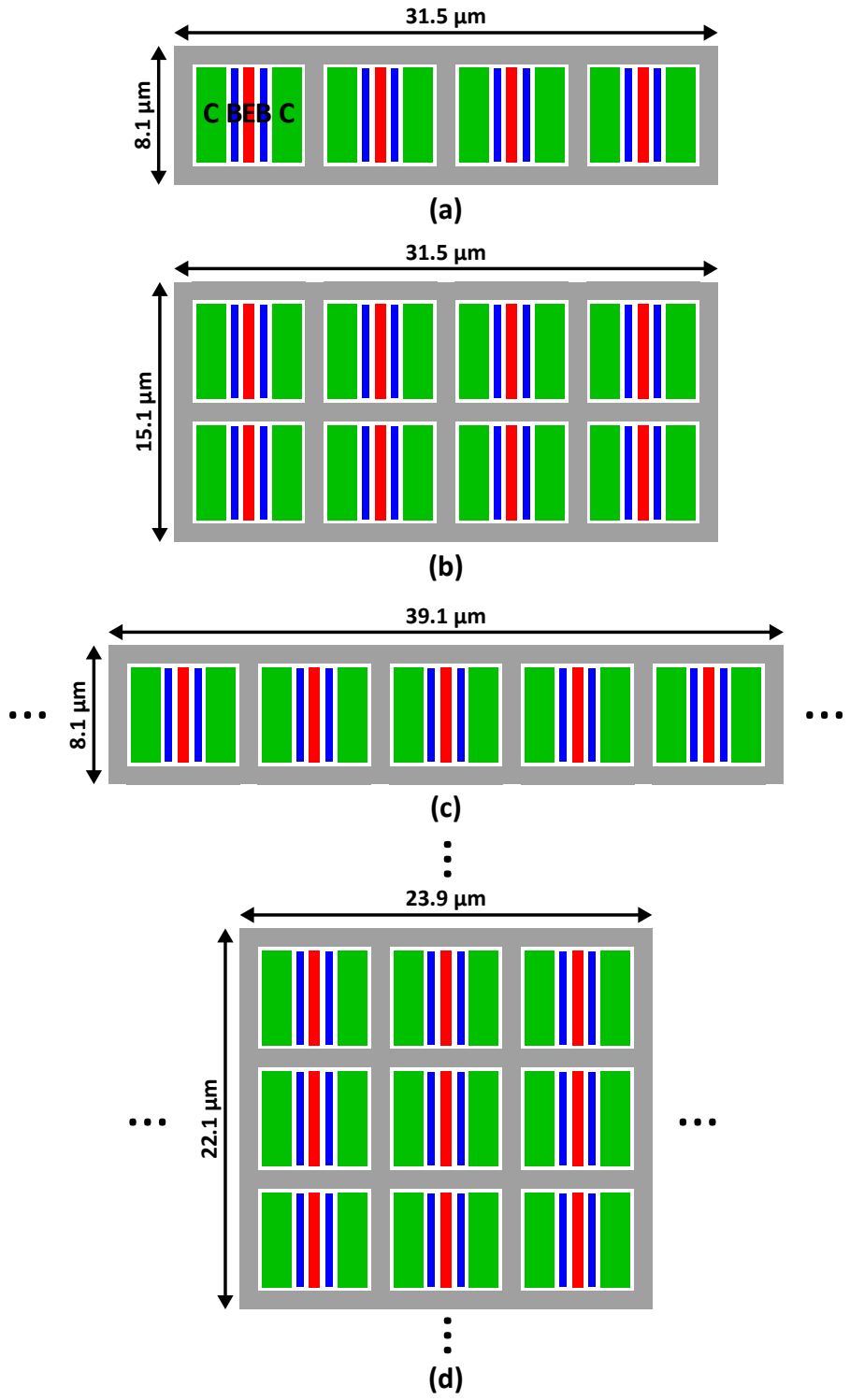


Figure 2.32: (a) Implemented 20 µm HBT layout (b) Implemented 40 µm HBT layout (c) Long 1D array of HBTs with the center transistor sharing 2 edges (d) Long 2D array of HBTs with the center transistor completely surrounded by adjacent, overlapping HBTs on all edges and corners.

Table 2.1: Extracted thermal impedance multipliers and thermal resistances for the 5 μm unit cells that make up the power transistors.

Parameter	**Value**
$m_{th,\mathrm{Q_0}}$	1.00
$m_{th,\mathrm{Q1}}$	1.10
$m_{th,\mathrm{Q2}}$	1.14
$m_{th,\mathrm{Q3}}$	1.34
$R_{th,5\mu m,\mathrm{Q_0}}$	2590 K/W
$R_{th,5\mu m,\mathrm{Q1}}$	2850 K/W
$R_{th,5\mu m,\mathrm{Q2}}$	2950 K/W
$R_{th,5\mu m,\mathrm{Q3}}$	3470 K/W
$R_{th,10\mu m,\mathrm{Q1}}$	1425 K/W
$R_{th,20\mu m,\mathrm{Q2}}$	738 K/W
$R_{th,40\mu m,\mathrm{Q3}}$	434 K/W

Extracted Thermal Models

The results of the custom thermal model extractions are presented here. Using DC measurement data from Ka-band and W-band PA circuits across different die samples, the effective thermal resistance of densely-packed 1D and 2D HBT structures is characterized and used in the modeling of the W-band transistor arrays.

Table 2.1 gives the extracted thermal multiplier correction factors for a 5 μm unit cell to account for mutual thermal coupling effects in the different implemented arrays. To obtain the equivalent thermal resistance of the aggregated transistor, as a whole, the $R_{th,\mathrm{eff}}$ must be divided by the number of unit cells within the array. So, while each of the eight 5 μm transistor within the Q3 array is modeled with $R_{th,\mathrm{Q3}} = 3470$ K/W, the aggregated Q3, as a whole, has $R_{th} = 3470/8$ K/W $= 434$ K/W.

$R_{th,\mathrm{Q_0}} = 2590$ K/W and $m_{th,\mathrm{Q_0}} = 1.00$ for a 5 μm unit-cell device modeled with only self-heating effects— no mutual thermal coupling. Transistors in the active bias circuits fall under this thermal model. For the first gain stage Q1 with two 5 μm unit cells joined side-by-side, each of the two HBT unit cells has a characteristic $m_{th} = 1.10$ of mutual thermal coupling, which results in the aggregate power cell with $R_{th,10\mu m,\mathrm{Q1}} = 1425$ K/W. For each of the four unit cells in the Q2 small 1D array, $m_{th} = 1.14$, and the aggregate power cell has an

Table 2.2: Effective junction temperature rise from mutual thermal coupling effects (over that of self-heating alone) for V_{CC} = 1.7–2.3 V at peak-f_T biasing current density (1.5 mA/μm).

Parameter	**Value**
$\Delta T_{j,\mathrm{Q1}}$	3.3–4.5°C
$\Delta T_{j,\mathrm{Q2}}$	4.6–6.3°C
$\Delta T_{j,\mathrm{Q3}}$	11.2–15.2°C

equivalent $R_{th,20\mu m,\mathrm{Q3}}$ = 738 K/W.

Finally, for each of the eight unit cells in the Q2 2D array, m_{th} = 1.34, and the aggregate power cell has an equivalent $R_{th,40\mu m,\mathrm{Q3}}$ = 434 K/W. The implemented 2D HBT packing in the Q3 40 μm power transistor avoids completely surrounding and enclosing any individual HBT cell with other adjacent cells. In other words, each cell in the power transistor has at least one un-shared edge. This arrangement reduces the mutual thermal coupling and thus temperature rise since all transistors have at least one un-shared edge. A completely surrounded unit cell may have a thermal multiplier closer to m_{th} 2.0, which would require additional emitter ballasting to stabilize thermally.

The resulting $R_{th,\mathrm{eff}}$ captures the aggregated average of the power-cell transistor. Outer unit cells will have less mutual thermal coupling than the inner cells, for example. For the 5 μm unit-cell transistors in the Q3 output power transistor near peak f_T biasing and 1.7–2.3 V collector voltage, this results in an average increase of about 11.2–15.2°C over nominal, non-coupled emitter-junction temperatures.

The equivalent average temperature rise in each unit-cell within an aggregated power transistor is given in Table 2.2.

To make simulations faster with more consistent convergence, each 5 μm unit-cell HBT within the same power-transistor array has the same m_{th} applied, even though HBTs at the end of the array have slightly different thermal profiles than the center ones. Applying multiplicity to the power-cell modeling allows considerable simplification of the simulations without appreciable errors.

The effects of increased thermal resistances are illustrated in Fig. 2.33. Maximum available gain (MAG) is plotted for the 40 μm aggregated power transistor across different mutual-thermal coupling coefficients m_{th} for the constituent unit-cell 5 μm HBTs. The solid lines represent expected MAG in with no layout parasitics.

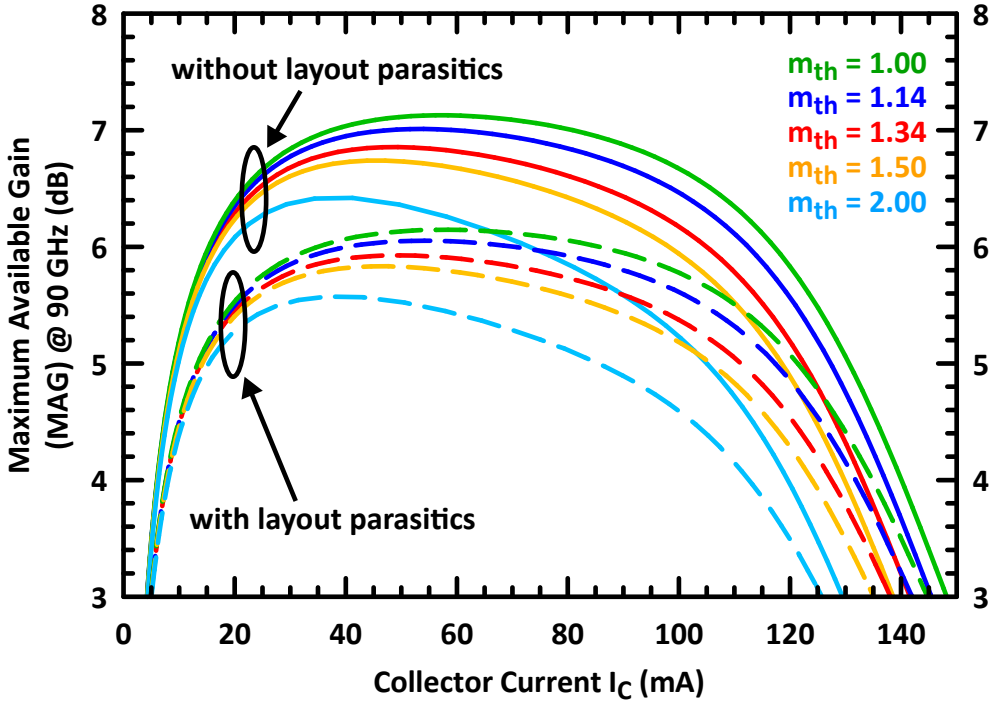


Figure 2.33: Maximum available gain (MAG) at 90 GHz for Q3 40μm HBT with mutual thermal coupling coefficient m_{th} from 1.00 to 2.00 with layout parasitics (dashed) and without layout parasitics (solid) for the constituent unit-cell 5 μm HBTs. V_{CC} = 1.7 V at room temperature, 25°C. External DC base-biasing resistance is 20 Ω (low).

In terms of gain at 90 GHz, the effects of layout parasitics dominate over thermal effects for 2D-packed 40 μm HBT, and a more compact power-transistor layout provides more improvement in gain than decreasing the thermal coupling between adjacent unit-cells elements of the power-transistor array. A more compact power-cell HBT layout will benefit the PA at 90 GHz as long as the resulting power-HBT is thermally stable. For much larger transistors that require more densely-packed 2D arrays, thermal stability may not be guaranteed without emitter ballasting resistor, and there is a trade-off between aggregated transistor size, thermal stability, and gain, and layout parasitics.

R_{th} increase from 1.0 to 1.22 only impacts the maximum available gain (MAG) by 0.3 dB at 90 GHz at a fixed collector current while wiring parasitics alone, assuming no thermal coupling, can impact MAG more than 1 dB. The penalty of decreased MAG under higher mutual thermal coupling is less than that of additional parasitics wiring larger transistor cells or transistor-level power splitting and combining.

Thermal Stability

Thermal stability in the HBTs must also be examined to ensure that an aggregated power device does not destroy itself in a thermal runaway condition. Variations in the current and voltage distribution of the emitter fingers can lead to a positive feedback process where the center emitter fingers are hotter, thermally, than adjacent finger and begin to steal current from its neighbors: current hogging. A critical point in this process results in all of the intended current going to the hottest emitter finger, destroying the device after internal silicon junction temperatures reach a failure point.

Stability in heterojunction HBTs is as not sensitive to thermal effects as homojunction BJTs since the DC forward base-to-collector current gain β_F drops slowly with increasing temperature, rather than increase rapidly with temperature [96]. This makes current-hogging in hotter transistor cells less of a stability issue, but thermal runaway can still exist under high-power dissipation conditions. Thermal conditions must be analyzed carefully to determine whether additional negative feedback measures are necessary for thermal stability.

Emitter ballasting is typically required within very large multi-emitter-fingered or very long emitter power devices [97, 98], but ballast-free designs have been proven thermally stable [99] under the right conditions. The unit-cell approach alleviates the ballasting requirements because of the electrothermal isolation between adjacent cells. The emitter area is split, effectively, among multiple cells with DT thermal isolation.

For a quick estimate of the required emitter ballasting resistance, equations can be derived from ideal bipolar junction equations. A simplified estimate of the required ballast resistance for guaranteed thermal stability in silicon bipolar homojunction transistors (BJTs) is given

in Eq. (2.16).

$$R_e \geq \frac{kT}{qI_c}(\alpha R_{th} V_{CC} - 1) \tag{2.16}$$

where $\alpha \approx 0.05 \times \mathrm{I}_C$ mA/°C for ideal silicon bipolar transistors [100].

For a standalone 5 μm unit cell with $R_{th} = 2590$ K/W (no mutual thermal coupling) operating at peak-f_T/f_{MAX}/MAG current of 7.5 mA at $V_{CC} = 1.7$ V, the estimate results in $R_{E,\mathrm{ballast}} \geq 2.2\ \Omega$. For the 5 μm npn unit-cell, the extrinsic emitter resistance $R_E = 2\ \Omega$, and the extrinsic base resistance $R_B = 46\ \Omega$, not accounting for additional wiring parasitics to the HBT cell. An additional external emitter wiring resistance of $> 0.6\ \Omega$ results in a total of $> 2.6\ \Omega$ built-in emitter ballasting from contact and wiring resistances per unit cell. Both emitter and base resistances, along with additional layout parasitics, provide localized voltage feedback around the HBT and help to stabilize the thermal feedback effects[11]

The simplified estimates of emitter ballasting requirements in Eq. (2.16) only apply directly to ideal, traditional silicon homojunction bipolar transistors with no mutual thermal coupling. Determining more accurate thermally-stable operating regions for modern HBTs requires circuit simulation with the extracted mutual thermal-coupling factors.

Simulations with additional thermal effects determine that $m_{th} > 1.7$ is amount of effective mutual thermal coupling required such that the built-in emitter ballasting within the 5 μm unit cells and local wiring can no longer guarantee thermal stability under nominal peak-f_T biasing conditions. Varying DC conditions are simulated to look for I_C vs V_{BE} flyback points[12] to determine points of potential thermal feedback instability.

At the extracted $m_{th} = 1.34$ for the Q3 output HBT, no thermal stability issues are expected, and none have been observed in any subsequent measurements of the W-band PAs. Omitting the requirement for individual emitter ballasting resistors preserves the precious gain at 90 GHz and greatly simplified the aggregated power-transistor layout.

2.6 Active Bias Circuits

As power amplifiers convert large amounts of DC power to RF power, the design of DC biasing networks is equally important as the RF design in mmW PAs. In particular, pushing the HBT to its voltage and power limits requires aggressive biasing schemes that are also adjustable in the lab. The ability to tune, precisely, DC currents and voltages in

[11] The effect of base ballasting resistance is much less than resistive emitter ballasting by a factor of approximately β_F.

[12] Flyback is the point where the I_C versus V_{BE}, at a given I_B biasing, bends back on itself, indicating a critical collector current where the device is potentially thermally unstable.

each gain stage enables more flexible optimization of the PA on the as collector voltages are pushed higher in measurements. This section covers the details of the active-bias circuits and networks that enable operational collector voltages $V_{CC} > 1.7$ V in the power HBT as well as stable PA operation across frequency and increasing collector voltages.

On-chip active-bias circuits allow power transistors to better track the reference bias circuits over process, voltage, and temperature (PVT). Bias circuits set and maintain a well-regulated DC operating point of the power transistors and supply necessary, stable DC base currents to the power transistors, especially under high RF drives. They must be designed to limit interference with RF signals and to be stable even outside the operating frequency of the amplifier.

Active biasing of the HBT power transistors refers to the use of active-transistor current mirroring (with localized feedback in bias circuits) and not simply purely resistive networks to set the DC operating currents & voltages. Integration of active on-chip bias is imperative for practical PA integration into a larger system, and biasing currents can be controlled precisely, even with expected transistor β_F[13] variations across the chip.

In this W-band PA design, the integrated-active bias circuits allow very low external base resistances to be presented to the power HBTs while maintaining fine current-tuning controls in the bias circuits. The power HBT's low external base resistance results in much higher BV_{CER} operational collector breakdown voltages over $BV_{\mathrm{CEO}} = 1.8\ V$, which allows more power to be delivered to the PA output, discussed more in Section 2.6.2. The active bias circuit, in conjunction with novel inverted-microstrip distributed transmission-line base-feeds, allows de-coupling of the DC base external resistance from the RF impedance presented to the power-transistor base. The base bias-insertion network allows extremely low DC resistance to be presented (20 Ω) while the RF impedance remains high (> 120 Ω) relative to the RF base input impedances. High standing currents in the active-bias reference HBTs and low HBT current-mirroring ratios allow for fine tuning of the power-transistor HBT currents as well as extremely stable bias references when the power cells are in compression (high-current operation).

In the prior Ka-band designs in this dissertation, 40 Ω external base resistance have been presented to the power transistors' base terminals. The low external base resistance of the bias circuit allows for higher collector-emitter breakdown voltages. Breakdown increases from the collector-emitter (with open base) $BV_{\mathrm{CEO}} = 1.8$ V closer toward the maximum collector-base (with open emitter) voltage $BV_{\mathrm{CBO}} = 6.0$ V. BV_{CBO} represents the absolute DC voltage limit on the collector-base junction, regardless of an npn transistor's external

[13] $\beta_F = \beta_{DC} = h_{FE}$ is used for DC base-to-collector current gain, $\beta_{AC} = h_{21} = h_{fe}$ is used for AC base-to-collector current gain.

terminal impedances [101].

The separate bias-voltage references allow fine tuning the bias currents of each power transistor. Moving each stage closer to Class A or deeper into Class AB can increase overall efficiency and output power while tuning the gain and harmonic distortion components.

2.6.1 Schematic Summary

Fig. 2.34 shows the simplified schematic for the active-bias circuit. This active-bias circuit is the unit-cell active-bias generation block across the entire PA. A voltage-feedback bias network sets a target reference current based on a diode-referenced voltage drop across a reference 250 Ω precision tantalum nitride (TaN) thin-film metal resistor within the back-end-of-line metal stackup. This network provides a stable, nearly independent with RF PA input powers, base-emitter voltage reference for the power transistors operating near peak-f_T biasing current densities. Each power transistor (Q1, Q2, and Q3) has its own, independent active-bias network with 5 μm Q_B active-bias transistor. Each of the three active-bias networks has an independent voltage-reference pin separate from the power-cell collector voltage pads. The bias-reference transistor is the 5 μm HBT unit-cell transistor while the power transistor is an array of m such 5 μm unit cells where $m = 2$, 4, or 8 for Q1, Q2, and Q3 gain stages, respectively.

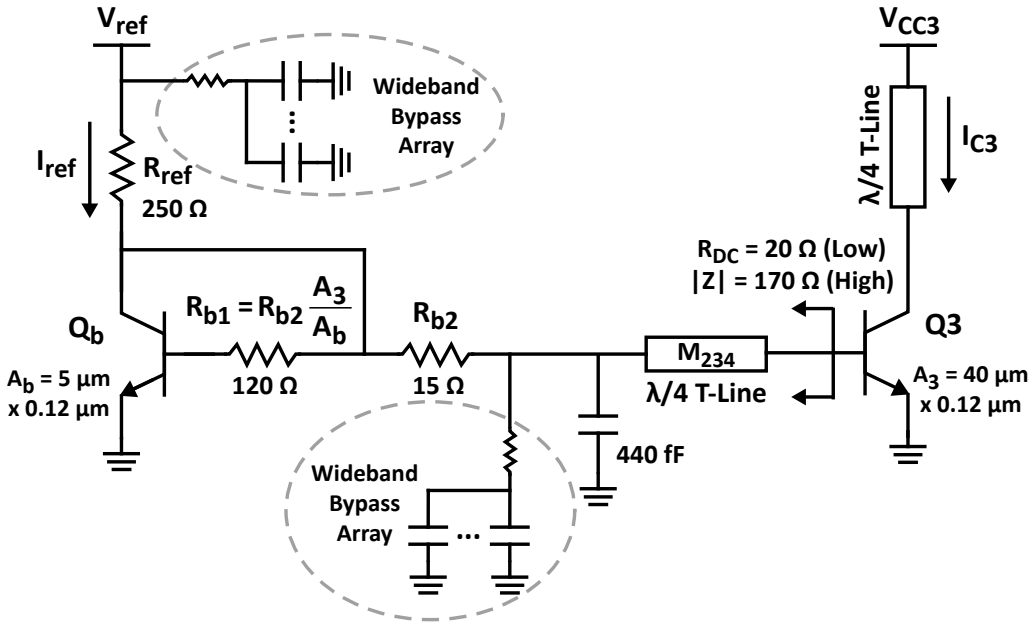


Figure 2.34: Active bias circuit.

$$I_{C3} \approx \left(\frac{\beta}{\beta+2}\right)\left(\frac{V_{\text{ref}} - V_{BE,\text{ON}}}{R_{\text{ref}}}\right) \tag{2.17}$$

In this bias-reference circuit, the external DC base resistance (ballast resistance) presented to power-transistor base will be dominated by R_{B1}. The biasing has additional R_{B1} and R_{B2} resistors for local feedback that results in improved β_F matching with the HBT power transistors. R_{B1} and R_{B2} are scaled inversely to the ratios of areas between the aggregated power-transistor array and the 5 μm bias-reference HBT. In other words, R_{B1} is scaled by the transistor mirroring ratio: $R_{B1} = m \times R_{B2}$. In the case of the Q3 HBT and its bias-reference circuit, $R_{B1} = 120\ \Omega$. At peak-f_T/f_{MAX}/MAG biasing current densities, the DC base-to-collector forward current-gain is $\beta_F = 170$–200.

The M_{234} quarter-wavelength base-feed transmission line connects the active-bias network to the power-cell HBT base terminal. It is made from stacking lower-metal copper layers M2, M3, and M4, which reside beneath the global RF ground plane metal. In conjunction with the 440 fF MIM capacitor, which is resonant at 90 GHz and acts as a low-impedance RF-shorting termination, the M_{234} transmission line becomes a quarter-wavelength resonator that presents a high-impedance (170 Ω at 90 GHz) shunt for DC-bias insertion at the HBT base terminal and 2.4 Ω DC series resistance. More details of the quarter-wavelength bias feeds are covered in Section 2.8.4.

Biasing HBTs near peak f_T & f_{MAX} results in usable 5–6 dB RF gains at 90 GHz, but the intrinsic HBT devices also have tremendous gain (> 40 dB) at low frequencies that must be stabilized. Two separate wideband MIM capacitor bypass arrays, discussed in detail in Section 2.7.3, stabilizes each active-bias network, which in turn, provides additional feedback stability to power-cell transistors at low frequencies below W-band. Additional lower-frequency (< 10 GHz) stability is achieved using off-chip de-Q'd network of SMT[14] capacitors on the needle probes very close to the DC bondpads. Finally, sub-MHz frequencies are additionally filtered by large electrolytic capacitors near the DC power supply terminals.

2.6.2 Biasing and Impact Ionization

The impact ionization base currents is pushed out to higher collector voltages if the generated base currents have a low-DC impedance ballast at the external base node. The active biasing technique in Fig. 2.34 presents 20 Ω, which is nearly a purely voltage-mode biasing. The reference current is set between the externally applied V_{ref} and the internal diode-voltage of Q_B across a precision TaN metal resistor R_{ref}. Resistors R_{B1} and R_{B2} provide localized voltage feedback to improve the DC forward-current base-to-collector β matching

[14]Surface-mount technology, SMT capacitors may are also known as SMDs (surface-mount devices).

since ±25% variations are expected, even between adjacent transistors. R_{B1} is sized by the mirror ratio, m, to split base currents evenly. R_{B1} and R_{B2} attenuate any unwanted RF/AC from entering the bias circuit, stabilize the bias circuit below W-band.

The active bias network presents a very low base impedance of 20 Ω to the external base terminals of the power transistors. This bias configuration is extremely aggressive and allows the power transistors to push higher collector voltages [101]. The effects of impact ionization on operation collector voltage is covered in Section 2.5.3 while the transmission-line structures are covered in Section 2.8.4.

The collector-emitter breakdown voltage, $BV_{\mathrm{CEO}} = 1.8$ V, is effectively pushed higher toward the absolute device-breakdown limit $BV_{\mathrm{CBO}} = 6.0$ V and leaves more headroom for RF/AC dynamic load-line swings. The operational collector voltage increases from V_{CC} = 1.3–1.6 V to V_{CC} = 1.7–2.3 V in the common-emitter configuration.

2.6.3 Bias Currents

Standing Bias Currents

A high standing bias current in the reference transistor of each active-biasing network is important for bias stability when using presenting such a low external base DC resistance (20 Ω), which is closer to voltage-mode biasing than current-mode biasing. The PA design uses higher currents in the active bias circuits than most analog/RF designs in order to create a low-impedance, stable bias reference that does not shift even when the PA is in gain compression. The biasing insures stable thermal operation to prevent runaway catastrophic breakdown in the power-transistor cell. Diode active biasing provides a very low-impedance voltage source reference with high standing DC quiescent currents. Bias stability is achieved with a high standing diode current such that a near-constant DC quiescent voltage is maintained at the base of the PA power cell.

The simple high standing-current (7.5 mA or 1.5 mA/μm) diode-referenced current-mirror biasing ensures high DC bias stability in the aggregated HBT power-cells. More complex biasing schemes with precise temperature compensation can be incorporated later as needed in the system design. The current mirror has base current splitting with TaN resistors ratioed properly to track the transistor ratioing. This provides local negative feedback around the reference transistors and improves the improve power transistor's current-density tracking to the active bias reference transistor.

For stability for low base-resistance external biasing conditions, no β-helping transistor structures are employed. Bias circuits with additional β-helper transistors are designed to source current into a transistor base, but they are not necessarily suitable for HBTs that

are operating well beyond $BV_{\mathrm{CEO}} = 1.8$ V, where HBT DC base currents may actually reverse due to impact-ionization ballasting. The external base terminal needs to have the ability to sink current (ballast) coming from the power-cell base terminal at high collector voltages. A β-helper structure typically assumes a current-mode (forced base) condition, not a voltage-mode, low node-resistance condition, which would sink much of the extra β-helper current in the form of increased power dissipation.

Low Mirroring Ratios

Typical mirroring ratios in large-signal circuits are between to 10:1 to 20:1 or more for bias efficiency, but unusually low mirroring ratios are used in this PA in order to maintain an extremely stable reference while presenting a low external base resistance. The bias transistor and power transistors are biased near the same 1.5 mA/μm peak-f_t/f_{MAX}/MAG current density as the power-cell HBTs for better transistor matching and bias tracking. The resulting 7.5 mA reference current in each active bias circuit results in 2:1, 4:1, and 8:1 mirroring ratios for Q1, Q2, and Q3, respectively, which is also directly related to the number of aggregated unit cells in each power transistor. The small mirroring ratios allow fine bias control of the collector currents while maintaining a stable reference voltage when the power transistors require more base currents at high RF input drives.

The bias-reference current is much larger than required base currents for the power-cell HBTs at any RF input drive level well into full PA compression. The extra standing bias current ensures that the biasing transistor tracks the power transistor and does not limit the compression characteristics of the power transistors in gain compression. The bias networks can supply/sink the required base currents throughout the RF input power range. The overall PA linearity also improves when bias currents stay constant over a varying AM envelope.

Bias Matching

The reference transistor and standing currents for Q1 and Q2 are particularly oversized since 2:1 and 4:1 mirroring ratios are extremely low, but the extra power dissipation is necessary to maintain the unit-cell design methodology so that all bias circuits and power-cell HBTs shift with process (and un-modeled effects) together. The penalty on efficiency is minimal since Q3 is still the dominant limitation on PAE and efficiency.

The low mirroring ratios in the current mirrors allow for very precise current settings in the HBT power-cells, especially at higher collector voltages (> 2.3 V) where collector biasing becomes very sensitive. Low mirroring ratios also results in less bias sensitivity to beta matching and transistor matching, which is particularly helpful since the bias-reference

transistor and power-cell transistors are not physically close to each other (separated by $\sim$ 300 μm) in order to implement quarter-wave transmission-line bias insertion networks.

Although the bias circuit and associated power-cell HBT are separated by ∼300 μm (straight-line distance) and not directly thermally coupled, the internal junctions temperatures do track together, on the first order, because the bias circuits are operating at the same quiescent current density. The mutual thermal coupling between a power transistor's densely-packed 5 μm cells causes thermal differences between power transistor and bias circuit that must be manually compensated using the independent reference-voltage provision in the active bias circuit. The packing arrangement of the unit cell, mutual thermal coupling effects, and distance arising from quarter-wavelength transmission-line biasing all contribute to device mismatches between the unit cells in the power-cell transistors and the unit-cell transistor in the reference bias.

The bias transistor and power transistors have similar current density, resulting in approximately m times the current in the power transistor compared to the reference current. However, this scaling relationship is not exact due to non-idealities in the current mirror which result in differences in effective DC β_F between bias and power transistors: (1) mutual heating of the m transistor cells causes uneven temperature increases across the cells and is not modeled in the transistor models. (2) Low early voltage, i.e., V_A = 2–5 V, with lower values at higher bias currents results in significant variation of collector current with V_{CE}.

DC forward β_F (base-to-collector current gain) is $\sim$ 170 around high-current, peak-f_T biasing, and β_F can vary as much as ±25% between adjacent, identically-sized HBTs and even more for different-sized HBTs that are separated from their reference transistors, as the case here with current mirrors and quarter-wave transmission-line base biasing. Lower mirroring ratios reduce transistor mismatches, and the bias structure scales the mismatches down further by approximately $\beta_F = 170$.

The bias circuits in the half-PA consume a total of 22.5 mA of DC quiescent current. This is 18% of the total quiescent DC power dissipation in the half-PA in light Class A/AB operation[15]. In overdriven Class A or Class AB operation near gain compression, the total PA DC power dissipation increases beyond that of the quiescent values, and the standing bias circuits in the active circuits, which effectively do not change with RF input powers, become a much smaller fraction of the power consumption.. The efficiency characteristics of the final Q3 output stage still dominates the overall efficiency of the PA.

[15]Light Class AB means that the Class AB operation is much closer to Class A than Class B in terms of conduction angles and associated DC biasing

Biasing for Device Aging

The HBT RF small-signal and large-signal characteristics are highly dependent on the DC biasing point. The RF performance can be normalized and tuned by adjusting the DC biasing voltages and currents to expected design values. Fixed offsets in expected DC biasing points arise from modeling shortcoming in device effects (e.g., thermal coupling, impact ionization). In contrast, shifts in biasing points with time (age) arise from long-term stress-related aging of HBT devices. As long as DC currents and voltages can be adjusted, RF performance can be re-tuned, and any stress-related HBT device shifts can be recovered.

Variation in DC β_F between two similar transistors (ideally identical, but different due to expected mismatch) does not appreciably affect RF performance as long as the devices have the same DC voltage & current biasing conditions V_{CE} and I_C. HBTs can also have varying DC β_F (h_{FE}) across the same die, but the active-bias circuit scale down these errors by approximately a factor of $\beta_F = 170$.

Base current is expected to change throughout the lifetime of the device. Although the precise mechanism is not understood yet by the research community, the shifts in base currents are thought to arise from the changes in oxide interface properties between the polysilicon and silicon crystal emitter. Transistor characteristics such as β_F can reduce by 5% over the expected lifetime of the device, and designs must be able to account for these aging DC characteristics [102] with proper bias provisions.

2.6.4 Biasing Tuning Provisions

Each gain stage has both collector current and collector voltage biasing provisions that allow the PA to be adjusted from pure Class A (and overdriven Class A) to various depths of Class AB with reduced conduction angles. The independent current-tuning provisions are essential to optimize the PA over the higher collector voltages on the measurement bench since simulations have limited convergence or limited accuracy beyond 1.7 V at W-band. The independent bias adjustments also allow the gain stages to be fine tuned separately to counter the effects of process variations or modeling errors.

The PA biasing design takes into account that collector voltages, V_{CC}[16], will be raised from the nominal 1.7 V to explore higher voltages and output powers where simulations under the VBIC model did not converge under large-signal conditions. As collector voltages are raised beyond $V_{CC} = 1.7$ V, the power-cell HBT DC-bias points become more and more sensitive to minor DC adjustments. Since proper RF performance is directly coupled to

[16]For common-emitter configurations with no external emitter ballasting/degeneration, the HBT V_{CE} is approximately the supply V_{CC} less the voltage (IR drop) across the collector bias feedlines.

obtaining target DC current & voltage operating points, having separate bias controls for each gain stage is essential for re-centering each gain stage as DC collector voltages are raised beyond 1.7 V.

The Early voltages for the HBTs are only 2–5 V near peak-f_T biasing and $V_{CC} \geq 1.7$ V. This makes the collector current sensitive to the applied collector voltage: increasing the collector voltage without adjusting the reference-bias provision increases the quiescent DC collector current beyond the target design values.

At higher collector voltages, the effective β_{DC} of the HBT increases because of the Early effect. The increased collector-base voltage increases the depletion width of the collector-base junction, effectively narrowing the base width, which causes the β_{DC} to rise.

$$\beta_F = \beta_{F0}\left(1 + \frac{V_{CE}}{V_A}\right) \tag{2.18}$$

where β_{F0} is the base-to-collector gain at zero collector bias, and the effective Early voltage, $V_A = 2$–5 for the 8HP HBT, depending on specific currents and voltages. V_A itself has a strong dependence on collector voltage and current.

The active bias circuits are tuned accordingly to set and fine-tune the DC quiescent collector currents independently in the power-cell HBTs. Re-centering the DC bias currents also re-centers the expected RF performance as collector voltages are increased beyond 1.7 V.

2.6.5 TaN Unit-Cell Bias Resistors

Precision tantalum nitride (TaN) back-end-of-line resistors are used as the unit-cell resistor throughout the design. All resistors on the chip are TaN, and similarly-functioned resistors (e.g. bias-reference resistors and local-feedback base resistors) are sized identically in dimensions in all the PA biasing networks of all three gain stages.

TaN thin-film metal resistors are high precision, have no voltage-dependent resistance and little parasitic capacitance, and are capable of handling relatively high currents found in the in the active bias-generation circuits. Of the available IC resistors, TaN thin-film metal resistors are the best tolerance resistor within GF 8HP with ±8% tolerances (compared to polysilicon resistors or substrate resistors at ±20%). TaN resistors have a low sheet resistance of $R_s = 60.5\ \Omega/\square$ at 25°C, which is appropriate for synthesizing the necessary small-valued resistances (i.e., 3–300 Ω). As a unit-cell component, similarly shaped TaN resistors offer the best matching across the PA chip, compared to any other resistor, with better than 0.08% matching between resistance values.

The dimensions of the precision 250 Ω reference bias resistor in the PA are $W = 20$ μm and $L = 82$ μm. These 250 Ω reference TaN resistors are kept far (> 50 μm) from the

reference-bias HBTs and other supplemental TaN resistors to ensure that heat generated from the resistors do not affect the reference HBT and vice versa. The self-heating effects and nearby copper-metal auto-filling effects (for local density requirements) are included in the foundry resistor models.

The current handling limits of this resistor is 0.5 mA/μm width, which translate to 10 mA for temperatures below 70°C or 7.5 mA at 100°C. The resistors used throughout the biasing circuits follow the 100°C limits for 100,000 power-on-hours reliability guidance.

Identically-sized, wide, 3 Ω, $W \times L = 25\ \mu\text{m} \times 3\ \mu\text{m}$ TaN resistors are used as de-Q resistive elements in all of the MIM wideband bypass networks discussed in Section 2.7.3.

2.7 MIM Capacitors

Metal-insulator-metal (MIM) capacitors are key components throughout the W-band PA design. Only MIM-type capacitors are used[17] in the PA because of their compactness and high capacitance density as DC blocks, lower-frequency bypass filtering, RF reactive matching elements, and RF low-impedance references (AC shorts) in both DC bias-insertion networks and RF matching networks. Over 55% of the total PA chip area is consumed by MIM capacitors and their associated wiring.

The MIM capacitor is the highest capacitance-per-area density device within 8HP. A 62 nm nitride insulator dielectric with relative permittivity $\epsilon_r = 7.0$ is sandwiched between two aluminum plates. The resulting capacitance density is 1 fF/μm^2, which is 10× the capacitance density achievable by using a silicon dioxide (SiO_2) capacitor from two vertically-adjacent metal-wiring levels (e.g. M3-to-M4 or M2-to-M3).

Eq. (2.19) is the equation for the low-frequency capacitance of a parallel-plate capacitor, applicable to the compact MIM capacitor away from self-resonance. The bottom-plate and perimeter parasitics are not taken into account here, and they are typically 6% of the nominal parallel-plate capacitance.

$$C = \epsilon_r \epsilon_0 \frac{A}{d} \tag{2.19}$$

The MIM capacitors are essential components in the unit-cell design approach. The MIM is chosen to be the only capacitor type used in the PA design for several reasons. Firstly, the high-density capacitance allows for compact in-line layouts that easily integrate with the low-loss top-metal AM transmission lines. Additionally, since MIM capacitors are located near the top-level metal, they can be easily connected between transmission lines

[17]The exception is a design variation experiment with metal-oxide-metal (MOM) capacitors in the matching networks, discussed further in Section2.10.1

with smaller wiring parasitics. Further, any process variations and modeling errors shift all MIM capacitors in the same direction, ensuring that the design does not severely de-tune because of component skews. Where possible, similar aspect ratios (width/length or W/L) are used for all MIM capacitors, regardless of absolute capacitance values.

Throughout the PA design, MIM capacitors are used for several functions:

1. as matching reactances and necessary interstage DC blocking;
2. as RF-shorts (zeroing capacitors) at their natural resonant frequency for terminating quarter-wavelength transmission lines and as high-impedance elements at the second harmonic ($2f_0$) for PA harmonic termination management;
3. as RF-shorts in resonant interstage matching network structures with DC blocking;
4. as elements in wideband bypass arrays for reducing gain outside of W-band (for RF and low-frequency stability) and for DC-bias noise filtering.

The MIMs in the matching networks and resonant interstage matching support the necessary voltage and RMS currents for reliable 100,000 power-on-hours operation at 100°C.

MIM Modeling

The MIM compact models from the foundry are only directly modeled with measured data up to 40 GHz, and some unmodeled second-order effects of losses at W-band are expected. Despite these limitations, MIM circuit modeling used for the PA relies heavily on the design-kit scalable SPICE-based model for several key reasons: (1) all the necessary physical constants of the MIM are not available to designers; (2) the foundry MIM model is based on many lots of measured hardware-correlated data with statistical distribution modeling; (3) design-kit model has integrated the effects of surrounding copper metal-fill and mandatory active-layer fill within the substrate; and (4) planar (or 2.5-D) EM simulators have difficulties with thin dielectrics ($\sim$62 nm) and dielectric bricks (non-infinite-plane dielectric layers). The EM modeling and simulation of the metal leads to the MIM capacitors are absorbed into the connecting transmission lines.

MIM Quality Factor

Compared to Ka-band, where the MIM capacitor quality factor Q is $\sim$ 20–30 @ 30 GHz, the MIM Q is just 7–10 @ 90 GHz for all implemented MIM capacitors in the W-band PA. The MIM capacitors have high operating voltage limits of 17 V. The operational current limitations are determined by the contact vias to the top plate, and the number of vias are

maximized for all implemented MIM layouts to ensure that reliability guidelines are met for with 100,000 power-on-hours (POH) reliability at 100°C.

Given the relatively low-Q nature of MIMs at 90 GHz compared to transmission-line distributed elements, as few MIM capacitors are used as possible in the RF portion of the PA (non-DC-bypass filtering capacitors). Matching networks are kept to as few elements as possible so that only the minimum number of capacitors are used to avoid excessive RF passive losses. The advantage of the MIM over other higher-Q capacitances is the extreme compactness, high precision, and device matching across the chip.

MIM Tolerances

The PA designs must take into account capacitor tolerances; otherwise, shifts in expected reactance values can result in severe degradation of RF performance. Since many capacitors are used as reactive matching elements and resonant circuit elements, the tolerance is particularly important, and the choice of using a lower-Q nitride MIM capacitor over a higher-Q ($\sim$20) metal-oxide-metal (MOM) capacitor is essential in the design.

The 8HP fabrication process results in MIM capacitances with tight $\pm$10% tolerances around their designed capacitance values. This is a small tolerance compared to custom metal-oxide-metal capacitors designed in the lower-level copper metals and SiO_2 dielectric. The copper damascene process in 8HP results in large variations of copper and SiO_2 thicknesses: the oxide thickness between two vertically-adjacent copper metal layers has a variation of $\pm$23%, and each metal plate can vary in thickness by $\pm$38%. The excellent tolerance and compactness of the MIM result in very predictable reactance values at 90 GHz, making the MIM preferred over higher-Q solutions like MOM capacitors (poor tolerances) or large transmission-line capacitors.

For this iteration of the PA design, the MIM capacitor is chosen over MOM capacitors since achieving precise matching networks across the entire PA is more important than loss penalty incurred from the lower MIM Q. More discussions of an experimental design implemented with both MIM and MOM capacitors is found later in Section 2.10.1. For the same capacitance, MOM capacitors are, at best, 10$\times$ the area (3.2$\times$ larger in each length & width linear dimension) compared to MIM capacitors. The precise matching among MOM capacitors on the same chip is not characterized, and local copper densities can affect the MOM capacitors differently, resulting in possible high component mismatches between MOMs on the same chip. In comparison, although MIM capacitances can vary $\pm$10% from design values, the capacitance matching across the chip is better than 0.06% for identical capacitors. The outstanding device matching allows matching-network reactances and resonant circuits to track together closely despite process skews, resulting in a more robust PA design.

The unit-cell approach takes advantage of the excellent MIM matching by re-using similarly-dimensioned and valued 91 fF MIM capacitors in the input, output, and Q1-to-Q2 interstage matching networks. A 440 fF unit-cell MIM capacitor is used as low-impedance (RF-short) references in all the base and collector quarter-wavelength insertion feeds. The same 440 fF MIM capacitor is used in a resonant LC interstage matching network.

The only MIM capacitor that is slightly different is the 130 fF Q2-to-Q3 interstage matching capacitor because a larger MIM device with additional via contacts is required to satisfy the current handling at the very low-impedance node ($\sim$4 Ω looking into the Q3 HBT power-cell input).

2.7.1 Matching-Network Elements

The MIM capacitors in the input, output, and interstage matching network act both as reactive elements as well as necessary DC blocks between gain stages, input, and output. Except for one larger 130 fF[18] Q2-to-Q3 interstage matching-network MIM and 440 fF RF-shorting MIMs, all other MIMs used as reactive matching elements are identical 91 fF unit cells, following the general unit-cell approach of the entire PA design. The capacitors have $Q = 10$, and the wide, low-inductance leads to the MIM plates result in > 350 GHz self-resonant frequencies.

The PA design pushes the MIM capacitor to its reliability limits as a 91 fF reactive matching element and series DC-block at 90 GHz. At the time of the design in the early 8HP foundry process, the 91 fF are the smallest MIM in 8HP that are compliant to handling the necessary RMS AC currents in the matching networks. MIM capacitors smaller than 91 fF would violate electromigration limitations set by the number of MIM top-plate contact vias. Process improvements since have allowed more vias to contact the top plate, alleviating any current-handling reliability concerns from RMS AC currents and electromigration.

Fig. 2.35 highlights all 91 fF MIM capacitor unit cells used as matching-network elements and DC blocks. The 130 fF MIM capacitor variant breaks from the unit-cell approach in order to remain electromigration compliant and provide more optimal Q2–Q3 interstage matching arising from the very low Q3 input impedance (4–5 Ω).

2.7.2 RF Short

A major difficulty of RF design at W-band is the lack of a compact low-impedance reference (RF short) that also blocks DC currents— a very large ideal capacitor that presents very low

[18] Quoted MIM capacitance values are for low-frequency operation where parasitic inductances, which reduce effective capacitances at mmW frequencies, can be ignored.

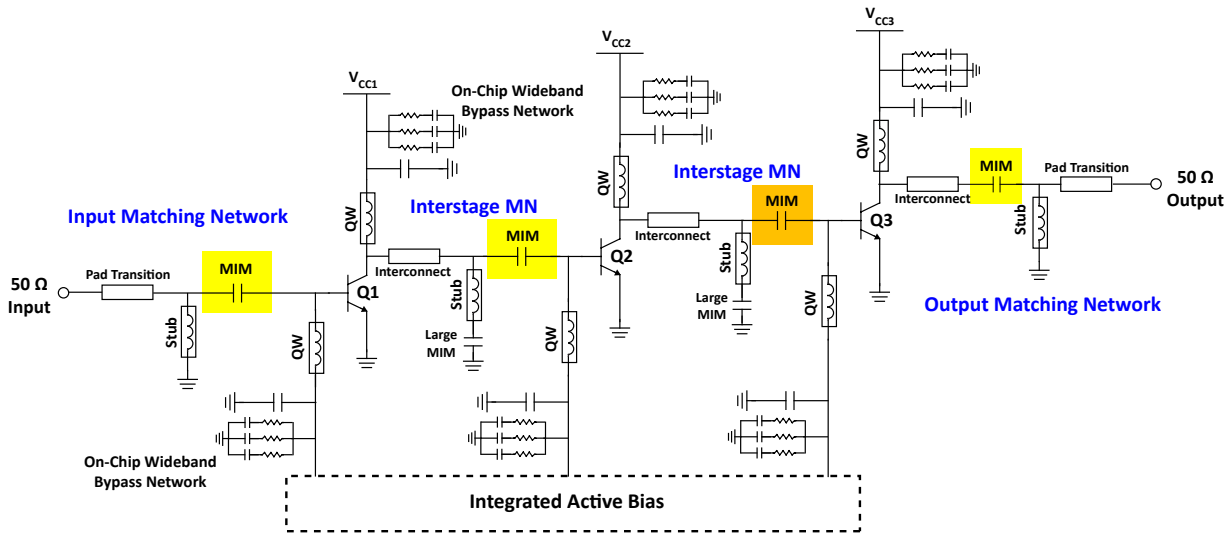


Figure 2.35: Half-PA simplified circuit schematic with all 91 fF unit-cell matching-network MIM capacitors highlighted in yellow and single 130 fF MIM variant highlighted in orange.

impedances within W-band. Inductive parasitic lead inductances in large capacitors (> 1 pF) resonate the capacitors at frequencies well below W-band, making non-ideal capacitors look effectively like inductances with relatively high impedance magnitudes ($> 5\ \Omega$) at 90 GHz.

To obtain a very low RF impedance reference at 90 GHz, a resonant MIM capacitor is used such that only a small residual real-part resistance exists— the equivalent series resistance (ESR). The MIM capacitor is the most compact— highest capacitance density— capacitor element, and its compactness makes it usable as a unit-cell RF-short termination (low-impedance reference) throughout the PA design. Although even lower impedances can be synthesized using distributed lines (e.g., half-wave resonator terminated in an open-circuit), the resonant MIM capacitor solution is extremely compact and can be used repeatedly across the PA chip with wide-bandwidth tolerances.

The aspect ratio of the capacitor influences the self-resonant frequency because of the lead inductances in the top and bottom plates. Fig. 2.36 illustrates the affect of MIM capacitor aspect ratio on self-resonant frequencies for a range of simulated 8HP capacitor values.

The low-frequency capacitance, from a reactance standpoint, appears as a larger effective capacitance at W-band higher frequencies because of parasitic inductances arising from contact vias and plate leads. Fig. 2.37 shows MIM effective capacitances at 90 GHz for different capacitor values and aspect-ratio configurations. Small effective capacitances, necessary for reactive impedance matching networks, become more difficult to achieve at higher mmW frequencies. Wider-aspect ratio (W/L) capacitors have less lead inductance and higher

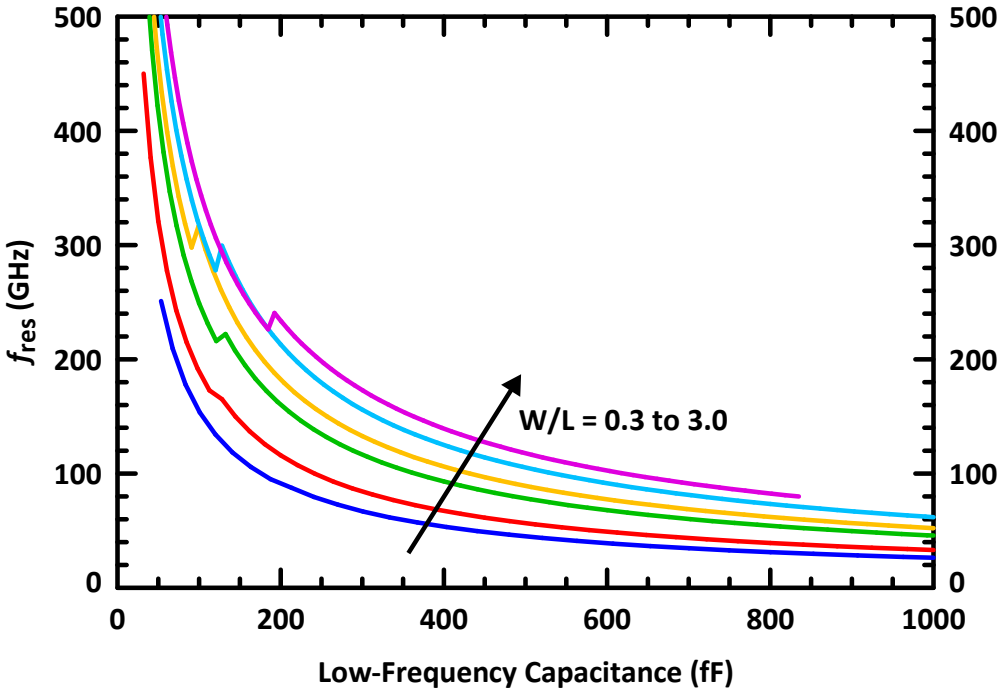


Figure 2.36: Simulated MIM resonant frequency f_0 vs. low-frequency capacitance. W/L aspect ratio = 0.33, 0.5, 1.0, 1.33, 2.0, and 3.0. For a given target capacitance, wider aspect-ratio MIM capacitors result in higher self-resonant frequencies. Discontinuities in the plot are attributed to discrete-stepped geometry changes for smaller MIM capacitors.

resonant frequencies. By choosing capacitor dimensions carefully and modeling the lead inductances in EM, a target resonant frequency can be attained.

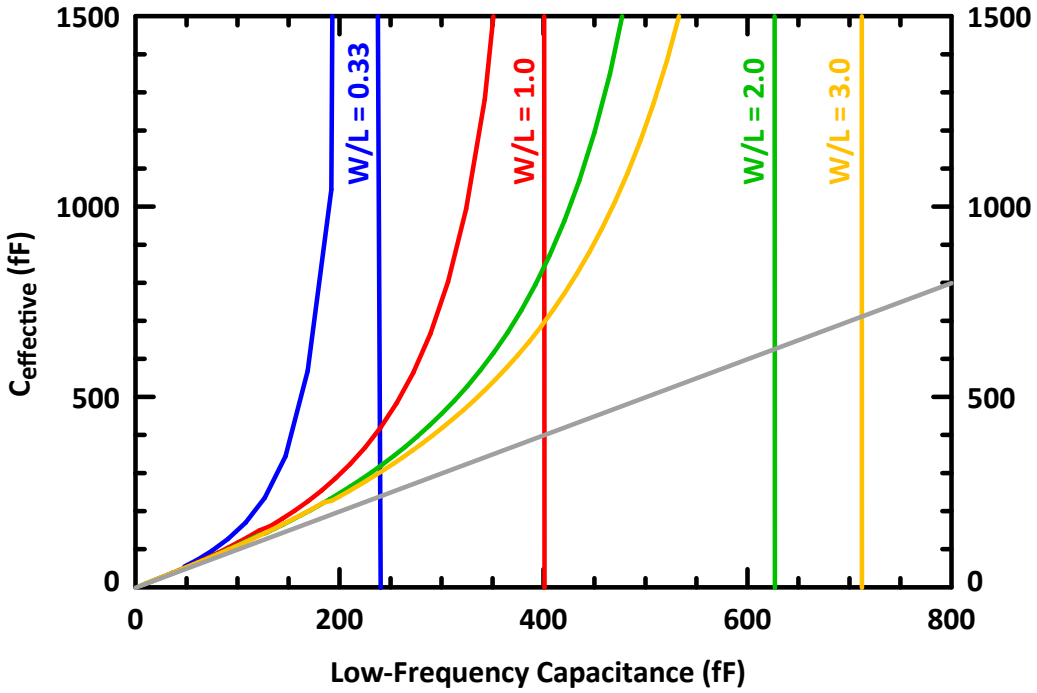


Figure 2.37: MIM effective capacitance at 90 GHz vs. low frequency capacitance. W/L aspect ratio = 0.33, 1.0, 2.0, and 3.0. Vertical lines indicate self-resonance points— MIM appears effective inductive @ 90 GHz beyond these lines.

440 fF MIM Capacitor

A compact, but very large-valued (relative to typical W-band capacitances ∼100 fF) 440 fF, $W \times L = 24\ \mu m \times 18\ \mu m$ nitride MIM capacitor is used near its natural self-resonant frequency as a shunt RF-shorting or zeroing capacitor— a very low-impedance ($< 1\ \Omega$) termination at W-band frequencies. Following the unit-cell design approach, the PA design uses identically-dimensioned 440 fF MIM capacitors at circuit nodes that require low-impedance terminations (RF-short) at 90 GHz. A total of 8 identical 440 fF MIMs are used in shunt as equivalent RF-shorts throughout the half-PA (16 in the full differential PA). The MIM's compactness and vertical proximity to the transmission-line top-metal above and global RF/DC ground below makes it flexible for fitting into tight layout nodes.

The 440 fF MIM allows an extremely compact W-band RF-short to be placed at the proper physical termination point on each quarter-wavelength line while allowing additional DC lines and larger capacitors to be placed beyond the 440 fF MIM capacitor. In a sense,

the 440 fF MIM capacitor acts a pre-fixed sliding RF-short design element.

To achieve maximize the capacitance and thus minimize the residual equivalent series resistance (ESR) at resonance in W-band, a W/L = 1.33 is used for the 440 fF MIM capacitors. Larger W/L ratios are allowed in the process up to 3.0, but the layout compactness and re-usability in different layout scenarios as a unit cell is reduced if the ratio is skewed too far from a square aspect ratio (W/L = 1.0).

The 440 fF MIM capacitor, by itself, self-resonates near 100 GHz. When placed within the PA layout, a MIM's additional inductive lead parasitics from wiring bring down the resonant frequency to 90–95 GHz— closer to W-band midband. The parallel-plate capacitance forms a series RLC resonator ($R = 0.6\ \Omega$, $L = 6$ pH, $C = 440$ fF) with its vias and lead/wiring inductances & resistances.

At the resonant frequency, the MIM capacitor's impedance is purely real at 0.6 Ω. As shown in Fig. 2.38, near the resonant frequency, the 440 fF MIM capacitor provides $< 1\ \Omega$ impedance over ±11% bandwidth (22 GHz) and below 1.5 Ω for a 38 GHz bandwidth, providing a wideband low-impedance reference point that covers all W-band frequencies. The MIM capacitance values have a ±10% tolerance over process variation, so a maximum center-frequency shift of ±5% is expected (95–105 GHz, before parasitics) since $f_0 \propto 1/\sqrt{C}$. Despite the ±10% absolute tolerance, the identically-dimensioned 440 fF MIM have better than 0.1% relative matching among themselves, so all RF-shorting capacitors are unit cells that track each other closely with process variations, ensuring that variations in resonant frequencies also track closely. Additional lead parasitics to the 400 fF MIM capacitor re-centers the resonant frequency of the entire structure close to 90–95 GHz, depending on the layout specifics.

The Q of the RF-short MIM capacitor is 7 @ 90 GHz. This includes the losses of the metal leads and vias to the MIM from the top-level (AM) metal. This Q is relatively low compared to distributed transmission-line techniques (Q = 20–30)and custom metal-oxide-metal (MOM) SiO_2 capacitors (Q > 20), but the MIM capacitors have significantly more compact and higher tolerance allowing critical W-band structures to be significantly smaller and to track closely as unit cells.

In conjunction with quarter-wavelength transmission lines in the collector transmission-line feeds and the base transmission-line feeds, the shunt 440 fF RF-shorting MIM capacitors terminate these transmission lines in a near short circuit, creating an effective transmission-line resonator that presents a high impedance (near open-circuit) at the opposite end at the feed-insertion points of the HBT power cells.

Fig. 2.39 highlights all self-resonant 440 fF MIM-capacitor unit cells used throughout the half-PA.

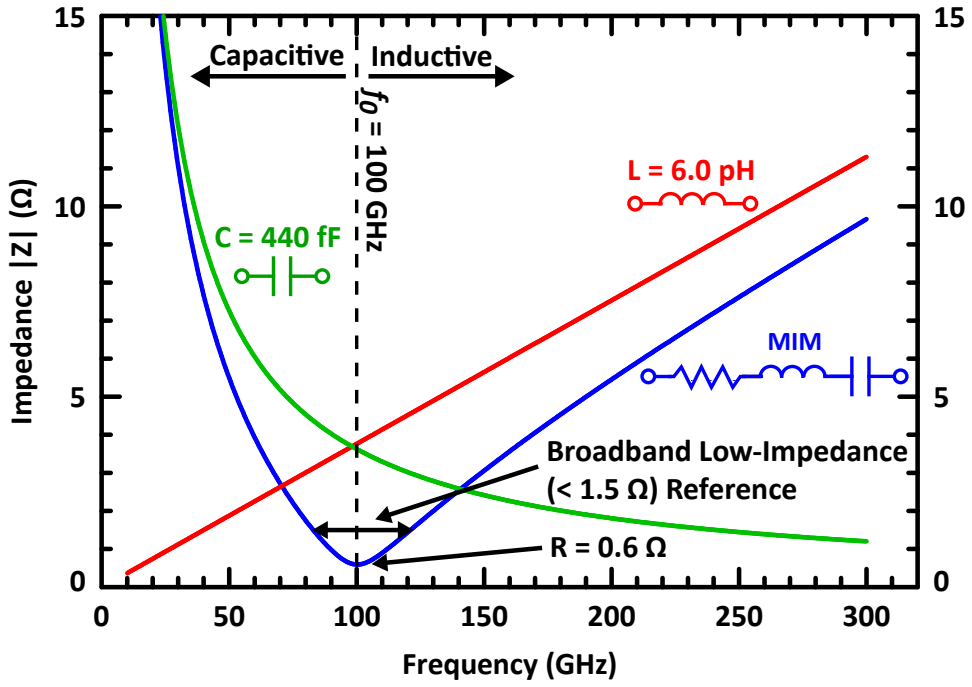


Figure 2.38: Impedance magnitude of 440 fF MIM RF-shorting capacitor and equivalent circuit responses: an ideal 440 fF capacitor (green), ideal 6.0 pH inductor (red), and simplified lumped equivalent model of the 440 fF MIM (blue). The 440 fF MIM capcitor has a natural self-resonant frequency of 100 GHz where it effective presents a purely resistive 0.6 Ω impedance and less than 1 Ω throughout W-band. Above 100 GHz, the 440 fF MIM capacitor response begins to look more and more inductive.

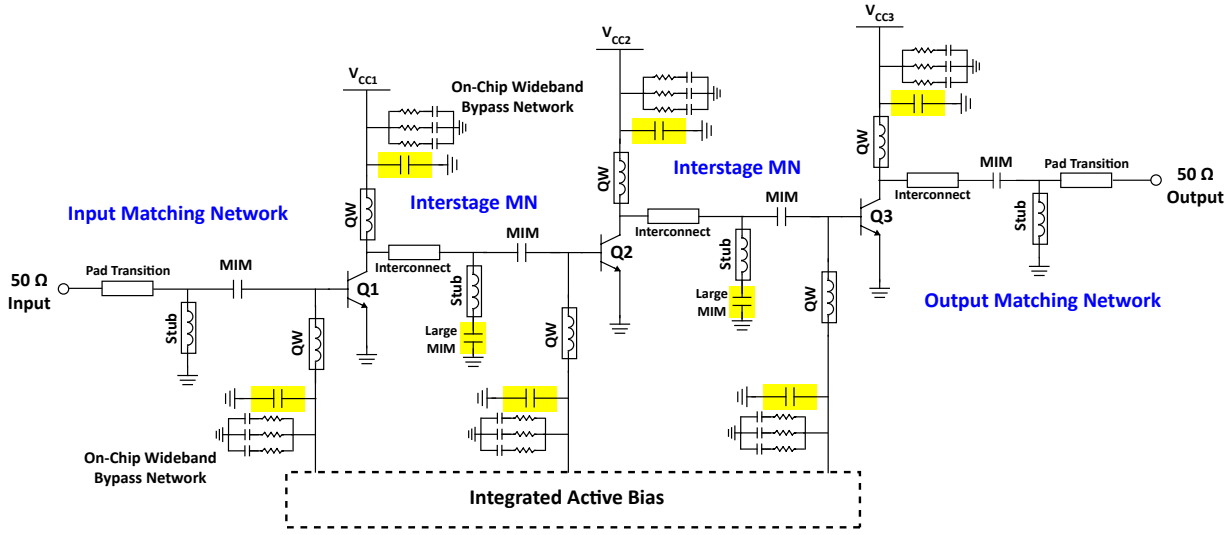


Figure 2.39: Half-PA simplified circuit schematic with all unit-cell 440 fF resonant MIM capacitors highlighted in yellow.

More details about the effects of MIM terminations in the transmission-line bias insertion networks is covered in Sections 2.8.3 and 2.8.4.

Perils of Multiple, Parallel Capacitors at Resonance

For circuit involving capacitors operating at or near resonance, arraying multiple resonant capacitors must be avoided. This issue is not normally a concern when using parallel MIM capacitors well below their resonant frequencies, but at W-band, all 8HP MIM capacitors > 600 fF have self-resonant frequencies < 100 GHz. Adding more similar, parallel resonant capacitors may seem to be an obvious means of reducing the RF impedance at resonance further by reducing the collective equivalent series resistance, but multiple capacitors close in resonance frequencies can easily experience anti-resonance in this configuration.

Small shifts in resonant frequencies among the arrayed capacitors near resonance can lead to anti-resonance conditions where the impedance becomes much higher than expected. Although adjacent MIM capacitance matching is excellent ($< 0.1\%$), very small differences (1-3%) in lead inductances either from fixed design differences or from process variation in the vias and lead metals can lead to high-impedance anti-resonance peaks because the poles of the two parallel RLC resonators interact when their resonant frequencies do not correspond exactly.

For example, between two nearly-identical 440 fF MIM caps, a small 3% shift— just 0.2 pH out of the total 6.0 pH parasitic— causes a high-impedance anti-resonance within W-band, as illustrated in Fig. 2.40 for the two-component case. Adding additional resonant capacitors in parallel only exacerbates the problem, resulting in multiple anti-resonance peaks close to the resonant frequency and directly in the bandwidth of interest where low impedance is desired. The in-band anti-resonances can lead to amplifier instabilities and oscillations as well as extreme notches in the gain response that lead to severe problems with signal linearity and integrity. Thus, only one resonant capacitor should be used in the RF-shorting structures (unless others are de-Q'd with resistors, as done with the wideband MIM array, presented in Section 2.7.3).

2.7.3 Wideband Bypass

A fully-integrated array of MIM capacitors is used as the unit-cell element to terminate nodes in the PA with low impedance over a wide range of frequencies well out of W-band, from 10 GHz to 200 GHz (at the PA's second harmonic). 9 identical instances of the array are used in the half-PA (18 total for the differential PA). Each array instance is approximately 250 μm × 200 μm, including the surrounding low-resistance grounding structures. Although

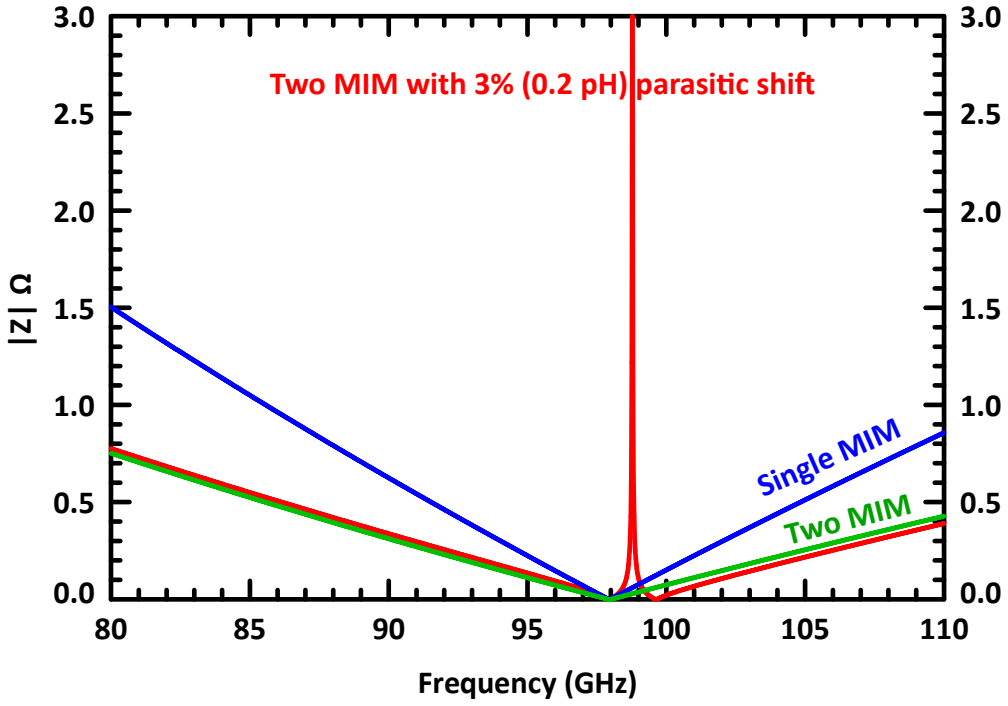


Figure 2.40: Resonant 440 fF MIM capacitor impedances for a single MIM capacitor (blue); two identical 440 fF MIMs in parallel ideally (green), and two parallel 440 fF MIMs with 3% (0.2 pH) parasitic shift from small differences in wiring or process skew (red).

the wideband-bypass MIM arrays consume large amounts of chip area, the fully-integrated capacitances allow for further circuit integration, especially tiling multiple PA instances in larger arrays, as shown later in Section 4.

The array of MIMs provide compact, low-frequency bypass and bias-noise filtering capacitance in the bias circuits and at each DC wirebond pad and within the active biasing circuits.

At mmW frequencies below W-band (10–75 GHz) where the intrinsic transistor gain increases significantly, the wideband MIM bypass arrays stabilize the W-band HBTs by presenting low-impedance terminations to all HBT power-cell transistors' collector and base terminals. At the W-band second harmonic $2f_0 = 180$ GHz, the MIM array, together with the 440 fF MIM, collector transmission-line feed, and base transmission-line feed present a $< 3\ \Omega$ impedance to the second harmonic content for Class AB harmonic control. (The quarter-wavelength base & collector transmission-lines at W-band are half-wavelength transmission lines around the second harmonic, 180 GHz.)

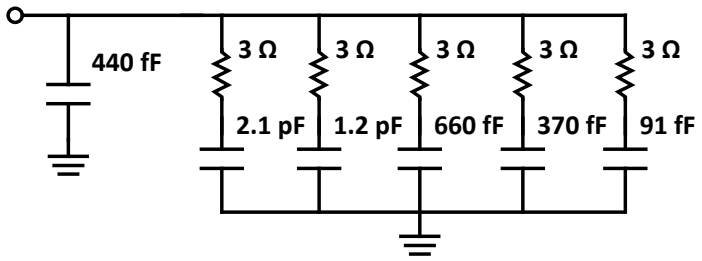


Figure 2.41: Wideband bypass array (unit cell) circuit schematic. All capacitors are MIM types. Resistors are unit-cell 3 Ω n^+ subcollector diffusions resistors, all identically dimensioned.

The wideband bypass array consists of 91 fF[19], 370 fF, 660 fF, 1.2 pF, and 2.1 pF MIM capacitors, each with its own 3 Ω de-Q resistor to provide a wideband response and to avoid high-impedance anti-resonance peaks in the AC impedance response.

The MIM capacitors within the array are weighted & scaled to give the desired wideband frequency response. The wide range of capacitors result in a wide range of low-impedance resonant frequencies where the impedance of a corresponding self-resonating capacitor is lowest. The MIM array avoids similar or repeated capacitor values in parallel to avoid anti-resonance peaks. MIM capacitance values are roughly scaled 2× to synthesize a wide range of self-resonant frequencies.

Series resistors are used in each individual MIM capacitor to de-Q the network in order

[19]This is the same unit-cell MIM capacitor used throughout the RF input, output, and interstage matching networks.

to widen its impedance response and prevent unintended impedance responses from coupling between multiple resonant structures that cause antiresonance peaks (spikes of high impedance) which can cause instabilities in the active circuits below W-band.

The 3 Ω resistor is an n^+ subcollector diffusion resistor formed from an unsilicided NFET source/drain junction. The substrate resistor sits on top of the silicon p-substrate and below the shallow-trench isolation (STI) oxide. The sheet resistance is only 8.8 Ω/□ and allows for a compact resistor in the capacitor array layout for which a precision TaN metal resistor is too large.

The entire bypass array sits in parallel to the 440 fF RF-shorting unit-cell capacitor, which does not have a de-Q resistor in order to provide the lowest possible 0.6 Ω impedance at self-resonance within W-band.

Fig. 2.42 illustrates the overall wideband capacitor array response along with the natural self-resonant frequency responses (without de-Q resistors) of the individual MIM components within the array. The wideband bypass capacitor array presents an impedance < 2 Ω from 15–190 GHz, < 3 Ω down to 10 GHz, and < 5 Ω down to 6 GHz, and < 10 Ω down to 3 GHz. The array aides in the second-harmonic termination (impedance < 2.6 Ω @ $2f_0$) necessary for Class AB operation.

The low-frequency ($<$ 3–6 GHz) capacitive filtering is distributed between both on-chip and off-chip capacitance networks. The off-chip capacitive networks are integrated into custom DC needle probes that provide bias to the PA's DC wirebond pads. Each DC probe needle has 120 pF of bypass capacitance to an adjacent ground probe needle. An additional 0.1 μF bypass capacitance (with a 5 Ω de-Q resistor) sits on each line of the DC wedge probe board. Additional very large capacitances (100s of μF) are part of the DC measurement cabling setup. (See Section 2.11.1 for more details on the off-chip capacitors on the DC wedge probes.)

2.8 Transmission Lines and Distributed Inductors

The unit-cell design concept is also applied to the transmission lines used throughout the W-band PA. Similar transmission-line structures, with similar characteristics, are used throughout the PA for RF input/output lines, shorted shunt-stub matching network elements, collector feedlines, and transmission-line elements in the interstage resonant matching networks. The transmission-line behaviors track together closely— matched— across the entire PA chip, and expected process shifts and minor modeling errors do not de-tune the sensitive PA design.

With the dielectric constant (relative permittivity) of silicon dioxide (SiO_2) at $\epsilon_r = 4.1$, transmission-line structures are approximately $\sqrt{4.1}$ or half that of free-space wavelengths.

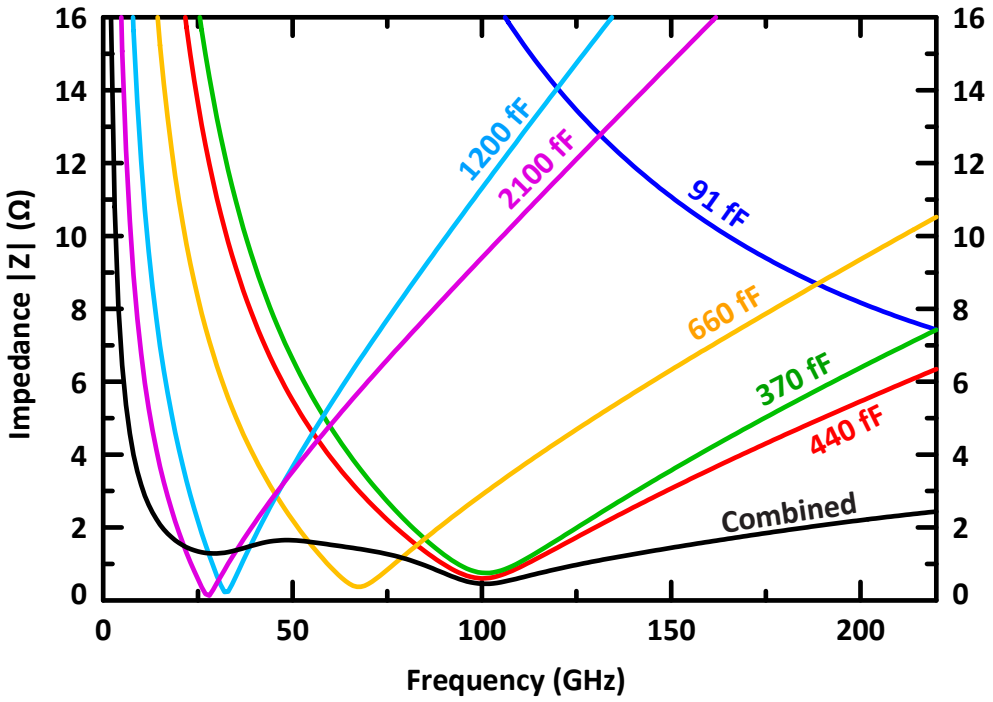


Figure 2.42: Impedance response of the wideband bypass array and the native impedance responses of the individual 91 fF, 370 fF, 440 fF, 1.2 pF, and 2.1 pF capacitors without the de-Q resistors. Off-chip capacitors included in the DC needle probes for used for further < 10 GHz bias filtering are not included here.

The precise characteristic guided wavelengths vary slightly with transmission-line cross sections and their effective guided ϵ_r, which are characterized in dedicated 2.5D EM simulations.

The characteristic guided[20] wavelength within a transmission line is given by Eqs. (2.20) and (2.21) [103].

$$\lambda_g = \frac{2\pi}{\beta} \tag{2.20}$$

$$\lambda_g = \frac{\lambda_0}{\sqrt{\epsilon_{r,eff}}} \tag{2.21}$$

where β is the transmission line's characteristic phase constant (imaginary part of the complex propagation constant).

Eq. (2.22) gives the phase velocity of the wave.

$$\nu_p = \frac{\omega}{\beta} \tag{2.22}$$

where ω (in radians) $= 2\pi f$ with f the frequency of operation (in Hz).

The transmission-line structure's effective dielectric constant is determined from comparing the free-space wavelength to the guided wavelength in Eq. (2.23).

$$\varepsilon_{r,eff} = \left(\frac{\lambda_0}{\lambda_g}\right)^2 \tag{2.23}$$

The nominal SiO_2 dielectric within 8HP has a dielectric constant $\epsilon_{r,\text{eff}} = 4.1$, but its effective value can be as low as 3.6 due to effects of forced copper-metal fill (or alternately, the effective dielectric thickness within copper layers shrinks from the small pieces of floating metal fill).

At W-band frequencies, skin depth plays an important role in the expected losses of the metals. The skin depth of a good conductor is given in Eq. (2.24) [62].

$$\delta = \frac{1}{\sqrt{\pi f \mu \sigma}} \tag{2.24}$$

where f is the frequency, μ is the material permeability, and σ is the conductivity. Within a chosen conductive material, the skin depth δ_s decreases with increasing frequency by a factor of $\sqrt{f}$. The skin depth of aluminum in 8HP (layers AM and LY) is $\delta_{s,Al} = 0.28$ μm @ 90 GHz while the skin depth of copper is $\delta_{s,Cu} = 0.23$ μm @ 90 GHz (layers MQ, M4, M3, M2, and M1).

[20]The "g" subscript is sometimes used in this text to indicate "guided" wave in order to distinguish from transistor parameters that also use β.

2.8.1 AM-Layer Transmission Lines

In the GF 8HP back-end process, the top-level metal (8HP layer: AM) is suitable for low-loss RF transmission lines at mmW frequencies. Fig. 2.43 illustrates top-layer aluminum AM-metal structure used in all AM transmission lines in the PA.

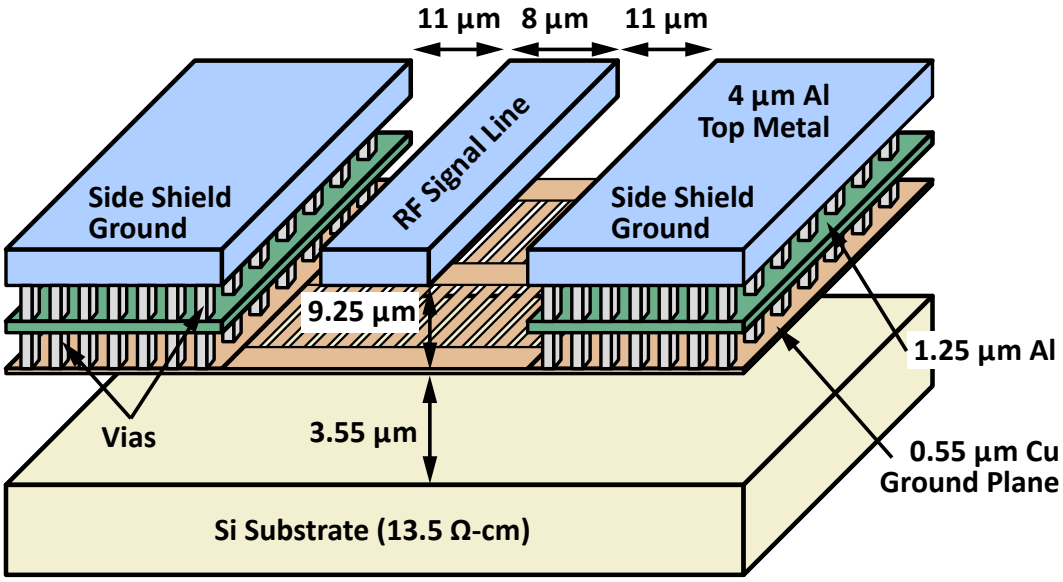


Figure 2.43: Transmission line cross section within GF 8HP.

Except for the collector-feed transmission line, which carries particularly high DC currents, all top-level (AM layer) transmission lines are 8 µm wide with 11 µm lateral spacing to side RF-ground shields and 9.25 µm vertical spacing to the global RF ground plane on layer MQ beneath. The resulting transmission line has a characteristic impedance of $Z_0 = 50\ \Omega$ in W-band. This is the unit-cell transmission-line used throughout the PA and for the 50 Ω input and output lines that interface to external 50 Ω (single-ended) microprobes.

The global RF ground plane is chosen to be copper layer MQ, which has a 0.55 µm thickness and sits below the top-level AM aluminum metal. The MQ RF-ground plane isolates the top-level (AM) transmission lines from the lossy p^- doped silicon substrate below with a bulk resistivity of $\rho_{sub} = 13.5\ \Omega{\cdot}$cm. The global RF-ground plane on layer MQ covers the majority of the PA area except over HBT transistors, MIM capacitors, TaN resistors, and RF & DC bondpads. The ground plane also is tied to additional DC-ground wiring on the uppermost two thick, low-loss aluminum metals (AM & LY) using arrays of vias throughout the PA chip.

Together, the wide side-shielding structures and large ground plane (below) form a fully

RF-shielded tub around the microstrip transmission line. (The top side of the transmission line is not covered/shielded.) Shielding for the top-level AM transmission lines is necessary to ensure that AM collector lines do not couple to adjacent collector lines on the same layer or to base-feed transmission lines and other structures underneath the AM transmission lines. The sidewall shields prevent unwanted lateral coupling to adjacent structures and allows the AM transmission lines to be packed closer together. The bottom shield (RF-ground plane) also prevents field penetration into the lossy substrate, increasing the overall quality factor of the transmission line.

The transmission-line side and bottom shielding increases the transmission-line loss and lowers its characteristic impedance Z_0 compared to an unshielded line, but the isolation is required to maintain PA stability across all frequencies, from near-DC to W-band. (Without side shielding or MQ ground plane, an microstrip transmission line on AM has a characteristic impedance closer to $Z_0 = 60\ \Omega$ at 90 GHz.)

Fig. 2.44 show EM simulations results of line-to-line isolation (from lateral coupling) versus separation distance for shielded and unshielded microstrip lines using top-level (AM) metal. The shielded transmission lines improve the isolation between adjacent lines by 15–20 dB over unshielded lines at 90 GHz. Two shielded microstrip AM transmission lines separated by 40 μm have a simulated isolation of > 40 dB @ 90 GHz.

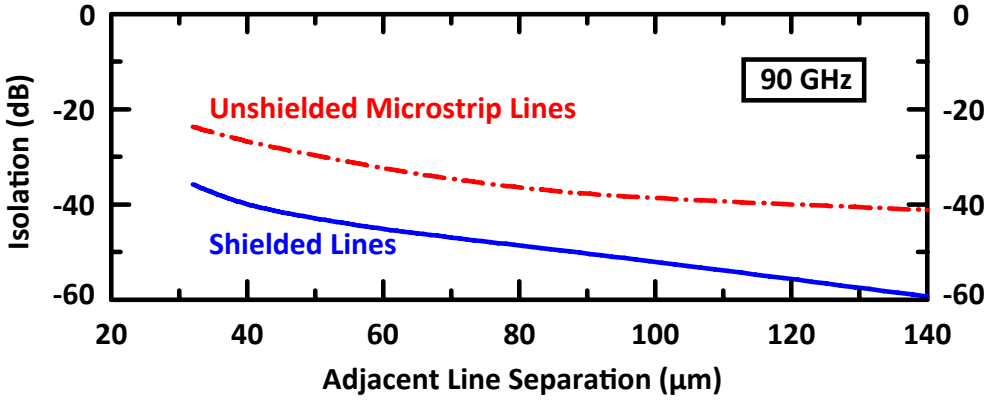


Figure 2.44: Isolation versus line separation for shielded and unshielded microstrip transmission lines on top-layer metal AM.

The transmission line sits in silicon dioxide ($\epsilon_{r,\text{eff}} = 4.1$), but the top side dielectric stack is composed of 1.35 μm layer of silicon dioxide (SiO_2), 0.45 μm layer of nitride ($\epsilon_{r,\text{eff}} = 7.1$), and 1.0–7.5 μm (2.5 μm nominal) layer of polyimide ($\epsilon_{r,\text{eff}} = 3.4$). The dielectrics above increase the $\epsilon_{r,\text{eff}}$ of top-level transmission-line structures to $\epsilon_{r,\text{eff}} = 4.3$. The transmission-line series insertion loss for the $Z_0 = 50\ \Omega$, 8 μm line with 11 μm side spacing (notated as 11–8–11 μm)

is 0.48 dB/mm @ 90 GHz. More details of the transmission line characteristics are covered in Section 2.8.3.

The global DC ground plane is formed from a via-connected stack of the top three metals AM, LY, and MQ. (MQ is also the global RF ground plane metal, and the global RF ground plane is connected to the global DC ground planes.) Most of the PA area, active and passive, is covered in the stacked grounding metals except for cut-outs necessary for the active and passive structures. The extent of the global DC & RF grounding can be in the AM top-metal coverage in Fig. 2.45.

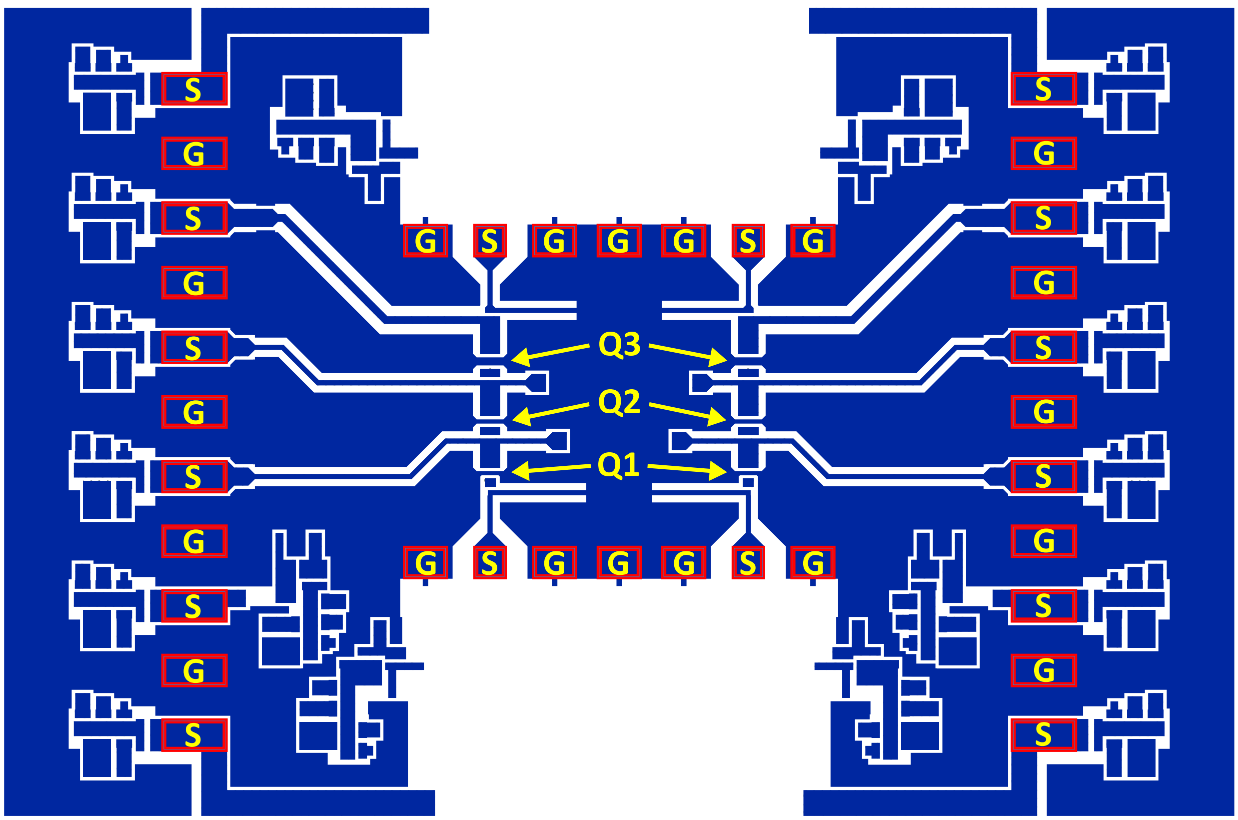


Figure 2.45: Top-level view of AM metal across the W-band differential PA with DC & RF I/O pads annotated.

Fig. 2.46 highlights all the unit-cell transmission lines used through the half-PA. Transmission lines of the same unit cell have the same cross-section, but not necessarily length. The quarter-wavelength base-biasing lines (in green) are the same length; the two quarter-wavelength collector-biasing lines for Q1 and Q2 are also the same length. The 16–12–16 μm (12 μm wide, 16 μm spacing to side shields) quarter wavelength collector-biasing line for Q3 breaks from the other unit-cells in order to be strictly compliant with electromigration guidance. It has similar transmission-line characteristics, however, as the 11–8–11 μm unit

cell (in yellow) used throughout the rest of the PA. The wide interconnects between each gain stage are also unit-cells with 11–32–11 µm cross sections that look similar to the 11–8–11 µm unit-cell. The lengths of the interconnects are different.

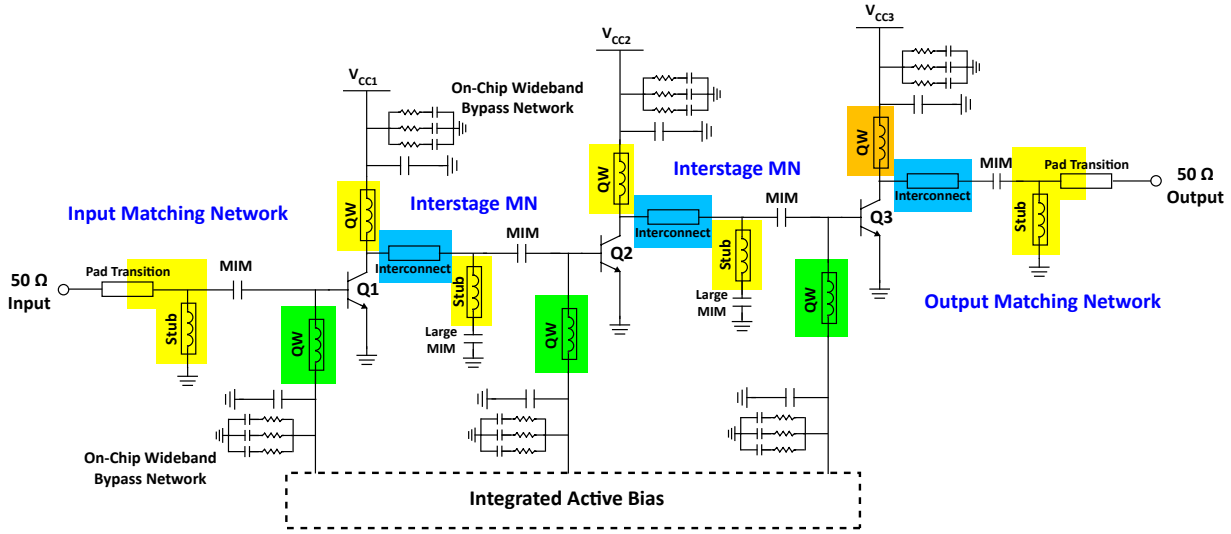


Figure 2.46: Half-PA simplified circuit schematic with all unit-cell transmission lines highlighted. 11–8–11 µm unit-cell transmission lines are in yellow, 11–32–11 µm unit-cell interstage interconnects are in blue, M_{234} stacked-copper inverted-microstrip unit-cell transmission lines for the base-biasing insertion networks are in green, and the Q3 16–12–16 µm collector feed transmission line is in orange.

2.8.2 Transmission-Line Inductors

Distributed transmission-line inductors are used throughout the W-band PA rather than small spiral inductors because large-valued (> 300 pH) spiral inductors have self-resonant frequencies well below W-band. High-Q inductors at 90 GHz can be synthesized from short-length ($l < \lambda_g/4$) transmission lines terminated on one end with an RF short-circuit.

The general expression for the complex input impedance Z_{IN} of a lossy transmission line (with characteristic impedance Z_0 terminated with complex load Z_L is given in Eq. (2.25) [104]. The Z_{IN} in Eq. (2.25) applies to the W-band PA interstage matching-network inductors, which are short transmission-line sections terminated in a 440 fF MIM capacitor that is self-resonant at 90 GHz (RF-short circuit) with some residual series resistance.

$$Z_{IN}(l) = Z_0 \frac{Z_L + Z_0 \tanh \gamma l}{Z_0 + Z_L \tanh \gamma l} \tag{2.25}$$

where γ is the complex propagation constant of the transmission line, given in Eq. (2.26).

$$\gamma = \alpha + j\beta = \sqrt{(R + j\omega L)(G + j\omega C)} \tag{2.26}$$

α and β make up the two components of the transmission-line complex propagation constant γ. β is the phase constant and α is the attenuation constant. R is the series resistance per unit length, L is the series inductance per unit length, G is the shunt conductance per unit length, and C is the shunt capacitance per unit length. The characteristic impedance Z_0 is related to the R, L, G, and C by Eq. (2.27).

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \tag{2.27}$$

Like Z_0, both α and β can be characterized over frequency for a given transmission-line structure. For low-loss lines ($R \ll \omega L$) with small dielectric losses ($G \approx 0$), α is simplified to Eq. (2.28) and β to Eq. (2.29).

$$\alpha \backsimeq \frac{1}{2}\left(R\sqrt{\frac{C}{L}}\right) = \frac{R}{2Z_0} \tag{2.28}$$

$$\beta \backsimeq \omega\sqrt{LC} \tag{2.29}$$

For the simplified case of a *lossless* line, the input impedance from Eq. (2.25) simplifies to Eq. (2.30).

$$Z_{IN,\text{lossless}}(l) = Z_0\frac{Z_L + jZ_0\tan\beta l}{Z_0 + jZ_L\tan\beta l} \tag{2.30}$$

When the transmission-line is terminated in a short circuit, $Z_L = 0$. The input impedance of a short-circuited, lossy transmission-line stub with length l is given in Eq. (2.31) [105]. The expression takes the transmission-line's characteristic loss α into account.

$$\begin{aligned} Z_{IN,\text{SC}}(l) &= Z_0\tanh[(\alpha + j\beta)l] \\ &= Z_0\frac{\tanh\alpha l + j\tan\beta l}{1 + j\tanh\alpha l\tan\beta l} \\ &= Z_0\frac{1 - j\tanh(\alpha l)\cot(\beta l)}{\tanh(\alpha l) - j\cot(\beta l)} \end{aligned} \tag{2.31}$$

where Z_0 is the characteristic impedance of the transmission line. β is the phase constant and α is the attenuation constant from Eq.(2.26).

Ignoring the losses (the real part of Z_{IN}) of the short-circuited transmission line, the input impedance of a short-circuited, lossless transmission line simplified to $Z_{IN,\text{lossless,SC}}$ in Eq. (2.32) [106].

$$Z_{IN,\text{lossless,SC}}(l) = jZ_0 \tan(\beta l) \tag{2.32}$$

A simplified equivalent inductance for a shunt short-circuited transmission-line stub of length l is given in Eq. (2.33) [107].

$$L_{eq}(l) = \frac{X_{in}}{\omega} = \frac{Z_0}{\omega} \tan(\beta l) \tag{2.33}$$

RF Ground Plane

Two types of ground plane structures are used throughout the PA chip for the transmission lines. Top-level AM transmission lines use pre-slotted ground plane structures (on the MQ layer) in order to prevent automatic cheesing (hole-punching) algorithms on the copper MQ metal for local-density compliance. For the custom M_{234} inverted microstrip transmission lines used for the base-biasing quarter-wavelength lines, a solid ground plane structure is used for higher density shielding to the top-level AM structure above. The final MQ copper local-density above the M_{234} transmission line will be similar to that of the pre-slotted MQ ground plane, but with smaller openings/holes more random in nature than the pre-slotted, parallel cutouts. This results in better overall shielding for the M_{234} line.

All transmission lines throughout the PA are EM-modeled for more accuracy in their specific use cases and terminations. Fig. 2.47 compares the EM-simulations results of an 11–8–11 µm (8 µm width, 11 µm side spacing) shorted shunt-stub AM transmission-line element with solid, contiguous MQ ground plane versus one with a pre-slotted MQ ground plane.

The EM simulation & modeling (via Keysight ADS Momentum) of transmission lines with solid MQ ground planes does not include the effects of the required patterned cheesing and filling of copper metals layers. The cheese (and fill, if there are voids) is unknown, precisely, to the user, so exact modeling is not possible.

The transmission line with pre-slotted ground plane has less loss, but is much more difficult to simulate at junctions and discontinuities. The transmission line with solid, contiguous MQ ground is far simpler to simulate, but the losses are slightly higher, and the contiguous MQ ground plane will be automatically cheesed with small holes punched into structure to comply with local-copper density requirements.

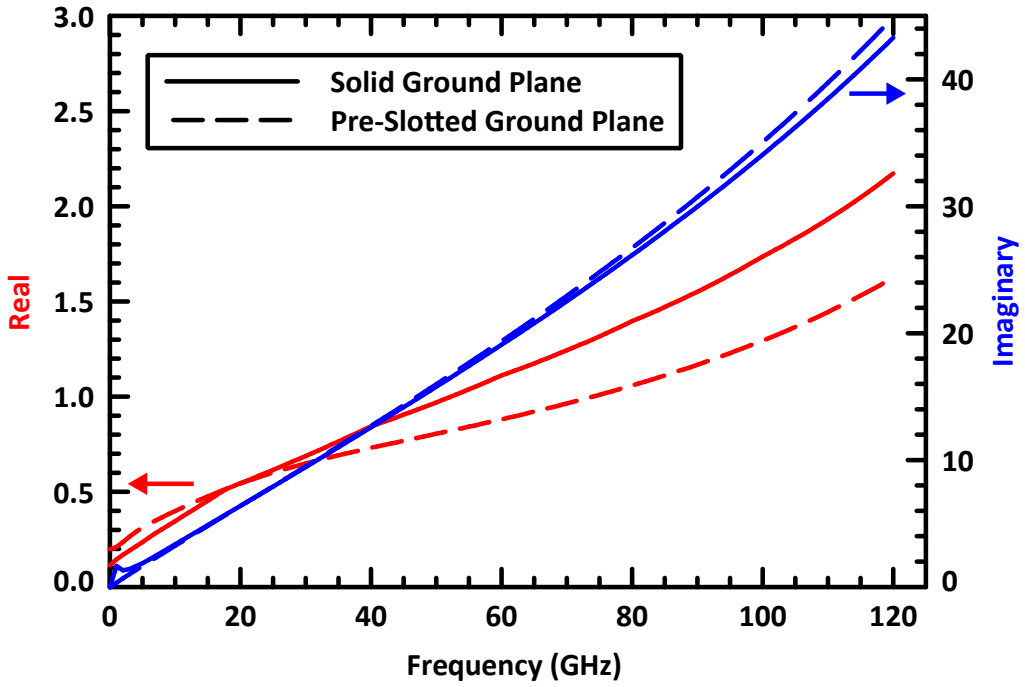


Figure 2.47: Simulated complex input impedance of a 135 μm AM, top-metal 11–8–11 μm transmission line segment (with MQ ground plane) terminated in a short-circuit on one end: real (red) and imaginary (blue) components of impedance comparing effects of a solid MQ ground plane versus a pre-slotted MQ ground plane.

Table 2.3: Inductance synthesized from shunt-stub, shorted transmission-lines using 11–8–11 μm unit-cell AM transmission lines. Reactance, equivalent inductance L_{eq}, and Q are given at 90 GHz.

Stage	**Length**	**Reactance**	$\mathbf{L_{eq}}$	Q
Input MN	135 μm	$+j28\ \Omega$	50 pH	57
Q1–Q2 Interstage	60 μm	$+j12\ \Omega$	20 pH	15
Q2–Q3 Interstage	30 μm	$+j6\ \Omega$	10 pH	8
Output MN	110 μm	$+j22\ \Omega$	40 pH	60

The PA design uses the pre-slotted ground plane structure for the input & output transmission line stubs, input & output transmission lines, quarter-wavelength collector lines, and interstage matching transmission-line stubs (terminated in resonant MIM capacitors for RF-short circuit). The solid MQ ground plane structure is used at junctions where multiple transmission-lines meet and above the custom M_{234} inverted microstrip line that runs beneath the RF ground plane (MQ).

All transmission line stubs use the same 11–8–11 μm transmissions line structure with characteristics listed later in Table 2.4 with α_g = 55.3 Np/m, β_g = 3900 rad/m, and Z_0 = 50 Ω at 90 GHz.

A summary of the shorted transmission-line stub structures is presented in Table 2.3. The input-matching network uses a 135 μm shunt element, shorted transmission-line stub with slotted ground plane. The shorted transmission-line stub presents an equivalent 50 pH ($+j28\ \Omega$) at 90 GHz with Q = 57. The output-matching network has a similar 110 μm shunt, shorted stub that presents an equivalent 40 pH inductance ($+j22\ \Omega$) at 90 GHz with Q = 60.

The Q1–Q2 interstage matching network uses a shunt 60 μm transmission-line section terminated with a resonant 440 fF MIM capacitor. The shorted transmission-line presents an equivalent 20 pH inductance ($+j12\ \Omega$) with Q = 15 when RF-shorted with the relatively low-Q 440 fF resonant MIM capacitor (Q = 7). The Q2–Q3 interstage matching network uses a shunt 30 μm transmission-line section terminated with a resonant 440 fF MIM capacitor. The shorted transmission-line is presents an equivalent 10 pH inductance ($+j6\ \Omega$) with Q = 8 when RF-shorted with the MIM capacitor (Q = 7).

See Appendix E for more discussions regarding metal-density limitations and their effects on the design of passive structures.

2.8.3 Quarter-Wavelength Collector Feeds

Compact, large-valued spiral inductors are not suitable for inductive RF-biasing chokes at W-band because of their low self-resonance frequencies limited by parasitic capacitances. (An RF-impedance of 170 Ω requires a 300 pH inductance at 90 GHz, for instance.) The collector bias-insert networks also require wide linewidths for electromigration current compliance, which results in additional parasitic capacitances that lower self-resonance even further.

All top-level AM transmission lines have a common characteristic RF impedance ∼50 Ω in order to follow a unit-cell approach and track together with process variations and residual modeling errors. Except for the Q3 collector feed line, the same cross-sectional dimensions are used (line width and spacing to side and bottom shields) throughout the design where distributed transmission lines are used. For the Q3 collector transmission line, which must carry more DC quiescent current when there is no RF input drive, the line width is increased to satisfy electromigration constraints for 100°C operation, but the transmission-line spacing to the side shields is adjusted so that the transmission-line characteristic impedance is kept nearly the same as the other unit-cells.

Separate quarter-wavelength (QW[21] or $\lambda_g/4$) transmission lines feed the HBT collectors of each PA stage. Although quarter-wavelength collector-bias feedlines, $\sim 400\mu$m in length, occupy significant PA area, they are low-loss, high-bandwidth distributed elements that present the frequency-dependent impedances required for Class AB PA operation: low resistance at DC, high impedance at $f_0 = 90$ GHz, and low impedance at the second harmonic $2f_0 = 180$ GHz.

The short-circuited quarter-wavelength transmission-line resonator inserts DC at the HBT collector node while presenting a high-impedance RF at the operating W-band frequencies, providing a low-loss broadband RF inductive choke, equivalently. The collector transmission-line is a quarter-wavelength structure at 90 GHz, and each Q1, Q2, and Q3 power-HBT collector has a separate quarter-wavelength line.

One end of the quarter-wavelength transmission-line terminates at the HBT collector while the other terminates near the DC wirebond probe pad with a low RF-impedance (high capacitance), creating a quarter-wavelength resonator within W-band. The broadband RF-short termination is created using a resonant 440 fF MIM capacitor and a wideband bypass capacitor array, discussed in detail in Section 2.7. The wideband capacitor array presents an impedance < 1 Ω within W-band and < 3 Ω from 10–200 GHz covering the necessary second-harmonic frequencies for Class AB operation. External DC voltage is applied by a custom probe with additional capacitance for low-frequency power-supply off-chip filtering.

[21] $\lambda_g = \lambda_{\text{guided}} = \lambda_{\text{free-space}}/\sqrt{\epsilon_r}$

Quarter-Wavelength Resonators

Transmission-lines, with their distributed effects across frequency and distance, provide the necessary quarter-wavelength resonator (with a supporting wideband bypass-capacitor array) bandpass structures for proper low-impedance harmonic terminations above W-band and low-impedance terminations below W-band for amplifier stability. The quarter-wavelength transmission lines are used in conjunction with resonant 440 fF RF-shorting MIM capacitors and wideband bypass capacitor arrays to form resonators that are used for DC bias insertion. (Quarter-wavelength resonators terminated in MIM capacitors are also used for the base-biasing insertion networks in all three PA gain stages, but the transmission-line is very different.)

Eq. (2.34) gives the input impedance of a lossy quarter-wave transmission-line with one end terminated in a perfect short— a quarter-wavelength transmission-line resonator [105].

$$Z_{in} = R = \frac{Z_0}{\alpha l} \tag{2.34}$$

The associated unloaded quality factor Q of the lossy quarter-wavelength[22] transmission-line resonator (shorted) is given in Eq. (2.35).

$$Q_{\mathrm{QW,lossy}} = \frac{\pi}{4\alpha_g l} = \frac{\beta_g}{2\alpha_g} \tag{2.35}$$

Quarter-wavelength resonators are used as shunt-element, bias-insertion networks at the base and collector HBT terminals, offering high bandwidth and RF impedances within W-band while simultaneously injecting DC with very low loss ($< 0.6\ \Omega$). For quarter-wavelength lines terminated in short circuits (QW resonators), the input impedance looking into the shunt resonator is high. Quarter-wavelength transmission-lines resonators with higher ratios of characteristic impedances to the HBT output-collector node impedance lead to wider high-impedance bandwidths, as illustrated in Fig. 2.48.

Higher characteristic impedance lines (Z_0) can result in higher shunt input impedances in quarter-wavelength resonators, leading to less shunt loss and higher bandwidths, but for a given metal & ground plane configuration within the stack-up, linewidths need to be narrower to increase Z_0, which result in higher characteristic line loss α as well as reduced current handling. The Q3 collector top-metal AM transmission line width is primarily dictated by DC current handling for electromigration while all other top-metal AM transmission lines, carrying less DC current, are sized for $Z_0 = 50\ \Omega$.

The quarter-wavelength bias-insertion lines are shunt elements with respect to the RF

[22]QW is the abbreviation for the quarter-wavelength ($\lambda_g/4$) transmission-line.

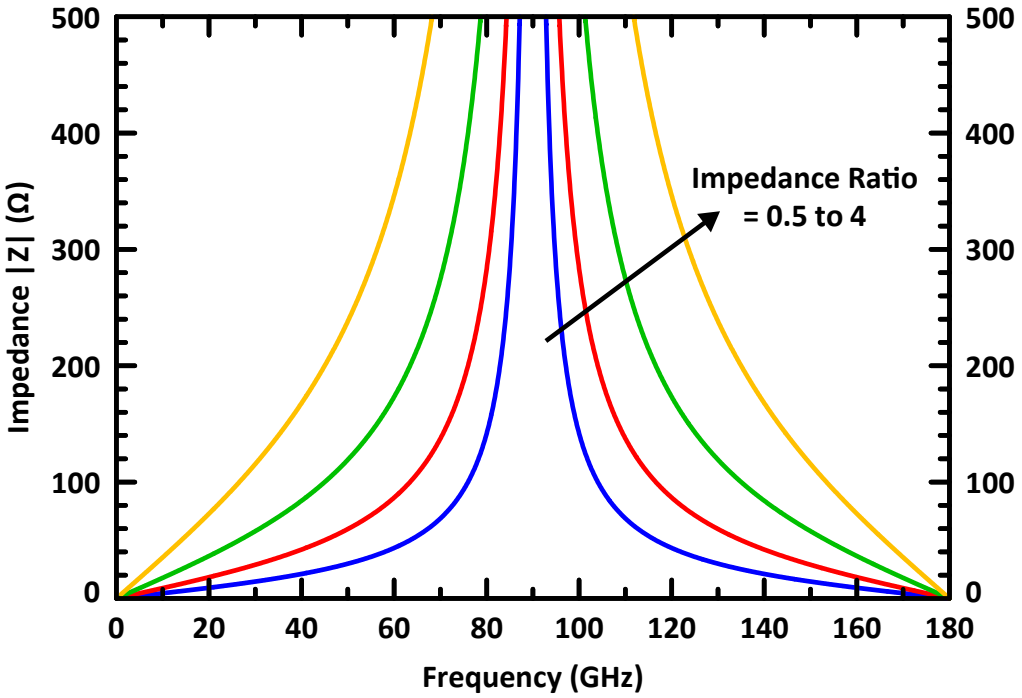


Figure 2.48: Impedance magnitude of an ideal quarter-wavelength transmission line resonator terminated in short-circuit for different impedance ratios $Z_{0,\text{line}}/Z_{\text{node}} = 0.5$, 1, 2, and 4.

Table 2.4: Q1 and Q2 collector bias-insertion transmission-line characteristics at 90 GHz: 8 μm width, 11 μm spacing to side shields (11–8–11 μm).

Parameter	**Value @ 90 GHz**
Z_0	50 Ω
Return Loss	> 35 dB (DC–220 GHz)
α_g	55.3 Np/m = 0.48 dB/mm
β_g	3900 rad/m = 0.224°/μm
ϵ_{eff}	4.3
λ_g	1610 μm
$\lambda_g/4$ Length	402 μm
$\lambda_g/4$ Series Insertion Loss	0.2 dB
$\lambda_g/4$ Resonator Q	35
$\lambda_g/4$ R_{DC}	0.3 Ω (DC)

signal path, so the associated loss within as a shunt resonator element is less than that of a series power-splitting/combining line. The PA's RF-signal loss attributed to the shunt quarter-wavelength transmission line (as a DC-bias-insertion feed) is less than 0.1 dB at each collector output of Q1, Q2, and Q3 at W-band frequencies. Shunt short-circuited quarter-wavelength resonant structures are more forgiving and result in high-bandwidths because they do not need to present an exact impedance, only a high impedance relative to the very low-impedance (5–47 Ω) bias-insertion nodes at the HBT terminals.

Q1/Q2 Collector Transmission-Line Feed

The 11–8–11 μm transmission-line is the unit-cell used throughout the W-band PA. It is an 8 μm-wide microstrip line on top-level AM metal with 11 μm separation to side shields and 9.25 μm separation above the RF ground plane (layer MQ). The cross-sectional structure is used for all transmission lines except the Q3 collector (which requires a wider line) and custom M_{234} base feeds. All transmission lines are fully characterized in EM simulations. Bends, junctions, step discontinuities, and other effects are taken into account with full EM modeling of the associated layouts. Table 2.4 summaries the 11–8–11 μm unit-cell transmission-line with additional entries for the quarter-wavelength collector characteristics.

The input impedance looking into the Q1 and Q2 quarter-wavelength-resonator collector

feeds is > 1400 Ω at 90 GHz and > 200 Ω from 78–108 GHz. The wideband high impedance ensures that the collector DC bias-insertion network does not load the RF output of the HBTs. The transmission-line resonator also terminates the second harmonic $2f_0$ in < 3Ω around 180 GHz, which in combination with the base-feed transmission-line resonator response, results in < 2 Ω 2nd-harmonic terminations for Class AB amplifier operation.

Q3 Collector Transmission-Line Feed

The Q3 collector-feed transmission line is a 12 μm-width top-level-metal AM line over MQ ground plane with 16 μm spacing to AM-sidewall shielding (16–12–16 μm) at 90 GHz. The transmission-line characteristics are summarized in Table 2.5. The transmission line has loss characteristics of α_g = 59.7 Np/m = 0.52 dB/mm, which corresponds[23] to a line loss of 0.52 dB/mm. The transmission-line β_g = 3900 rad/m = 0.224°/μm, which corresponds to a guided wavelength λ_g = 1610 μm and $\epsilon_{r,\text{eff}}$ = 4.3. The RF characteristic impedance is Z_0 = 48 Ω. The return loss is > 36 dB from DC to 200 GHz (to the W-band second harmonic).

The resulting quarter-wave transmission-line length is 402 μm at 90 GHz with an associated resonator Q = 33. The series insertion loss of the 402 μm is 0.2 dB at 90 GHz[24]. The DC series resistance of the quarter-wavelength 402 μm line is 0.4 Ω.

The Q3 HBT collector-feed is equivalent to a quarter-wavelength transmission-line resonator terminated in a non-ideal short provided by the lumped-element self-resonant (at 90 GHz) 440 fF MIM capacitor and wideband bypass-capacitor array. The quality factor of the quarter-wavelength transmission line resonator terminated in an ideal RF short is Q = 33 @ 90 GHz. In the practical case where the RF termination is not a perfect short at W-band, Q = 24 @ 90 GHz. Although the resonant 440 fF MIM capacitor terminating the quarter-wave resonators has a Q of only ∼7, the combined resonant structure still has a Q ∼ 24. Effectively, the resonant capacitor is in series with the quarter-wave resonator—not in parallel—, so the residual real, small resistance terminates one end of the quarter-wave line in a low resistance that gets transformed to a high impedance (400–1000 Ω) at the bias insertion point near the HBTs.

At 90 GHz, the collector's combined quarter-wave transmission line and terminating MIM structure (RF-short) effectively look like an extremely large inductor or RF choke. Fig. 2.49 shows the EM-simulated results of the input-impedance magnitudes of the Q1, Q2, and Q3 quarter-wavelength collector transmission lines terminated in a resonant 440 fF MIM capacitor and accompanying wideband bypass capacitor array. These are the effective

[23] 1 Np = 1 Nepers ≈ 8.69 dB.

[24] Note that the transmission-line is used as a shunt resonant element in the PA and not a series line with respect to the RF signal path.

Table 2.5: Q3 collector bias-insertion transmission-line characteristics at 90 GHz: 12 μm width, 16 μm spacing to side shields (16–12–16 μm).

Parameter	**Value @ 90 GHz**
Z_0	48 Ω
Return Loss	> 35 dB (DC–220 GHz)
α_g	59.7 Np/m = 0.52 dB/mm
β_g	3900 rad/m = 0.224°/μm
ϵ_{eff}	4.3
λ_g	1610 μm
$\lambda_g/4$ Length	402 μm
$\lambda_g/4$ Series Insertion Loss	0.2 dB
$\lambda_g/4$ Resonator Q	33
$\lambda_g/4$ R_{DC}	0.2 Ω (DC)

impedance seen by the HBT collector terminals.

The input impedance looking into the Q3 quarter-wavelength collector feed is > 1000 Ω at 90 GHz and > 190 Ω from 80–104 GHz. The Q3 transmission-line resonator also terminates the second harmonic $2f_0$ in < 2Ω around 180 GHz. The input impedance looking into the Q1 and Q2 quarter-wavelength-resonator collector feeds is > 1400 Ω at 90 GHz and > 200 Ω from 78–108 GHz.

The impedance responses are slightly higher than 90 GHz because additional parasitic wiring inductances (to the HBT collector-output interconnects) are not included. The implemented physical transmission-line lengths are slightly shorter (10–20 μm) than a full quarter-wavelength in order to account for additional wiring and via inductances to interface the HBT collectors on one end and the 440 fF MIM RF-shorting capacitor on the other end.

Although wider lines can provide even higher-Q quarter-wave resonator feeds, further increasing line width does not have appreciable effects on the collector HBT losses since the bias-insertion network is a shunt, resonant, high-impedance element rather than a series element.

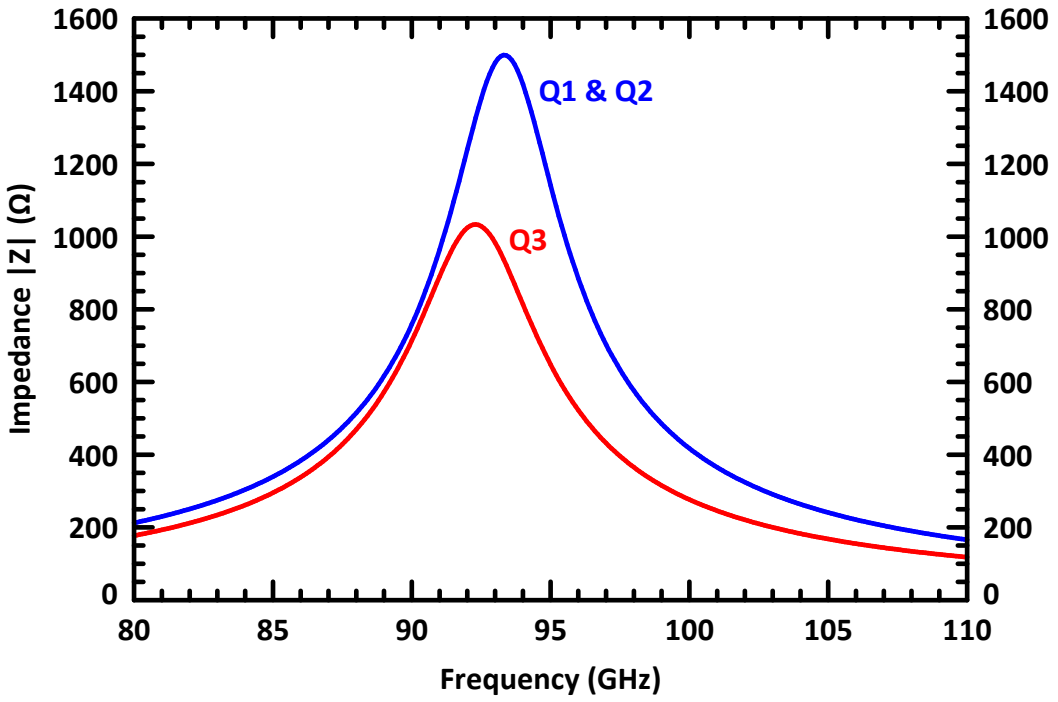


Figure 2.49: Input impedance magnitude of the Q1 & Q2 11–8–11 μm quarter-wavelength transmission-line resonator (in blue) and the Q3 16–12–16 μm quarter-wavelength transmission-line resonator (in red). All resonators are terminated on one end by a 440 fF MIM capacitor and a wideband bypass capacitor array.

2.8.4 Quarter-Wave Base Feeds

The bias insertion network is designed to extend the collector-to-emitter breakdown voltage from $BV_{\mathrm{CEO}} = 1.8$ V closer to the fundamental $BV_{\mathrm{CBO}} = 6.0$ V limits. As described in Section 2.6.2, the extension in operational breakdown voltage BV_{CER} can be accomplished by ballasting the excess DC base current caused by the impact ionization (weak avalanche) process.

The approach used in this work is to use a quarter-wavelength transmission line resonator in the external base-biasing insertion network. The quarter-wavelength bias-insertion network acts as a wideband RF inductive choke, which presents low impedance at DC and high impedance at RF (W-band) at the bias insertion point at the HBT power-cell base. This design provides a very low DC impedance without affecting the RF. This methodology also improves low frequency stability of the transistor. A custom transmission-line structure using stacked lower-level M2, M3, and M4 metals that share the RF ground plane *above* it with the top-level AM-metal transmission-line structures.

HBT External Bias Resistance

Base-biasing feeds in sub-mmW, small-signal bipolar amplifier designs are typically implemented with resistors or resistor networks that present multi-KΩ values to the HBT base terminal— the external DC base resistance $R_{\mathrm{b,bias}} > 1000\ \Omega$. With large power transistors ICs at mmW frequencies, the $R_{\mathrm{B,bias}}$ can be lowered further since HBT input impedances drop significantly. At 90 GHz, 8HP HBTs input impedances are $< 30\ \Omega$ even for relatively small (10 μm) transistors operating at peak-f_T current densities. Lowering the external base resistance $R_{\mathrm{B,bias}}$ is possible in order to push operational collector breakdown voltages BV_{CER} from the open-base case ($R_{\mathrm{B,bias}} \rightarrow \infty$, $BV_{\mathrm{CEO}} = 1.8$ V) toward $BV_{\mathrm{CBO}} = 6.0$ V, as discussed in Section 2.6.2.

Prior works have demonstrated base biasing with resistors $R_{\mathrm{B,bias}} > 300\ \Omega$ to increase operation BV_{CER} [14]. In the presented W-band PA, one of the main goals is to push operation collector breakdown voltages even higher by using extremely low external-base resistances presented to the HBT base terminal.

The W-band PA here pushes operational collector voltages even higher by reducing the external base resistance down to $R_{\mathrm{B,bias}} = 20\ \Omega$, the lowest-known resistance used in SiGe mmW PAs. Using such low DC resistances pushes operational collector voltages higher, but since 20 Ω is also on the order of the HBT input impedances, transmission-line bias-insertion structures in the base feedlines are necessary to reduce the RF-loading impacts.

W-band frequency is too high for large-valued spiral inductors as RF biasing chokes

(e.g., 300 pH for 170 Ω impedance at 90 GHz) because the parasitic capacitances reduce the inductor resonant frequencies to well below W-band. Using a distributed transmission line in the base feeds allows a fixed, low DC resistance to be presented while maintaining a high impedance at the operating frequency.

The function of the base-biasing feed line is to supply the appropriate DC quiescent base current to the power-transistor cells while presenting a high impedance to RF signals at the base node to prevent injection of RF into the DC biasing circuitry. The ideal base biasing feed is an RF choke that present high-impedance to the RF signal and low-impedance to the DC bias current. Quarter-wavelength bias-insertion networks used in the HBT-base biasing allows for low external DC resistances presented to the HBT base terminal while maintaining high RF shunt impedances at the HBT input, which minimizes unwanted RF-loading. Lowering the external DC base resistances (e.g., from $>$ 1 KΩ down to 20 Ω) extends the operational collector-emitter breakdown voltage BV_{CER} from $BV_{\mathrm{CEO}} = 1.8$ V closer toward $BV_{\mathrm{CBO}} = 6.0$ V. All external base terminals in each gain stage have a < 20 Ω DC resistance and RF impedance from DC to W-band, ballasting excess holes in the base from DC and any large-signal carrier effects.

The quarter-wavelength transmission-line resonator, similar in operation to the collector ones discussed in Section 2.8.3, are wide-bandwidth bias-insertion structures that allow the DC resistance and RF impedance presented to the HBT base terminal to be set independently: low 20 Ω resistance (with added series resistance) at DC, 170 Ω RF impedance at 90 GHz, and low impedance at $2f_0 = 180$ GHz. Furthermore, the base-feed transmission-line resonator presents a low impedance at lower frequencies (outside the operating bandwidth) arising from filtering effects of the large wideband bypass capacitor arrays in parallel to the 440 fF RF-shorting capacitors. The low-impedance terminations on all three HBT terminals stabilize the HBT where its available gain is highest at sub-mmW frequencies.

In typical resistive biasing schemes, 300–2000 Ω resistors are used, typically, which result in both DC and RF impedances to be 300–2000 Ω, respectively, over wide bandwidths— DC to W-band to 2nd harmonic and beyond. With a distributed transmission-line structure terminated in a proper low-impedance, the DC resistance presented to the HBT base terminal can be set low (20 Ω) while the RF impedance at W-band remains high (170 Ω) relative to the node impedances ($<$ 5–20Ω). Furthermore, at the 2nd harmonic, the quarter-wavelength line at 90 GHz becomes a half-wavelength line at 180 GHz, transforming the low-impedance (near short-circuit) termination at W-band to a low-impedance termination at $2f_0$, which is used for harmonic termination control in Class AB PA operation.

The quarter-wavelength (at 90 GHz) base line is terminated in a resonant 440 fF MIM as RF-short and a wideband MIM capacitance array. A 15 Ω series TaN inductor further

de-couples the RF section from the active-biasing networks and also helps to set the 20 Ω total external base resistance. The active-bias circuits have further MIM bypass capacitances to ensure stability and reduce RF-signal injection. Below 15 GHz, the presented RF impedance is < 10 Ω, which along with the collector feeds and the high-pass matching networks, helps to stabilize the low-frequency transistor response and guarantee stability.

By decoupling the RF impedance and the $R_{\mathrm{B,bias}}$ DC resistance presented to the HBT base terminals, the $R_{\mathrm{B,bias}}$ DC-resistance does not affect the maximum available gain MAG but does improve the large-signal breakdown characteristics over 300 Ω external-resistance biasing.

$\mathrm{M_{234}}$ Line Characteristics

To support a differential PA layout— a symmetric layout that is mirrored about a center line and has all the DC & RF I/O on the peripheries—, a special transmission line is necessary to support a quarter-wavelength base-biasing feed & insertion network for each gain stage. Without a differential layout requirement, quarter-wavelength base-biasing feeds can simply be done on the top-level AM metal in the area where the mirrored half-PA sits in the differential layout. (Moreover, if collector voltages did not need to be pushed higher, simple high-resistance biasing could have been used).

The lack of many additional metals above the MQ ground plane necessitates use of lower-level metals *below* the MQ RF-ground plane for quarter-wavelength bias feeds. These transmission lines share the RF ground plane with the top-level transmission lines above. A custom, non-standard transmission-line structure is created by stacking— connecting vertically-adjacent metal layers together with arrays of vias— three of the lower copper-metal layers to form an inverted transmission line that sits below the global RF ground plane on layer MQ.

$\mathrm{M_{234}}$ refers to the quarter-wavelength base-feed transmission-line formed with a stack of M2, M3, and M4 lower copper metal layers (excluding M1) tied together electrically with densely-spaced vias between each metal layer. The $\mathrm{M_{234}}$ line is used in an inverted microstrip fashion with the global MQ ground plane sitting *above* it and shared, indirectly, with the top-level AM transmission line structures. Each HBT power-cell transistor has a separate $\mathrm{M_{234}}$ transmission line connecting to its HBT base terminal for DC bias insertion.

The copper-layer conductivity (M1, M2, M3, M4, and MQ) is 1.5× that of aluminum layers (LY & AM) with 5.35 ℧/m vs. 3.57 ℧/m), respectively. However, the lower M1–M4 copper metal layers are much thinner metals with 0.29–0.32 μm thicknesses or equivalently, only $1.5\delta_{s,\mathrm{Cu}}$ at 90 GHz[25], resulting in relatively high metal sheet resistances ($R_s = 0.058$ Ω/□)

[25]The skin depth of copper is $\delta_{s,\mathrm{Cu}} = 0.22$ μm at 90 GHz.

5–9× higher than that of the top-level AM aluminum ($R_s = 0.007\ \Omega/\square$).

To overcome the relatively high metal sheet resistance of the individual copper layers, 3 copper layers are stacked for improved RF losses over a single metal layer. The stacked copper transmission-line is formed with layers M2, M3, and M4, which is referred as the M_{234} base-feed transmission line. (The conductivity of the dense layer of vias between each copper layer is about 1/5th that of the copper metal layers.) The stacked metal configuration decreases the losses at 90 GHz and results in increased transmission-line Q at 90 GHz.

The cross-section of the implemented custom base-feed transmission line is illustrated in Fig. 2.50. The stacked M_{234} base-feed transmission lines are sized 2.24 μm wide, which is the maximum width for stacked copper wires to be exempt from foundry auto-cheesing[26]. The lowest-level M1 is not used in the M_{234} copper stack because of its close proximity (0.60 μm) to the lossy 13.5 Ω·cm SiGe p^- substrate. Avoiding M1 decreases the losses from RF currents induced in a lossy substrate and prevent unwanted coupling into structures within the substrate. Additional shallow-trench isolation (STI) in the substrate underneath the base transmission line provides an extra 0.35 μm of oxide isolation from the substrate, allowing the M_{234} structure to sit 22% higher above the silicon substrate at 1.59 μm. Deep-trench (DT) and shallow-trench lattice networks along the M_{234} signal path further reduce induced eddy-flow RF-currents at the substrate surface and further decreases coupling to other nearby structures.

The global RF copper ground plane sits 0.65 μm *above* the stacked M_{234} transmission line, competing the inverted microstrip-line structure. The M_{234} lines sit in SiO_2 dielectric with an effective dielectric constant $\epsilon_{r,\text{eff}} = 3.9$, supplied by the foundry, which takes into account extra copper-metal fill-effects throughout the dielectric layers.

The transmission-line characteristics are extracted from EM simulations and presented in Table. 2.6. The M_{234} inverted microstrip transmission line has a characteristic impedance of $Z_0 = 30\ \Omega$ within W-band. The EM-simulated return-loss of the transmission line in a $Z_0 = 30\ \Omega$ system is better than 22 dB from DC to 200 GHz ($2f_0$).

The M_{234} characteristics transmission-line loss is given by $\alpha_g = 416$ Np/m = 3.6 dB/mm @ 90 GHz. At 90 GHz, the transmission line has a characteristic $\beta_g = 4340$ rad/m = 0.249°/μm, which implies a guided quarter-wavelength of $\lambda_g/4 = 362$ μm. The corresponding effective dielectric constant (effective relative permittivity) is $\epsilon_{r,\text{eff}} = 5.3$. The effective dielectric constant $\epsilon_{r,\text{eff}}$ is higher than that of the surrounding SiO_2 because some fields are penetrating into the silicon substrate below ($\epsilon_{r,\text{eff}} = 11.7$ for Si), but most of the fields are

[26]Stacked lines greater than 2.24 μm can theoretically provide additional design freedoms, but in practice, they are difficult to model accurately in EM simulations because stacked lines > 2.24 μm are aggressively cheesed (reducing local density up to 30–50%) by the foundry mask-preparation algorithms to meet strict copper local-metal density requirements.

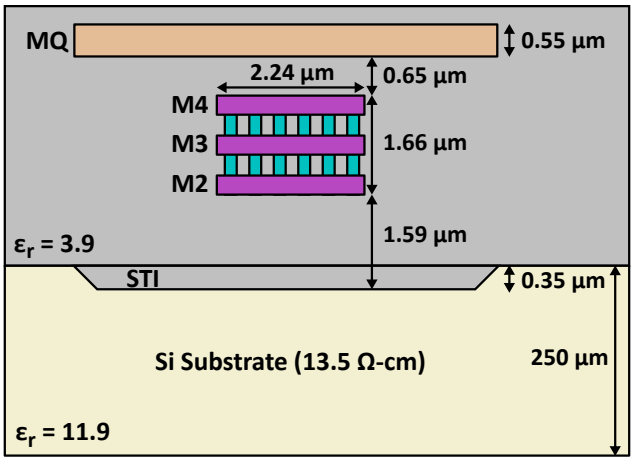


Figure 2.50: Cross section of the custom inverted microstrip transmission-line structure on stacked metal layers M2, M3, and M4. The stacked M_{234} line is internally connected with dense vias. The line runs beneath the global RF ground plane MQ.

Table 2.6: M_{234} base-feed transmission line characteristics.

Parameter	**Value @ 90 GHz**
Z_0	30 Ω
Return Loss	$>$ 22 dB (DC–220 GHz)
α_g	415.9 Np/m = 3.61 dB/mm
β_g	4340 rad/m = 0.240°/µm
ϵ_{eff}	5.3
λ_g	1448 µm
$\lambda_g/4$ Length	362 µm
$\lambda_g/4$ Series Insertion Loss	1.3 dB
$\lambda_g/4$ Resonator Q	5.2
$\lambda_g/4$ R_{DC}	2.4 Ω (DC)

still contained between the RF ground plane (above on layer MQ) and the stacked M_{234} copper lines. The slightly higher effective dielectric constant also increases the effective transmission-line phase constant β_g compared to top-level AM-metal transmission lines.

Fig. 2.51 plots the M_{234} transmission-line characteristics across W-band: α (loss), β

(phase), and quality factor Q.

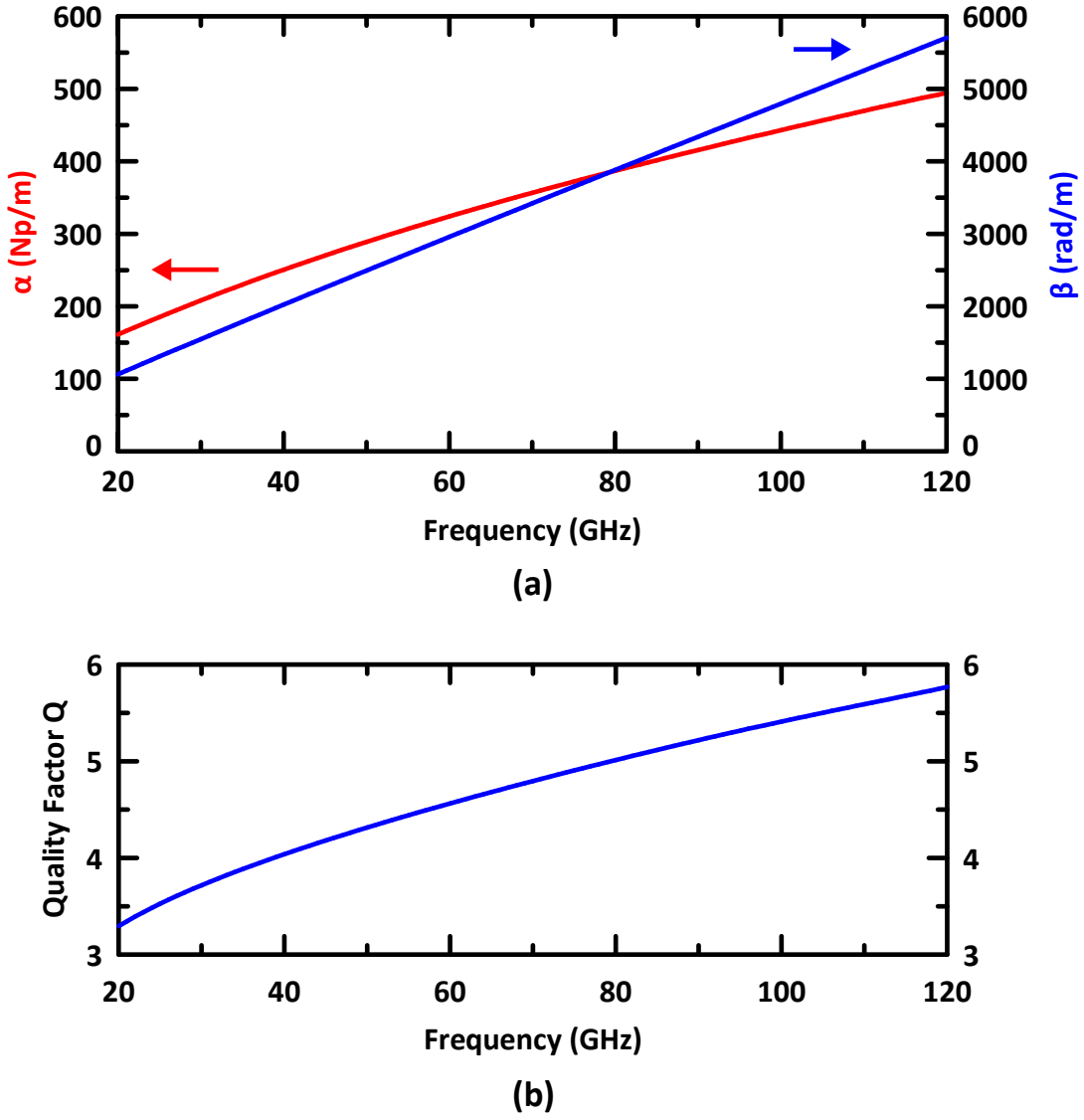


Figure 2.51: Inverted microstrip M_{234} transmission line (a) loss (α), phase (β) and (b) quality factor Q.

M_{234} RF Ground Plane and Isolation

Since the collector bias-insertion transmission-lines share the same RF ground plane on layer MQ as the M_{234} base-feed transmission lines, the signal coupling through the ground plane must be examined to ensure HBT feedback stability at all frequencies. The MQ ground plane is a 0.55 μm-thick copper metal that results in a sheet resistance $R_s = 0.034\ \Omega/\square$ and a skin depth of 0.23 μm at 90 GHz.

Rather than use pre-slotted MQ ground planes, as done with the top-level AM transmission lines, the stacked M_{234} lines use a solid MQ ground plane that is marked for automatic cheesing at the foundry. During the foundry mask generation, small holes are punched into the MQ ground plane to reduce the local density to 50% within any 200 μm × 200 μm window. The cheesed solid MQ ground plane still results in a denser MQ ground plane than pre-slotting metal.

The worst-case M_{234} to top-metal AM line isolation occurs if M_{234} runs directly underneath a collector transmission line and shares the same immediate ground plane metal on MQ. The EM-simulated isolation between the AM transmission line and M_{234} line in this case is still > 40 dB within W-band and > 55 dB at 30 GHz. The side shielding and side-walls of the top-level AM transmission line structure aid in reducing vertical field penetration into the RF ground plane underneath. However, for additional signal isolation across *all* frequencies, the collector lines and M_{234} are placed such that they do not run coincidentally— on top of one another. M_{234} lines only cross underneath top-level AM transmission lines in very small cross sections— almost orthogonally— to avoid coupling together through the RF ground plane.

Side shielding of the M_{234} lines is difficult to implement because of metal-density and other copper layout-rule constraints. Since base-feed lines are not side-shielded, they are kept at distances where line-to-line base-feed coupling between is negligible. The M_{234} lines come as close as 80 μm to each other, in the worst case, between the Q1 and Q2 base feeds. Even without lateral/side shielding, the EM-simulated isolation between the M_{234} lines is better than > 60 dB from DC to 220 GHz since the distance to the adjacent M_{234} is much larger than the 0.65 μm vertical separation from the MQ global ground plane.

A series of orthogonal deep-trench isolation lines, within the substrate, along the M_{234} transmission line axis helps to isolate any possible coupling from one gain stage to the next and reduces induced RF currents in the substrate caused by the base-biasing lines.

A 3-dimensional view of the W-band PA layout is shown in Fig. 2.52. The magenta base feeds are shown routed under the AM top-layer metal. MIM capacitors used for matching, resonant RF-shorts, and wideband bypass are shown in yellow-orange.

M_{234} Quarter-Wavelength Resonator

The 350 μm quarter-wavelength M_{234} transmission line has a characteristic impedance $Z_0 = 30\ \Omega$. When terminated on one end with a 440 fF MIM (RF-short) and wideband bypass capacitor array, the quarter-wavelength base-feed line becomes a transmission-line resonator. The resulting RF impedance at the bias-injection node at the HBT base is 120–170 Ω within the W-band operating frequencies 80–100 GHz while the DC series resistance is 2.4 Ω. Around the second harmonic $2f_0 = 180$ GHz the M_{234} line is effectively a half-wave

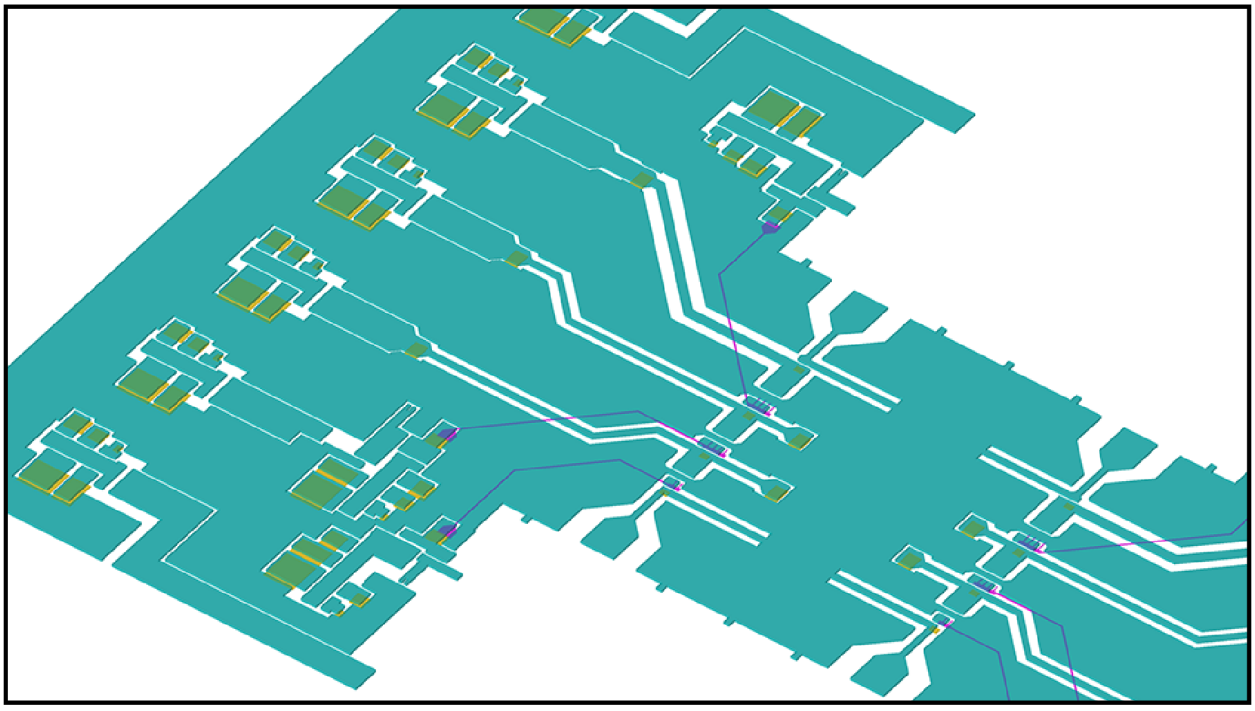


Figure 2.52: 3-D view of W-band PA layout. AM top-layer metal is shown in blue-green, base feeds shown in magenta, and MIM capacitors shown in yellow-orange.

structure, transforming the terminating wideband RF-short near the DC bondpad to a low impedance at the HBT. At 180 GHz, the M_{234} line presents impedances $< 7\ \Omega$, which in parallel combination with the collector-feed transmission-line resonators, present $< 2\ \Omega$ to the harmonics at $2f_0$ for proper Class AB harmonic terminations.

Additional 15 Ω TaN metal resistors and wiring resistances in the active bias circuits increase the DC resistance to the desired 20 Ω, the total external DC resistance presented to the HBT base terminal. The feed length is slightly shorter than the $\lambda_g/4 = 362\ \mu$m quarter-wavelength in order to account for additional wiring and via inductances at the connection points to the MIM capacitor at one end and the HBT base feeds at the other end.

The lower characteristic impedance Z_0 and higher RF losses α_g of the M_{234} custom transmission line results in lower RF input impedances when used in quarter-wavelength resonator, as shown earlier in Eq. (2.34). As the earlier Fig. 2.49 shows, the higher Q, higher Z_0 collector transmission-line resonator results in higher input impedances. The M_234 lower impedance magnitudes, however, are acceptable since HBT input impedances are very small, even smaller than their associated output impedances, and the base feed is a unique situation where the higher DC loss of the line is acceptable for HBT base biasing (compared to requirements in the collector biasing) since a 15 Ω series resistance is added within the

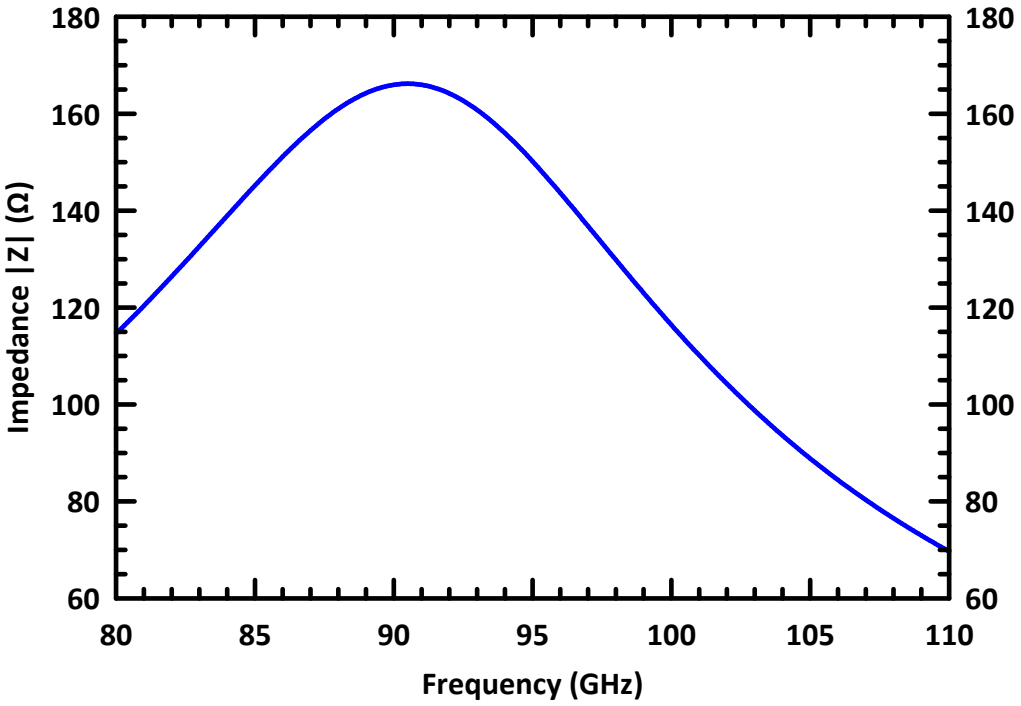


Figure 2.53: Input impedance magnitude of the 350 μm M_{234} base-feed quarter-wavelength transmission-line resonator. The transmission-line resonator is terminated on its other end with a 440 fF RF-shorting MIM capacitor and a wideband MIM bypass capacitor array. The other end connects to the HBT base terminal for DC bias injection.

active biasing circuit, anyway, to set the external base impedance to 20 Ω.

Using M_{234} as a quarter-wavelength resonator terminated in a non-ideal RF short-circuit results in a resonator $Q = 5.2$ at 90 GHz. Although the resonator Q of the base line is not high relative to other passive components, its effects are reduced because its resonant impedance is still high relative to the base-input node impedances.

The series insertion loss for 362 μm line is 1.3 dB at 90 GHz, but since the line is used as a shunt-element (resonator), the RF loss is much less. The associated loss of the shunt M_{234} base-insertion quarter-wave resonator, including wideband capacitor termination, is 0.50 dB at 90 GHz for Q1, 0.25 dB for Q2 stage, 0.13 dB for the Q3 stage. The M_{234} 120–170 Ω impedance is relatively large compared to the 20 Ω, 10 Ω, and 5 Ω input impedance of Q1, Q2, and Q3 biased at peak-f_T/f_{MAX} current densities.

Although 0.5 dB is not an insignificant loss, the penalty occurs at the input to the PA, which impacts gain, but not saturated the saturated output power. In future designs, a more optimal tradeoff (higher impedance or even lower loss line) can be synthesized to recover additional gain at the PA input. In this first W-band iteration, the design choice has been made to synthesize a custom line that is more predictable (via EM-simulation) over incremental performance improvements that are not as predictable (e.g., wider M_{234} lines that are automatically cheesed heavily (up to 50%) into local-density compliance, which are particularly difficult to simulate since the user does not have control over the cheesing algorithms).

2.8.5 RF Input/Output Pads

For the W-band design, small, custom RF input/output bondpads are designed to reduce pad parasitics at 90 GHz. (The earlier Ka-band designs use the larger standard wirebond pads measuring $W \times L = 61$ μm × 101 μm the RF input and output.) The custom W-band RF input/output pads are designed with 50 μm × 50 μm top-level metal (AM) bondpads for landing external 50 Ω measurement probes in ground-signal-ground (G–S–G) configuration at 100 μm probe pitch.

The CPW-to-microstrip RF-pad transitions are illustrated in Fig. 2.54. The 50 μm-wide on-chip RF-probe pads taper (over a distance of 21 μm) down to the 50 Ω input transmission line with 8 μm-wide signal conductor. The wirebond RF-ground pads also taper into the input-transmission-line side-shielding, separated 11 μm from the signal conductor. The EM-simulated return loss the back-to-back pad transitions is > 22 dB within W-band, and the associated back-to-back (total) losses of the input and output pads, combined, is < 0.2 dB in W-band.

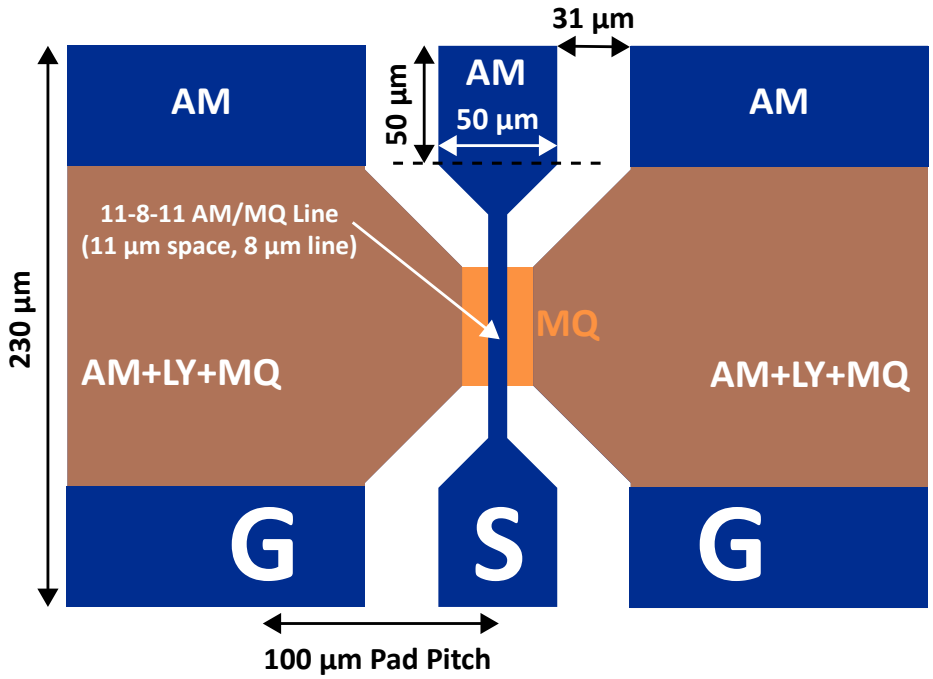


Figure 2.54: On-chip microstrip to external CPW G–S–G probe transition on top-level AM metal. A calibration thru structure is shown.

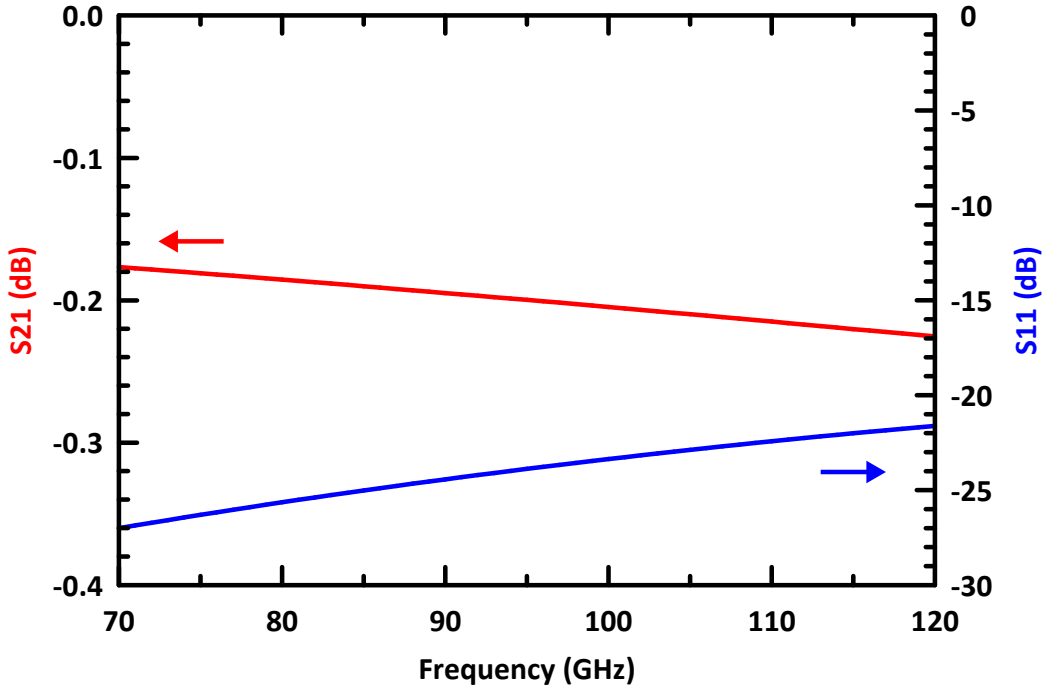


Figure 2.55: On-chip microstrip to CPW probe transition (back-to-back): insertion loss and return loss at W-band.

Special extra-long glass-cut openings, parallel to the G–S–G pad openings that open beyond the top pad metal, provide extra clearance for the angle-of-attack of the external G–S–G microprobes. The additional passivation openings aid in the reproducibility & precision of the measurements while reducing the probability of probe damage. The lower-angle of attack to the RF pads accommodates more consistent probing and probe overdrive– the amount of skid after first contact with the aluminum pad metal. Overall, measurements are more repeatable over many probings without wearing down or damaging the delicate microprobes.

The PA design absorbs the on-chip RF input and output pads into the input and output matching networks so that the PA has 50 Ω input and output impedances referenced to the probe pads. In future further-integrated designs with additional on-chip input drivers and output structures, the pads and their transitions can be removed, eliminating the associated 0.2 dB losses.

The shorted shunt-stub transmission-line matching-network elements at the immediate input and output of the PA provide shunt ESD paths to very large DC grounds, and no additional ESD-protection diodes are required in RF signal path. ESD on the RF inputs and outputs is possible, as demonstrated in [108, 109], but normally prohibitive because of the degradation in RF performance: additional RF loss and signal non-linearities with large signals.

The RF input/output pads are fully ESD compliant because of the shorted shunt-stub transmission-line matching-network elements at the immediate inputs and outputs of the PA that provide shunt ESD paths to very large DC ground. No additional ESD-protection diodes are required in RF signal path. Placing ESD diodes on the RF inputs and outputs is possible, as demonstrated in [108, 109], but normally prohibitive because of the degradation in RF performance: additional RF loss and signal non-linearities with large signals, especially at 90 GHz where available gain is low.

The ESD-protected RF I/O allows the PA to drive a load like an antenna, directly, without handling issues. In the current design, the DC bias pads are not fully ESD protected from all possible ESD paths on the chip, but dedicated ESD-diode clamps can be easily integrated at these DC bias pads in a future design iteration. The RF effects of the ESD clamps would be negligible to the RF the RF design because of the isolation provided by the quarter-wavelength resonator bias feeds.

2.8.6 Electromigration

One of the principles of the PA design is to adhere to reliability guidelines so that resulting designs are practical IC implementations. Reliability guidelines are especially significant in

high-powered designs because of the increased stresses of high RF currents & voltages, high DC currents, and large thermal dissipations.

The expected PA lifetime is set by electromigration constraints in the metals and not by aging or stressing of active devices within the substrate [110]. The top-level aluminum AM carries the highest DC currents on the PA chip and set the overall expected chip lifetime. Collector quarter-wavelength transmission-line widths are sized to comply with DC and RF electromigration guidelines for continuous 100,000 (11.4 years) power-on-hour operation at 100°C, which covers worst-case operating conditions for commercial electronics. Operation below 100°C results in much longer expected chip lifetimes. For example, at 25°C room temperature operation, the maximum allowed DC currents $I_{DC,\max}$ for a given width aluminium interconnect (AM layer) increases by 40× compared to the worst-case 100°C design target. This operational electromigration headroom results in reducing overall PA stresses.

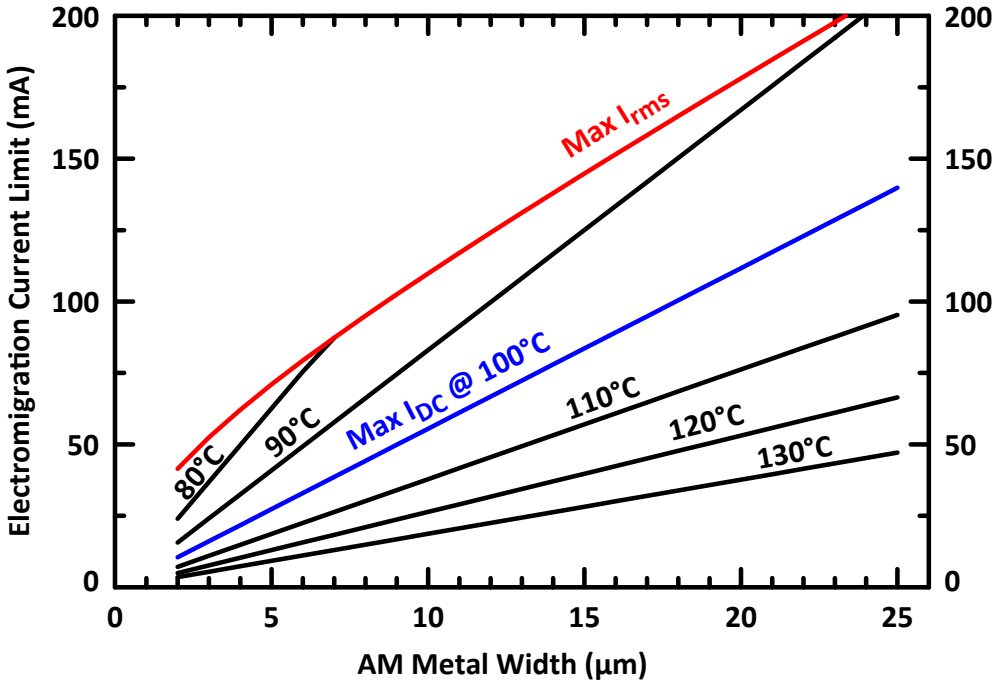


Figure 2.56: Calculated electromigration current limits for top-level AM metal at different operating temperatures. I_{DC} is the DC current on the wire. I_{rms} is the root-mean-squared AC current on the wire. Implemented designs target 100°C current limitations (in blue). Below 70°C, maximum DC and RMS currents must follow the $I_{rms,\max}$ limitations (in red).

Fig. 2.56 shows the electromigration current limits for the top-level (AM) aluminum at different line widths across chip operating temperatures [110]. Currents limits are dictated

by the line widths as well as temperatures of operation. Above 70°C, $I_{DC,\max}$ becomes the current limit, and minimum AM linewidths must be sized appropriately in high-current PA designs, particularly for the output collector. Below 70°C, the temperature-independent limit for current on a wire current is given by the maximum RMS AC current, $I_{\mathrm{rms,max}}$. Neither I_{rms} AC currents nor I_{DC} DC currents may exceed this current limit. AC RMS current limits are dictated by local heating effects from losses in the metal while DC electromigration is a function of both local heating from resistive losses as well as material drift from directional current flow.

The 8 μm-wide unit-cell AM-transmission line structure, used throughout the PA RF signal lines in everything but the Q3 collector line and the base feeds, has a 44 mA DC current limit at 100°C. The output power transistor's collector feed line is designed with an AM metal width of 12 μm which results in a 67 mA current limit. The Q3 DC collector current (half-PA) is 60 mA, nominally, and less at higher collector voltages because of deeper Class AB biasing for efficiency optimizations. The 3-stacked M_{234} base-feed inverted transmission line carries very low DC currents (< 400 μA), and the combined 2.24 μm-widths lines can handle DC currents up to 20 mA at 100°C.

2.9 Large-Signal Design

In this section, the details of the large-signal design are discussed, including the RF output matching network, interstage matching networks, input matching network and RF input/output wirebond pads & transitions. For simplicity, only half of the differential PA—single-ended schematic— is considered in the discussions since the mirrored half contains identical circuitry mirrored about a center DC & RF ground plane.

Figure 2.4 shows the simplified circuit diagram of the implemented baseline W-band PA design (design W1).

The large-signal designs are optimized around $V_{CC} = 1.7$ V in Class AB operation near P_{1dB} compression. Biasing voltage and current tuning provisions allow the PA large-signal operating point to be adjusted to higher collector voltages and deeper Class AB operation (reduced conduction angles). A unit-cell approach is applied at the PA system level by using and re-using key active and passives structures.

The lack of dedicated large-signal power-HBT RF models based on measured data of dedicated large power transistors within W-band, makes power amplifier design challenging since the compact, scalable SPICE-based device models struggle to model the transistor behavior at the current and voltage limits of the HBT. Accurate transistor RF-models are especially challenging to extract directly at W-band, and parameters are typically fitted from

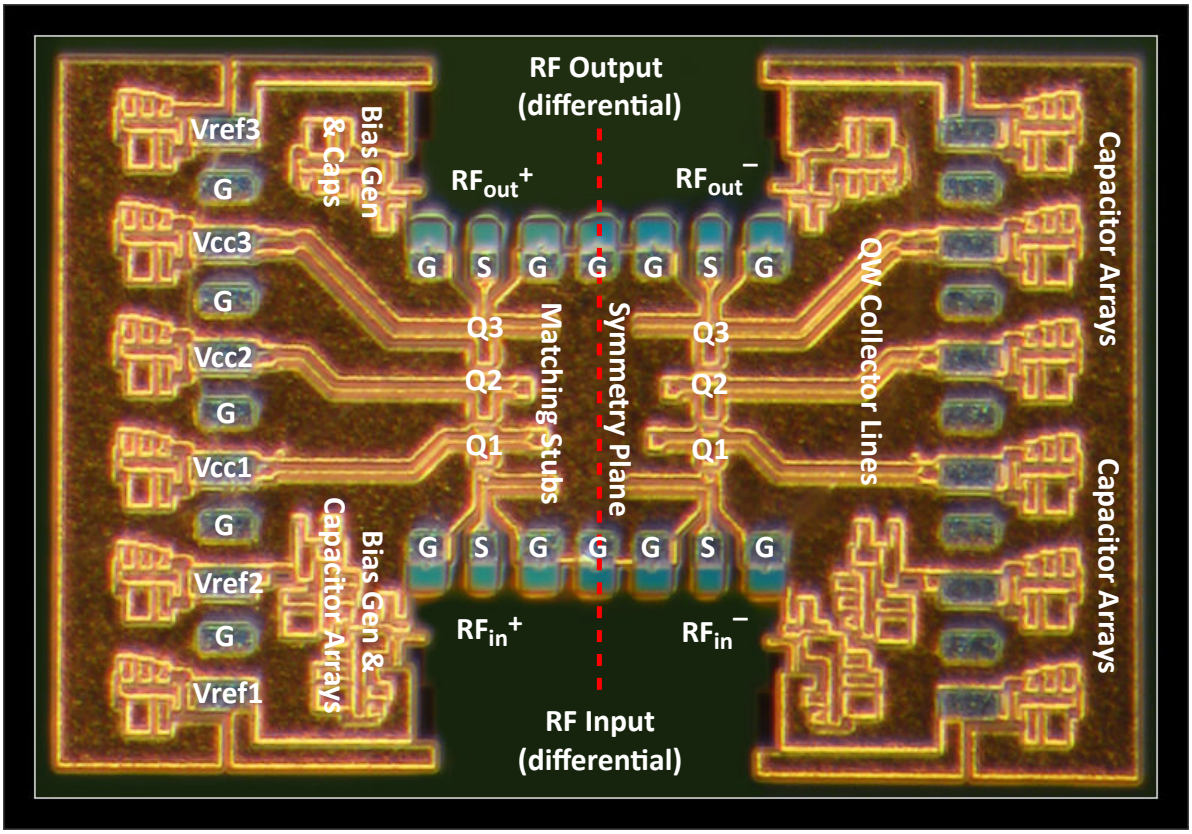


Figure 2.57: Chip micrograph of the balanced PA. The total area, including pads and large integrated MIM bypass capacitors arrays, is $1.9 \times 1.25 = 2.4$ mm^2. The right and left halves are identical and symmetric about the vertical center line.

a core RF-model extra below 30 GHz.

The baseline PA design is optimized for operation at 1.7 V in Class AB operation. Optimizing around higher collector-voltage operation was not possible using the older VBIC models at the time because of large-signal simulation convergence issues. The biasing provisions allow pushing the PA into higher voltage operation as well as deeper Class AB operation (reduced conduction angles with lowered quiescent DC bias currents), resulting in higher output powers and higher peak-PAE operation over nominal 1.7 V operation.

2.9.1 Large-Signal Design Approach

The W-band follows traditional Class AB, iterative large-signal design approach using large-signal harmonic balance simulations to determine optimal load and source impedances for high output powers and peak PAEs within each transistor stage with special focus on the output transistor Q3. Interstage matching networks must be iterated to find proper impedance-matching trade-offs between the optimal source impedance of one stage and the optimal load impedance of the preceding transistor stage.

Large-signal design incorporates aspects of traditional small-signal design, but takes into account large signal swings that push amplifiers into highly non-linear operating regions. Any PA design applying small-signal techniques, such as S-parameter-based matching, must also take into account power levels throughout the PA gain stages. From a simulation perspective, this means sweeping S-parameters over a range of input/output power levels throughout the design process and using large-signal harmonic balance techniques to capture effects of non-linearities caused by signal compression and distortion.

The three-stage PA consists of common-emitter transistors biased at quiescent currents resulting in Class AB, linear operation. Since no foundry power-transistor layout cells are available, a power transistor is implemented by aggregating smaller, high-f_T npn standard-cells.

Measured load-pull data is not available for the standard-cell transistors or aggregated power transistors, but the wideband implementation and independent bias tuning provisions in PA design allow re-centering the PA despite minor optimal-load mismatches, modeling inaccuracies, and process variations.

The HBT's changing S-parameters over RF input powers causes the matching networks to have input-power dependencies, as non-linear capacitances are very different at the quiescent DC bias point (small-signal levels) versus large-signal amplifier compression, where high RF input powers result in highly non-linear behavior and even shifting of HBT bias points. Circuits may be analyzed using a small-signal S-parameter approach only when an amplifier

operates purely in Class A, but for overdriven Class A (into HBT gain compression) and Class AB operation, large-signal analysis is necessary to set the proper RF matching network targets.

2.9.2 Gain

The HBT of each stage is sized such that headroom is left to compress each stage into gain compression for overdriven waveform clipping in either Class A or Class AB operation. Variable bias control allows earlier gain stages to bias closer to Class A while later stages bias deeper into Class AB for efficiency. Interstage, input, and output matching networks shape the gain function of the overall PA, and the Q1 and Q2 sizing ensure that the driver stages do not compress earlier than the output Q3 transistor.

The power transistors are biased near their peak-f_T/f_{MAX}/MAG current density at a quiescent current of 1.5 mA/μm (12.5 mA/μm^2). Each common-emitter stage provides 4–6 dB of small-signal gain, and each transistor stage is scaled by twice the area over the previous-stage transistor to ensure that the output transistor Q3 limits the large-signal compression characteristics. The first-stage power transistor Q1 consists of two aggregated 5 μm × 0.12 μm HBT unit cells while the second stage power transistor Q2 is made from 4 units. The output power transistor Q3 consists of 8 units cells connected in a dense two-dimensional array to form a 40 μm emitter-length device. Each 5 × 0.12 μm^2 unit-cell transistor has a single emitter finger, dual-base fingers, and dual-collector fingers (C–B–E–B–C). Within the power-transistor arrays, each unit-cell transistor is surrounded by deep-trench isolation structures. Low-loss, 30 μm-wide transmission lines connect the unit-cell collectors together and feed into the inter-stage matching networks.

More available gain leads allows the use of larger output devices (for a fixed target gain) or higher PAE or a combination of the two. To a certain point, increasing gain for a fixed target output power leads to higher PAEs when the PA is gain-limited (below 20 dB, for instance). For a fixed target current density, normally around peak-f_T/f_{MAX}/MAG, larger HBT devices have lower input and output resistances, which leads to higher source-impedance and load-impedance transformation ratios, making matching networks more narrowband and more sensitive to component tolerances and IC process variations.

HBT power cells are sized large, relatively, for their frequency of operation in order to achieve more output power, and the practical limitations of the device size fall on available device gain as well as impedance-matching bandwidths and sensitivities to component tolerances. Both gain and impedance at the input and output of the HBT decrease with larger power device sizes. The input impedance of large devices, biased at near peak

f_T/f_{MAX}/MAG, are even lower than their output impedance, and the interstage matching networks have the most difficult elements to implement precisely due to the required small component values and the limited area between transistor gain stages.

In the W-band PA design, the transistors are designed to have at least 3 dB gain near P_{1dB} so that the Class A/AB amplifier can be sufficiently driven into full compression for maximum output powers P_{sat}. To ensure that driver transistors Q1 & Q2 are not the limiting the Q3 HBT output-power capabilities, Q1 and Q2 HBTs are sized for PA output powers near P_{sat} so that both transistors can deliver enough available power to fully compress Q3. To meet the input-power requirements, a transistor's area is sized to be the area of the following (next-stage) transistor scaled down by the gain of the following transistor, i.e., $A_n = A_{n+1}/G_{n+1}$. Each transistor is sized to provide sufficient power to drive the following stage into P_{sat}. The Q1 and Q2 power HBTs are thus conservatively oversized according to Fig 2.58.

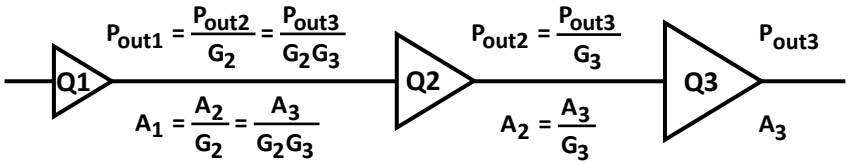


Figure 2.58: Power Transistor Sizing

With a target minimum gain of 3 dB in each stage, each transistor is sized to be half the size of the next stage transistor. The transistors are geometrically scaled with each subsequent stage twice as large in total area. This scaling approach guarantees that the final stage of the PA, Q3, is the limiting stage when the PA enters gain compression at high output powers. Q2 is 20 μm × 0.12 μm, and Q1 is 10 μm × 0.12 μm. Since the actual gain per stage is higher than 3 dB at the 1-dB gain compression point, P_{1dB}, this sizing methodology results in slightly oversized transistors, but the headroom allows for current and voltage tuning of the operation point, giving flexibility in tuning the RF matching by changing DC points, as described more in Section 2.9.6. In future iteration, reducing the size of the transistors in the first two stages can increase gain, reduce matching-network losses by easing the matching conditions, and reduce overall DC power consumption, resulting in improved PAE, but at the expense of headroom for bench-tuning the final PA in the lab.

The available gain (MSG & MAG) for the Q3 40 μm HBT (output transistor) is plotted versus collector current for various collector voltages, V_{CC}, in Fig. 2.59. Biased at 1.7 V and the peak-f_T DC quiescent current of 60 mA, the 40 μm npn HBT has a maximum available gain of only 5.6 dB at 90 GHz. Above 20 GHz, the 40 μm transistor is unconditionally

stable, and maximum stable gain (MSG), where $K = 1$, is plotted for frequencies below 20 GHz. The simulated gain responses take into account additional wiring capacitances of the aggregated power-cell HBT layout as well as the emitter ground inductance. The effects of transmission line feeds into and out of the power cell are not modeled here, but are taken into account at later stages of the PA design within EM simulations and distributed models.

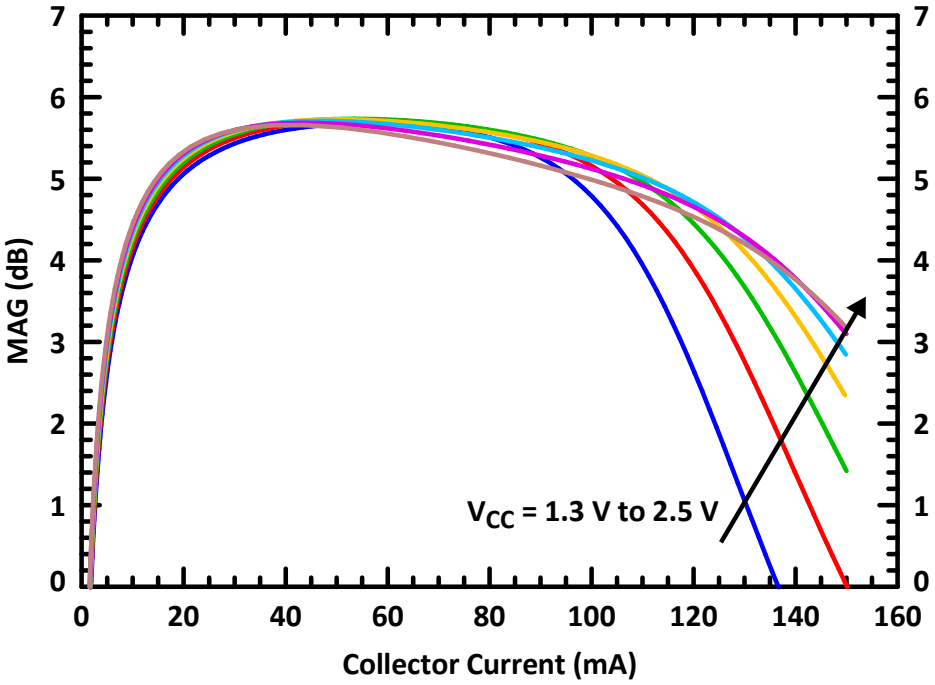


Figure 2.59: Q3 40 µm output transistor MAG @ 90 GHz vs. DC collector current I_C. V_{CC} = 1.3 V to 2.5 V in 0.2 V steps.

At high current densities, the MAG rolloff is slower at higher collector voltages. (Biasing for peak f_T also coincides nearly to peak f_{MAX} and peak MAG biasing.)

Since there is greater than 20 dB small-signal gain below 20 GHz and since the transistor is potentially unstable, the response of the transistor must be managed. Decreasing the low-frequency gain outside of W-band of the PA increases the overall amplifier stability. Collector bias lines, base bias lines, and the input & output matching stubs all assist in presenting low RF/AC impedances to the HBT terminals below W-band frequencies, which prevents unwanted gain and bias instabilities at lower frequencies. Since the four matching networks are all high-pass type and the base & collector quarter-wavelength transmission-line resonator feeds with wideband bypass capacitor arrays are bandpass (high impedance) at W-band, but low impedance outside W-band, the resulting gain roll-off at the lower end of the PA bandwidth is very steep, and gain outside W-band is attenuated significantly for

Table 2.7: W-Band Transistor Values. Q_B is the unit-cell transistor also used in the active-bias reference.

Transistor	**# Unit Cells**	$\mathbf{L_{emitter}}$ **(μm)**	$\mathbf{W_{emitter}}$ **(μm)**	**Emitter Area (μm²)**	**Current Density (mA/μm²)**	$\boldsymbol{I_C}$ **(mA)**
Q_B	1	5	0.12	0.6	12.5	7.5
Q1	2	10	0.12	1.2	12.5	15
Q2	4	20	0.12	2.4	12.5	30
Q3	8	40	0.12	4.8	12.5	60

transistor stability. The emitter of all HBTs are connected directly to a low-inductance DC grounds.

Table 2.7 summarizes the chosen transistor sizing and biasing for all three W-band PA gain stages.

The GF 8HP maximum recommended DC forward-biased emitter current density for high-reliability operation is 15 mA/μm^2. The maximum recommended DC collector-base voltage to avoid avalanche effects is $V_{CB} = 1.5$ V which translates to a $V_{CE} = 2.4$ V since $V_{BE,ON} \sim 890$ mV near peak f_T current density. All DC current densities are identical and below the 15 mA/μm limit recommended by the foundry for reliability and 100,000 POH at 100°C.

2.9.3 Output Transistor Q3

A 40 μm output HBT is close to the limits of HBT transistor sizing at W-band for reasonable system gain and efficiency. In general, larger transistors have less gain than smaller ones due to additional wiring parasitics, and beyond a certain size, impedance transformations become large, narrowing the small-signal and large-signal matching bandwidth and making the design very sensitive to small component variations as the impedance transformations rise above 5:1. Transistors even larger than 40 μm are also impacted by more thermal issues from increased power dissipation and increased mutual thermal-coupling between unit cells because of even higher 2D packing densities to achieve large aggregated power cells.

Fig. 2.60 illustrates the small-signal constant-gain circles at 90 GHz without parasitics or mutual heating effects at nominal 1.5 mA/μm current biasing at $V_{CC} = 1.7$ V. The $G_{p,max}$ matching point results in the maximum available gain (MAG) at 90 GHz, but the G_p matching point does not take any input mismatch into account. Some additional Q3 input

mismatch will exist when the interstage matching network is optimized and iterated with the load pull of the prior Q2 transistor. In addition, > 0.5 dB of small-signal gain is lost when the wiring parasitics and additional mutual thermal coupling are considered in the transistor.

Fig. 2.61 shows the maximum stable gain and maximum available gain for the designed 40 μm output power transistor Q3. The simulation includes EM-extracted parasitics related to inter-unit-cell wiring, but does not take into account electro-thermal coupling effects. The 40 μm transistor is unconditionally stable above 22 GHz (Rollett $K > 1$). On-chip capacitive bypass networks terminating the quarter-wavelength bias insertion network at both the collector and base nodes stabilizes the transistor. Small-signal MAG at 90 GHz with extracted parasitics is just 5.7 dB. Losses from the biasing networks and interstage & output matching networks are not taken into account, and realizable Q3 stage gain will be less than 5 dB.

For $K < 1$, MSG is defined in Eq. (2.36). (See Appendix A for more details of K).

$$G_{MSG} = \frac{|S_{21}|}{|S_{12}|} \tag{2.36}$$

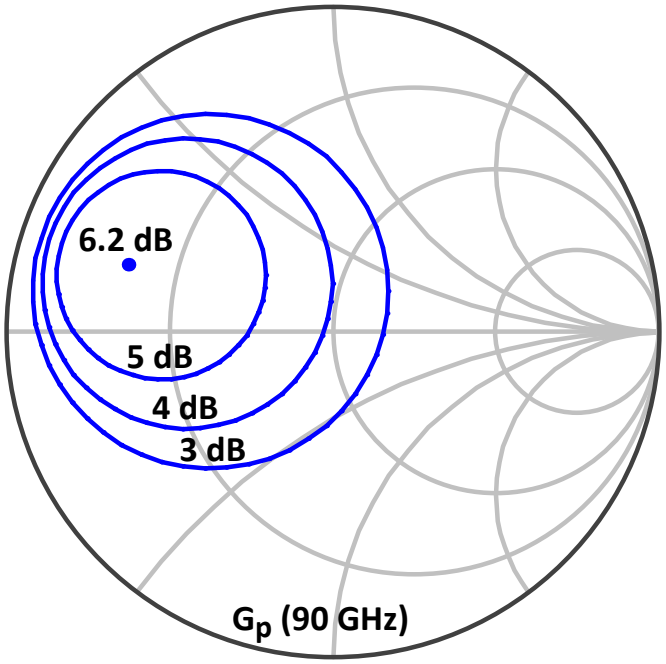


Figure 2.60: G_p output Smith chart for 40 μm transistor.

Fig. 2.62 illustrates the maximum available gain (MAG) at 90 GHz versus collector current over various collector voltages $V_{CC} = 1.5$–2.5 V for the implemented 40 μm Q3 aggregated power-cell HBT.

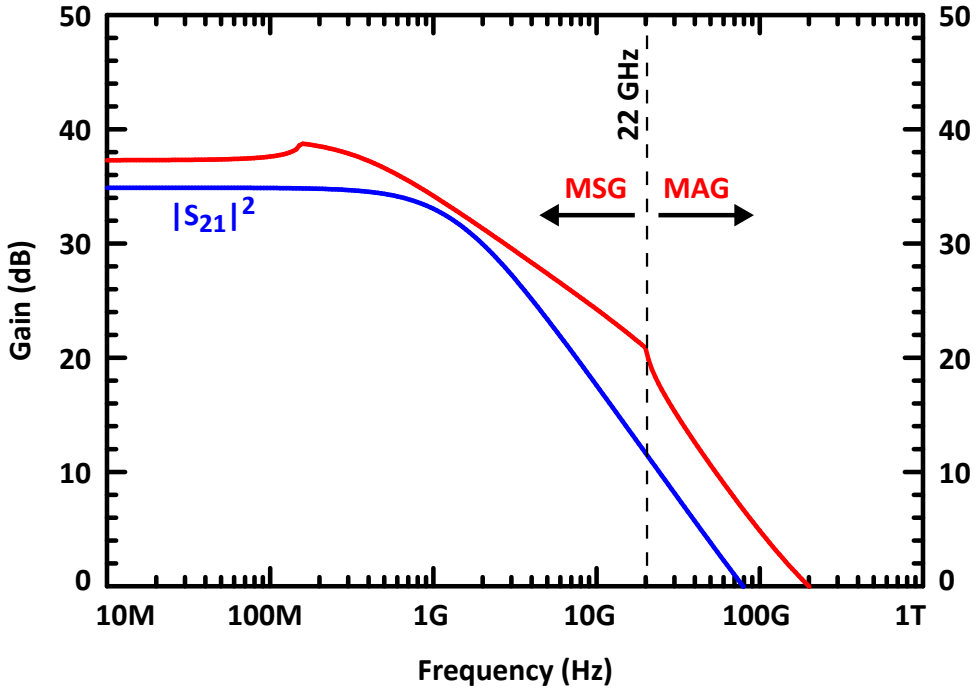


Figure 2.61: Maximum stable gain (MSG) and maximum available gain (MAG) for Q3 = 40 μm, V_{CE} = 1.7 V.

Associated HBT wiring parasitics, inter-power-cell parasitics, and mutual thermal coupling within the power-cell is modeled, but additional inter-stage wiring parasitics and losses are not included. Inter-stage parasitics and transmission-line losses (Q2–Q3 and Q3 to output matching network) are not included, but they are absorbed into EM simulations and included in the final large-signal design optimizations. Peak-MAG biasing coincides with peak-f_{MAX} and peak-f_T biasing at 1.5 mA/μm. At the peak-MAG DC biasing and 1.7 V collector voltage, the 40 μm device draws 60 mA and has a corresponding MAG of just 5.7 dB at 90 GHz. Accounting for additional design-implementation losses, the overall Q3 gain stage achievable MAG will be < 5 dB at 90 GHz. MAG is within 0.5 dB from its peak from 30–100 mA currents, but DC quiescent biasing above 72 mA violates reliability guidelines for the HBT (15 mA/μm^2 or 1.8 mA/μm current densities).

Parasitic Model

Fig. 2.63 shows the simplified model for the aggregated 40 μm Q3 transistor. $R_{E,\mathrm{ext}}$ = 0.03 Ω, $L_{E,\mathrm{ext}}$ = 1.5 pH, C_{BE} = 21 fF, C_{BC} = 8 fF, and C_{CE} = 0.6 fF. Additional

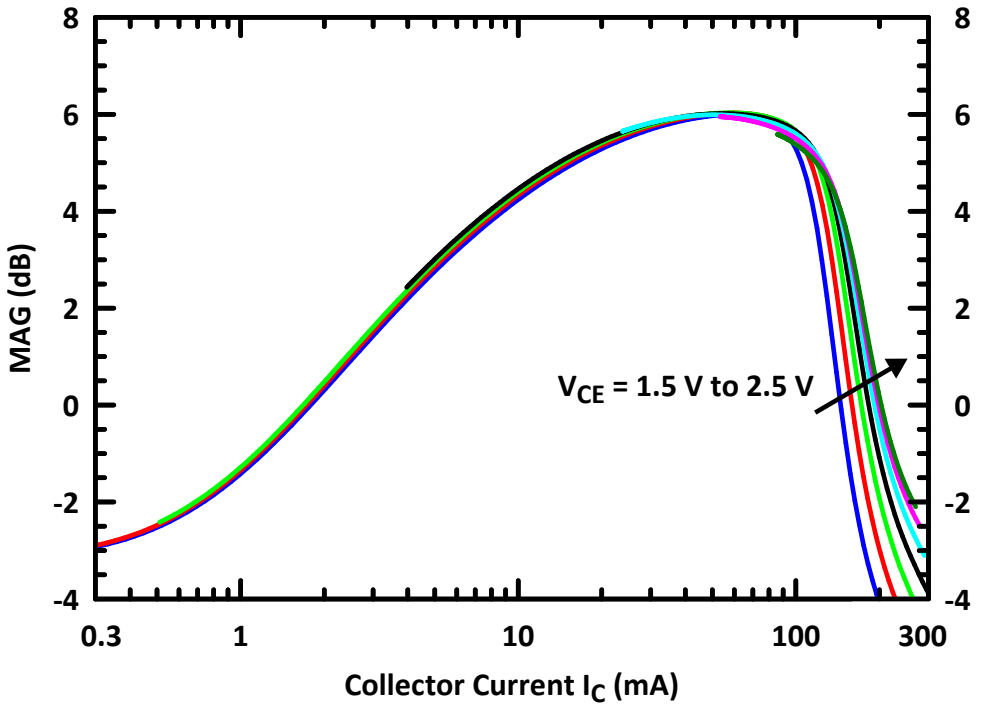


Figure 2.62: MAG vs. collector current I_C for 40 μm npn model with parasitics and mutual thermal-coupling effects. V_{CE} = 1.3 V to 2.5 V in 0.2 V steps.

external R_C and R_B resistances and other external[27] , distributed parasitics are absorbed into simulated EM models (S-parameter blocks) of the transmission-lines and interconnects that connect to the HBTs' external terminals. The additional parasitics around the 40 μm HBT drop the peak f_T below 180 GHz and peak f_{MAX} below 200 GHz.

No emitter ballast resistors are used in order to preserve the gain at 90 GHz. The dense layout of the aggregated 40 μm HBT power cell and increased mutual-thermal coupling between its 5 μm unit cells does not lead to any thermal runaway conditions, as discussed in Section 2.5.6. The custom mutual-thermal coupling coefficient, extracted from DC measurements, m_{th} = 1.34 (from Section 2.5.6) completes the large-signal model of the Q3 HBT power-cell transistor.

The external $R_{B,bias}$ resistor in the schematic represents the DC external base resistance which ballasts the excess holes generated from operating the HBT near and above BV_{CEO} = 1.8 V. $R_{B,bias}$ is 20 Ω in the W-band PA design for all HBT gain stages.

[27]Intrinsic HBT resistances are within the physical transistor cell (within substrate and not to any metal contacts). Extrinsic HBT resistances are modeled within the HBT parameterized cell (p-cell). External HBT resistances are external to the p-cell (additional layout resistances from wiring and also external bias resistances/resistors).

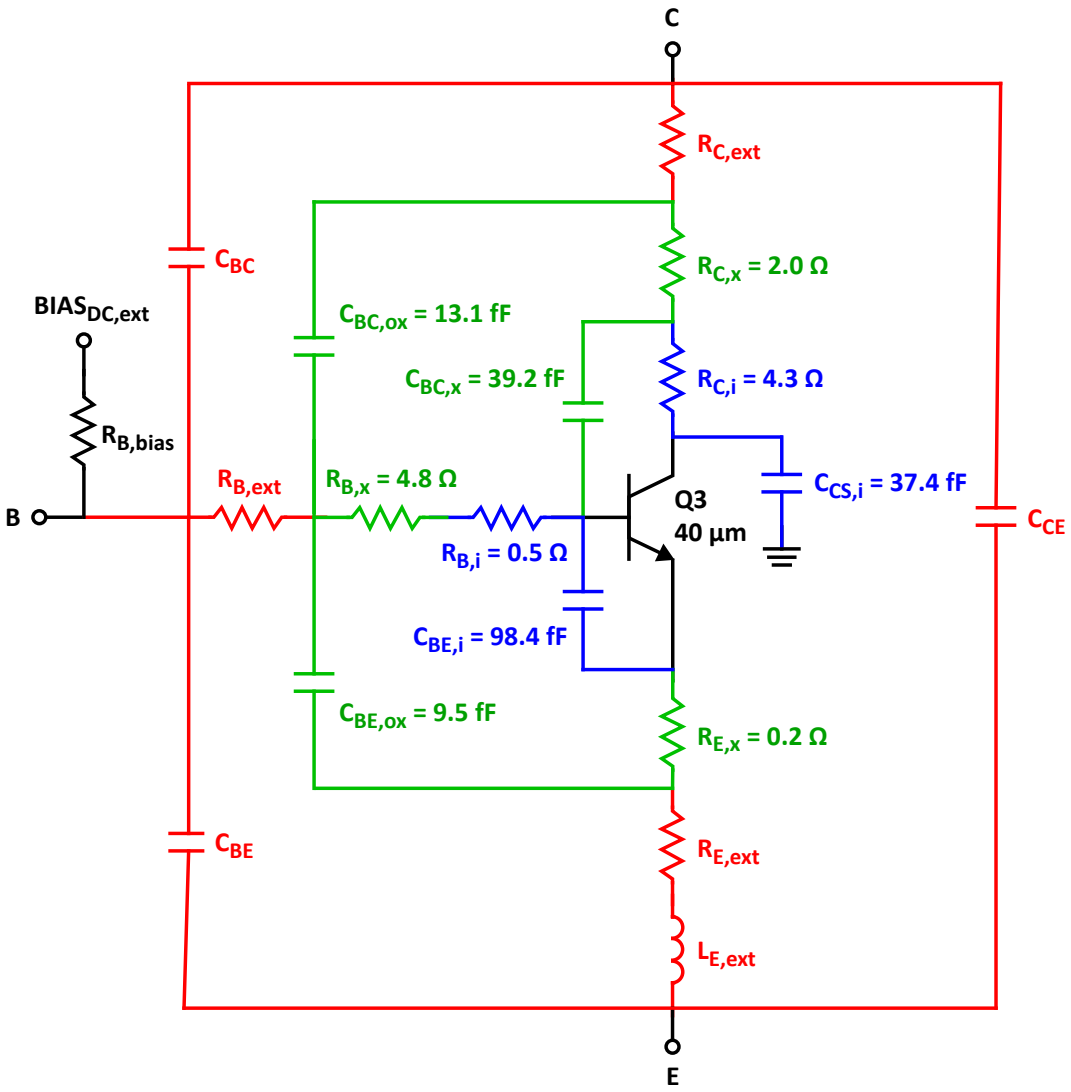


Figure 2.63: Simplified model for aggregated 40 µm Q3 transistor. Intrinsic components are in blue, extrinsic components are in green, and external components are in red. The internal HBT model is contained within Q3 (in black). $R_{\mathrm{B,bias}}$ is the external DC resistance presented to the base terminal by the base-bias insertion network.

Bilateral HBT

The Q3 power HBT cell is a bilateral transistor device within W-band. The unilateral figure of merit U is derived from 2-port S-parameters in Eq. (2.37) [104].

$$U = \frac{|S_{12}||S_{21}||S_{11}||S_{22}|}{\left(1 - |S_{11}|^2\right)\left(1 - |S_{22}|^2\right)} \tag{2.37}$$

The output 40 μm Q3 HBT power-cell has a simulated $U = 0.17$ within W-band, which includes parasitics and thermal-coupling effects. The error in the predicted transducer gain[28] by treating the bilateral HBT device as an ideal unilateral device (i.e., $|S_{12}| = 0$) is bounded by expressions involving U, as detailed in Eq. (2.38).

$$\frac{1}{(1+U)^2} < \frac{G_T}{G_{TU}} < \frac{1}{(1-U)^2} \tag{2.38}$$

G_T/G_{TU} is the ratio of bilateral transducer power gain to the transducer power gain under unilateral assumption ($|S_{12}| = 0$). The expression can be converted into log (dB) terms[29], shown in Eq. (2.39).

$$G_{TU,\mathrm{dB}} - 20\log_{10}(1+U) < G_{T,\mathrm{dB}} < G_{TU,\mathrm{dB}} - 20\log_{10}(1-U) \tag{2.39}$$

For the 40 μm Q3 HBT, simplified design approaches that assume a unilateral HBT results in a possible gain error from -1.4 dB to +1.6 dB. With a 3.0 dB range in predicted gain when the HBT only has a MAG of 6 dB at 90 GHz, the Q3 HBT must be treated as a bilateral device— unilateral S-parameter simplifications do not apply well. The bilateral nature of the HBT also implies that load matching will affect the optimal source matching, which in turn affects the load matching of the prior-stage HBT, and so forth. The final large-signal design needs to be optimized and iterated in the simulator to take the bilateral behaviors into full account.

2.9.4 Matching Networks

All matching networks— input, Q1–Q2 interstage, Q2–Q3 interstage, and output— use 2-element L-matching networks with shunt inductances and series capacitances. The capacitance comes from compact 91 fF unit-cell MIM capacitors, and the inductance is synthesized from shorted unit-cell 11–8–11 μm 50 Ω transmission-line stubs used throughout the PA. The one exception to the MIM 91 fF unit cell use is the larger 130 fF MIM Q2–Q3 interstage

[28] Transducer gain is defined as G_T (in dB) $= 10\log_{10}(P_{\mathrm{load}}/P_{\mathrm{available}})$.
[29] The $20\log_{10}(1-U)$ term on the right side is a negative value, so the number adds to $G_{TU,\mathrm{dB}}$

Table 2.8: HBT small-signal input and output resistances at peak-f_T/f_{MAX} biasing currents and $V_{CC} = 1.7$ V.

Power Transistor	**Emitter Length (μm)**	**Input** $Re\{Z_{in}\}$ **(Ω)**	**Output** $Re\{Z_{out}\}$ **(Ω)**
Q1	10	20	46
Q2	20	10	23
Q3	40	5	12

matching capacitor that is necessary to handle the larger interstage RMS/AC currents for reliability compliance at the very low-impedance 5 Ω Q3 HBT base node.

The MIM capacitors (1 fF/μm^2 density) are described in detail earlier in Section 2.7. The transmission-line stubs are detailed more in Section 2.8.1 and Section 2.8.2.

All RF four matching networks[30] in the half-PA are high-pass-type matching networks. The 91 fF capacitor is always the series element, and the distributed transmission-line inductor is always the shunt element terminated in short circuit (or 440 fF resonant MIM capacitor as equivalent RF short circuit at 90 GHz).

Using high-pass filter matching networks compensates for the natural low-pass behavior of the HBT gain rolloff (f_T/f_{MAX}/MAG rolloff) at mmW frequencies and results in higher operating bandwidths and less gain peaking. Gain roll-off at higher frequencies is due to the intrinsic capacitances within the HBT device while roll-off at the half-PA lower band edge is primarily set by the 4 high-pass matching networks and 6 quarter-wavelength transmission-line resonator feeds (terminated on one end in wideband low RF-impedances).

The transmission-line stubs in the matching networks allow simple adjustments to the design without requiring major re-work. The shunt stub can simply be made slightly longer or shorter to adjust the design's midband frequency and gain-shaping characteristics.

Table 2.8 lists the real-part of the small-signal input impedances, including wiring parasitics, of the aggregated power cells Q1, Q2, and Q3 biased at peak f_T/f_{MAX}/MAG current density of 1.5 mA/μm at 90 GHz.

Fig. 2.64 shows the impedances of key elements of the PA. All impedances are given

[30]The term "matching network" is used generically in this text and does not necessarily mean maximum power-transfer matching. Strictly, "matching" implies complex-conjugate matching for maximum power transfer while "impedance transformation network" is technically more accurate terminology for networks that simply transform one impedance to a different target impedance for any reason: maximum output power, minimum noise, maximum power transfer, etc.

at 90 GHz unless specified otherwise. The HBT impedances are given in $Re\{Z\}$ while transmission-line impedances are given in $|Z|$.

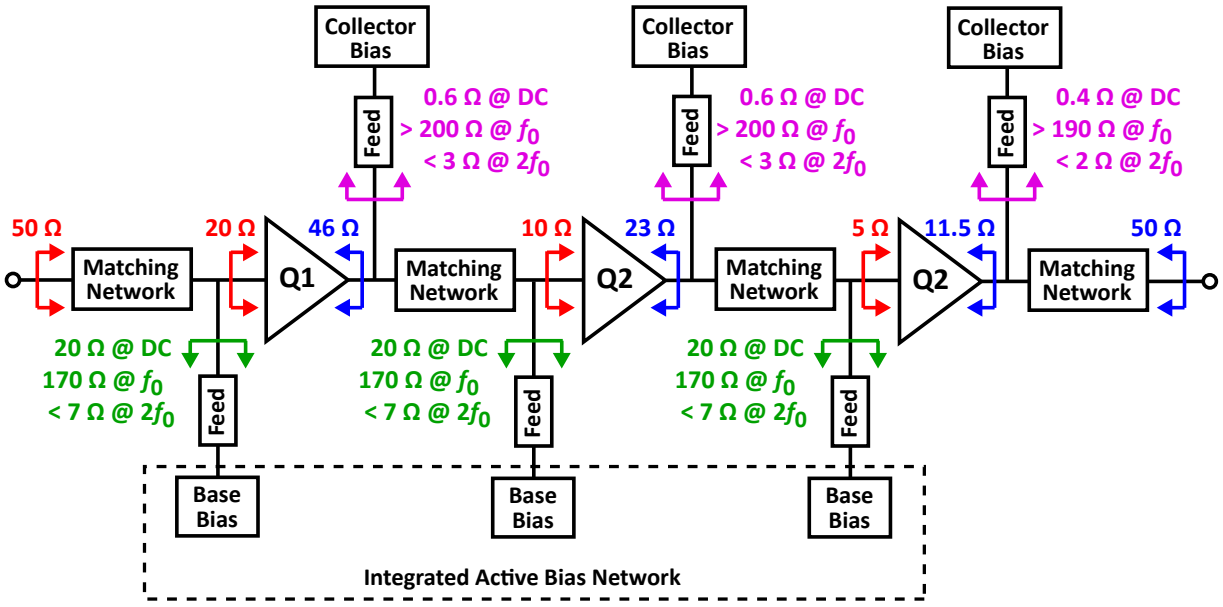


Figure 2.64: Key input and output impedances for the W-band half-PA. All impedances given at 90 GHz unless specified otherwise. Transistor impedances are Re{Z} while quarter-wavelength impedances are $|Z|$.

At the nominal PA biasing (1.5 mA/μm) and collector voltage ($V_{CC} = 1.7$ V), the real part of the Q1 small-signal input and output impedances are 20 Ω and 46 Ω, respectively, which includes wiring parasitics and mutual thermal coupling effects. The real part of the Q2 small-signal input and output impedances are 10 Ω and 23 Ω, respectively. The real part of the output Q3 small-signal input and output impedances are 5 Ω and 12 Ω, respectively. Small-signal HBT impedance scale down approximately 2× with each subsequent stage since transistors are sized 2× larger while operating at the same peak-f_T 1.5 mA/μm collector current density. Complex-conjugate matching would result in 4.6:1 impedance transformation ratios between transistors (interstage) and approximately 4.2:1 output matching transformation, but power matching via load-pull and source-pull iterations move matching networks away from ideal complex conjugate-matching conditions. Only the PA input, a 2.5:1 impedance transformation, is matched closer to standard complex-conjugate matching conditions.

As an example, even the optimal input matching of Q1 varies depending on the design goal. For the capacitor in the two-element impedance-transformation network, 80 fF is best for small-signal input matching (close to conjugate matching for maximum power transfer), but under large signal RF inputs close to gain compression, HBT capacitances change, and

the optimum source matching for Q1 output power moves the optimal MIM values to 110 fF. The chosen 91 fF MIM is optimal for the Q3 output matching for power and about the midpoint between small-signal and large-signal power matching for Q1. Using the same 91 fF capacitor at both input and output follows the unit-cell approach, and device mismatches track extremely well. The input shunt stubs are adjusted accordingly at both the input and output to complete the matching/impedance transformation networks based on large-signal load-pull and source-pull sweeps over power, frequency, and collector voltages.

While smaller inductances can be readily synthesized with shorter transmission-line sections, MIM capacitors face reliability concerns below 65 fF, because of their limited number of vias for small MIMs. Only one or two vias, rather than four or more, contact the MIM top plate, and any via failure or degradation may result in large shifts in MIM operation. Moreover, with fewer than four vias, the vias begin to violate maximum allowable peak AC RMS current densities at the PA output powers. In order to remain electromigration compliant for reliability, larger MIMs ($>$ 65 fF) must be used or MOM capacitors must be employed. MOM capacitors, discussed later in Section 2.10.1, are much larger than MIM and more difficult to model because of the forced auto-cheesing of the plates (up to 50% local density reduction) and force auto-filling (around 10% local copper density) between the plates. In terms of RMS/AC current compliant, MOMs are not via-limited since they are much larger structures, but they have considerably worse tolerances ($\pm$38% metal thicknesses) because their plates and dielectrics are involved in the copper damascene process

2.9.5 Matching Network Losses

The loaded Q of a 2-element L-matching network with 2 lossless, reactive elements matching a complex load impedance to a real load (purely resistive) is given by Eq. (2.40), where $R_2 > R_1$.

$$Q_{\text{MN,loaded}} = \sqrt{R_2/R_1 - 1} \tag{2.40}$$

For the interstage matching networks, which transform one complex impedance to another complex impedance, the loaded Q of a 2-element L-matching network is given by Eq. (2.41) [111]. (See Appendix B for detailed derivations.) The matching network is illustrated in Fig. 2.9.5.

$$Q_{\text{MN,loaded}} = \sqrt{\frac{R_2}{R_1} - 1 + \frac{X_2^2}{R_1 R_2}}. \tag{2.41}$$

The condition $R_2 > R_1$ is assumed. For the Q2–Q3 interstage matching network in the W-band PA, $R_2 = 23\ \Omega$ (Q2 output), $R_1 = 5\ \Omega$ (Q3 input), and $X_2 = -j12\ \Omega$ @ 90 GHz. The

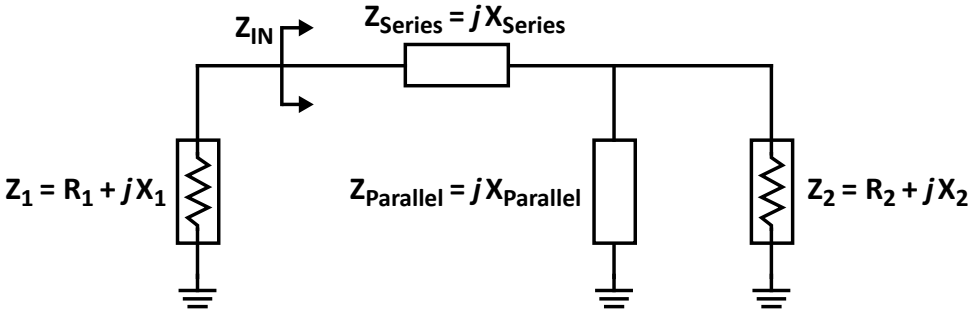


Figure 2.65: Two-element L-matching network with complex-load and complex-source impedances. $R_2 > R_1$.

term $X_2^2/(R_1R_2) = 1.25$, and $Q_{\text{MN,loaded}}$ becomes 2.2 for ideal, 2-element, interstage conjugate matching conditions. (Note that analyzing only the real part of the complex impedances ReZ using Eq. (2.40) gives $Q = 1.9$ for the 4.6:1 impedance transformation network.) For the simplified case of matching a complex source impedance to a purely real load resistance, (e.g., $R_2 = 50\ \Omega$, $X_2 = 0\ \Omega$), a 4.2:1 impedance transformation ratio at the half-PA output results in $Q_{\text{MN,loaded}} = 1.8$ while a 2.5:1 impedance transformation ratio at the half-PA input results in $Q_{\text{MN,loaded}} = 1.2$.

Ideal loaded Q of each matching network are given in Table 2.9. Lower loaded-Q_{MN} matching networks implies smaller impedance transformation ratios within the matching network, which generally results in wider matching bandwidths are less sensitivity to component tolerances arising from IC process variations. The highest loaded Q comes from the interstage impedance-transformation networks, and the interstage gain response limits the bandwidth of the overall PA small-signal response.

Since the HBT power cells are ratio 2× in area per stage, the impedance ratios and loaded interstage Q remain approximately the same, not accounting for the differing wiring parasitics. In practical terms, these complex impedance values shift ±15% with RF input powers, and the matching networks must be optimized iteratively for large-signal performance targets where mismatch away from complex-conjugate matching (maximum power transfer) results in higher output powers and efficiencies.

The total transducer loss (PA input to output) from the finite-Q, lossy MIM in the 2-element matching networks is approximately 5.8 dB at 90 GHz, assuming ideal small-signal conjugate-matching conditions, which change for large-signal load-pull and source-pull optimizations, discussed in Section 2.9.6.

The conservative, low 2:1 transistor-area ratio from one gain stage to the next results in lower loaded Q 2-element matching networks, which translates to higher bandwidths. The

Table 2.9: Loaded quality factors for ideal, lossless matching elements and for dominant lossy MIM ($Q = 10$) in the W-band PA matching networks at 90 GHz.

Network	**Lossless** $Q_{\mathbf{MN,loaded}}$	**Lossy** $Q_{\mathbf{lossy,MN,loaded}}$	L_T **(Transducer Loss)**
Input MN	1.2	1.1	1.0 dB
Interstage Q1–Q2	2.2	1.8	1.7 dB
Interstage Q2–Q3	2.2	1.7	1.7 dB
Output MN	1.8	1.5	1.4 dB

trade-off is increased total power dissipation and lowered efficiency. As mentioned earlier in Section 2.9.2, each gain stage power-HBT is sized 2:1 in area to the prior HBT in order to ensure that the last gain stage (Q3) is the limiting transistor when the PA is deep gain compression. The 2:1 ratio ensure that each preceding HBT stage can provide the required input power to fully drive Q3 into deep compression.

For the case of lossy reactive elements, $Q_{MN,loaded}$ decreases, resulting in a more wideband response, but at the undesired expense of resistive losses (transducer losses) and overall, lower system gain. When the MIM capacitor with relatively low $Q_{MIM} = 10$ (unloaded) dominates the loss of the matching network (since transmission-line elements have unloaded $Q \approx 30$), the effective unloaded Q of the matching network elements is approximately Q_{MIM}, the unloaded quality factor of the MIM capacitor alone [112].

The loaded Q of the resulting lossy matching network is given by Eq. (2.42).

$$Q_{\text{lossy MN,loaded}} = \frac{Q_{\mathrm{MN,loaded}}}{1 + (Q_{\mathrm{MN,loaded}}/Q_{\mathrm{MIM}})} \tag{2.42}$$

The loss factor for the 2-element L-matching network with finite-Q, lossy elements is given in Eq. (2.43) [1].

$$LF = \frac{Q_{\mathrm{MIM}}}{Q_{\mathrm{MN,loaded}} + Q_{\mathrm{MIM}}} \tag{2.43}$$

The transducer loss of the 2-element L-matching network with finite unloaded-Q elements with losses dominated by the MIM capacitor is given by Eq. (2.44).

$$L_T \text{ (in dB)} = 20\log_{10}\left(\frac{1}{LF}\right) = 20\log_{10}\left(1 + \frac{Q_{\mathrm{MN,loaded}}}{Q_{\mathrm{MIM}}}\right) \tag{2.44}$$

$Q_{\mathrm{MN,loaded}} = 1.2$–$2.2$ is the loaded Q of the implemented 2-element matching network and

$Q_{\text{MIM}} = 10$ is the effective, unloaded Q of the 2-element network dominated by the MIM capacitor loss.

With all input, output, and interstage matching networks as high-pass types and all bias-insertion networks as high-impedance wideband quarter-wavelength resonators, there is little midband reactive peaking, no severe gain ripples, and smooth gain roll-offs at each band edge. The loaded Q of the matching networks are 1.2–2.2, which results in a wideband matching and smooth gain response.

Adding additional L–C stages (ladders) to the matching network can make the matching wider in bandwidth, but in practical implementations constrained by layout, the loss of multiple L–C ladders is more than that of a single 2-element L–C section because of the relatively low Q MIM capacitors. Metal-oxide-metal capacitors with $Q > 20$ are discussed more in Section 2.10.1. For the same desired capacitance, MOM capacitors are 10× larger in area than MIM capacitors and have much worse capacitance tolerance values, so they are only used experimentally in the same 2-element matching networks to push the PA design above 100 GHz.

The practical implementation of the matching networks involved detailed simulations that sweep frequency as well as RF-input powers. The source and load impedances change with frequency and power, and they change at different rates— the source and load impedances in the preceding matching-network examples do not track together over frequency and power. Additional parasitic wiring and EM effects from transmission-line discontinuities must also be taken into account in the simulator.

Small-Signal Output Matching

Approximate values for a 2-element L-matching network can be found for the PA output matching to 50 Ω at 90 GHz using the lossless matching network assumption. The impedance match from 50 Ω to 11.5 Ω (a 4.3:1 impedance match) corresponds to $Q_{MN,loaded} = 1.83$. When using a 2-element L-matching network with shunt-inductor and series-capacitor, the matching-network element values are found to be $L_{\text{shunt}} = 4350/f_{\text{GHz}}$ (in pH) and $C_{\text{series}} = 7560/f_{\text{GHz}}$ (in fF) where $f_{\text{GHz}} = 90$ for 90 GHz. For a fixed *impedance-transformation ratio*, the component values scale inversely with frequency (i.e., $1/f$). For a fixed center *frequency*, the required capacitance increases with increasing impedance transformation ratio while the required inductance decreases.

For maximum output powers, the output of the power amplifier is not conjugately matched to the output transistor's output impedance. Rather, the load is deliberately mismatched to provide the required voltage and current dynamic load lines for maximum power. (The output matching network becomes an impedance transformation network— it is not matching

conjugately, but providing a necessary impedance transformation.) The large signal output-matching is discussed more in Section 2.9.6. The optimal output-power match must consider the changing HBT source and load impedances under high input & output powers as well as the effects of increasing collector voltages $V_{CC} > 1.7$ V.

One of the considerations for the limits of the output Q3 HBT size is the required impedance transformation and component tolerances. Impedance transformations beyond 5:1 (50:10 Ω) or $Q_{MN} > 2$ with a 2-element matching network at the output makes the PA design much more sensitive to small shifts in values within the reactive matching networks, particularly the MIM capacitor. Effects of the matching-element component variations are amplified by the transformation ratio. Although larger impedance transformation ratios

Interstage Matching

The interstage-matching networks consist of a shunt series-resonant-LC circuit followed by a series MIM capacitor. The simplified circuit is shown in Fig. 2.66. The series-resonant LC structure consists of a short transmission line in series with a 440 fF MIM capacitor. The LC structure is used just beyond its series-resonant frequency in order to present a small inductive reactance (10–20 pH) while maintaining a DC block.

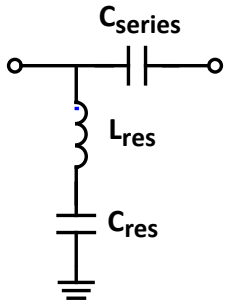


Figure 2.66: Interstage matching network that steps down the HBT output impedance (left side) to the HBT input impedance of the next stage whose HBT is 2× larger in area.

The compact interstage matching networks allow the interconnect loss and parasitics between gain stages to be reduced by keeping keeping the Q1, Q2, and Q3 power transistors close together. The only additional series element required between the HBT gain stages is DC-blocking MIM capacitor, which is also part of the reactive matching network

The 2-element L-matching networks sit between the power-cell HBTs: between Q1 & Q2 and Q2 & Q3. These networks transform the required output load impedance of one stage to the input source impedance of the following stage. The Q2–Q3 interstage matching is

particularly challenging because of the low HBT base input impedance (5 Ω at 90 GHz) such small values of effective inductance or capacitance are required.

Like the input and output matching networks, the interstage matching network is comprised of a shunt inductive element and a series capacitor element. The interstage matching networks mimic the same high-pass matching network approach as the input and output matching networks. The two-element interstage matching network is, strictly speaking, bandpass type, but in conjunction with the collector quarter-wavelength transmission line terminated in a wideband short, the overall response is still similar to a high-pass network with very low impedance terminations below W-band.

Setting the optimal transformation network is highly iterative in the design process. Optimizations using iterative simulations are required, as the changes in a power-cell HBT's both source and load terminations affect PA characteristics. The bilateral/non-unilateral nature of the HBT amplifier at W-band, especially at high output powers, means that source terminations affect optimal source terminations for maximum power (or gain or PAE) and vice versa. The effects of a termination of one stage propagate throughout the PA gain stages all the way to the terminal output-matching and input-matching networks.

The Q1–Q2 interstage has a 60 μm RF-shorted 11–8–11 μm unit-cell transmission-line stub followed by a unit-cell, series 91 fF MIM capacitor. The Q2–Q3 interstage has a 30 μm RF-shorted 11–8–11 μm unit-cell transmission-line stub followed by a series 130 fF MIM capacitor, which is larger than the unit cell in order to maintain reliability compliance for the high AC/RMS currents at the Q3 HBT base nose.

Since the transmission-line stub (inductive element) sits at the collector output node, it cannot shunt DC current to ground— the stub cannot be DC-grounded, but it needs to be AC/RF-grounded for proper response as a small (10–20 pH) inductance. The unit-cell, self-resonant (in W-band) 440 fF RF-shorting capacitor is used in series with the shunt transmission-line section to synthesize the necessary RF-short while remaining DC-blocked.

The Q2–Q3 interstage shunt-stub transmission-line inductor and series 440 fF resonant RF-shorting capacitor are shown in Fig. 2.67.

With the 2-element L-matching network approach, in general, to match to impedances with smaller real parts, smaller inductances are required. At mmW frequencies with large power-cell HBTs running at high current densities, input impedances are small (5–10 Ω), requiring very small inductance values (< 20 Ω reactances). The shunt inductive matching element is mostly responsible for stepping down the node impedance, so its value becomes extremely small at W-band (10–20 pH) for interstage HBT-input impedances.

While series L-C filter structures as shunt elements in matching networks are not new, this configuration is unique because the capacitor, alone, provides no reactance at the target

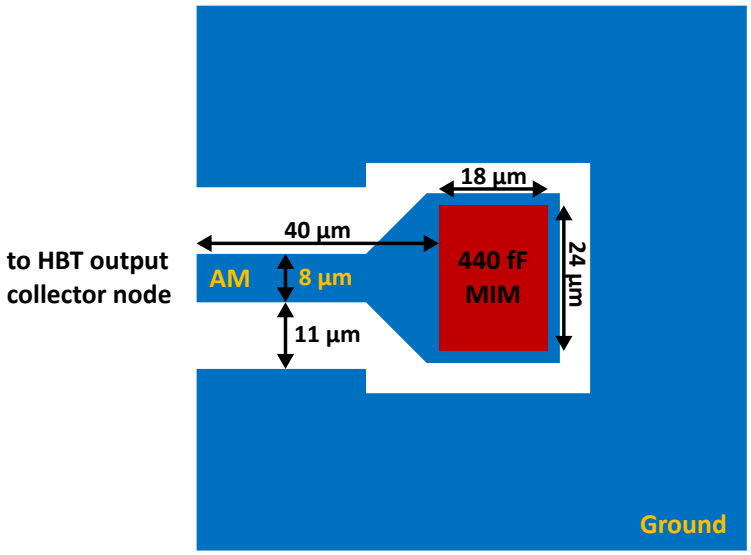


Figure 2.67: Layout for interstage MIM capacitor. The top plate is connected to the (11–8–11 µm), 50 Ω transmission line. Not shown: the global ground plane underneath the entire structure to which the bottom plate of the MIM is connected; and the vias that connect the MIM top plates to the top-level metal (AM).

frequency of operation, 90 GHz. It looks like an RF/AC short circuit with small parasitic residual resistance (equivalent series resistance, ESR = 0.6 Ω) at 90 GHz.

The combined transmission-line inductance and RF-shorting MIM capacitor can also be analyzed as an equivalent LC structure that is used past its combined series-resonant frequency since the 440 fF MIM capacitor is already self-resonant at 90 GHz.

At the resonance frequency (40 GHz for Q1–Q2 LC structure, 48 GHz for Q2–Q3) the shunt LC structure appears as an RF short-circuit with residual real-part ESR = 0.6 Ω. Beyond the series resonance frequency, the structure provides reactance ($+jX$) equivalent to a very small inductance.

Q1–Q2 interstage 60 µm transmission-line stub terminated in the 440 fF MIM resonates at 40 GHz and results in a 30 pH equivalent inductance at 90 GHz with $Q = 15$. The Q2–Q3 interstage 30 µm transmission-line stub terminated in the 440 fF MIM resonates at 48 GHz and results in a 10 pH equivalent inductance at 90 GHz with $Q = 8$. Both resonant structures are DC-blocked, which is necessary to prevent collector DC collector currents from shorting to ground.

The interstage matching network utilizes the same unit-cell resonant 440 fF MIM capacitors as the ones used to terminate all quarter-wavelength resonators in the collector-bias lines as well as in the base-bias lines. Any MIM process skew allows all the all the matching networks

and quarter-wavelength resonators to track together. Back-modeling is also simplified The matching between the identically-dimensioned RF-shorting capacitors is better than 0.1$, so all 16 440 fF MIM capacitors in the differential PA (or 8 in the half-PA) will shift the same in process variations, and the unit-cell 11–8–11 μm transmission lines, defined by the aluminum top-metal dimensions, have equally excellent matching among all unit-cell lines.

Since the transmission-line section, in series with the resonant 440 fF MIM capacitor, is equivalently an LC series-resonant circuit, the design-question arises whether a larger MIM capacitor could be used to synthesize the same equivalent inductance (a target $+jX$ reactance past the LC resonance). The larger the MIM, the shorter the required transmission-line to synthesis the same inductance, making the layout more compact. The issues with this MIM-size maximization are two-fold. First, most importantly, moving away from the unit-cell resonant 440 fF MIM capacitor approach allows mismatch (from process variations or model errors) between non-identical resonant capacitors to de-tune the PA response. Secondly, larger-valued MIM capacitors do not have higher Q at 90 GHz because the associated resistance is scaling directly with the capacitance. Thus, larger-valued MIM capacitor Q is low compared to short transmission-lines ($Q > 30$), so the overall Q of the inductive stub terminated with MIM capacitor drops and approaches the Q of the MIM capacitor as the transmission-line section becomes shorter, as shown earlier in Table 2.3.

2.9.6 Large-Signal Matching

The PA matching networks must take into account the target operational region of the PA: large output powers. For these large-signal designs, the PA is optimized into mild gain compression at output powers near P_{1dB} and beyond.

Since the original VBIC models have severe convergence issues with large-signal harmonic balance above 1.7 V and into deep gain compression (closer to P_{sat}), the large-signal optimizations are centered around simulated load lines at 1.7 V near P_{1dB}, the nominal operating voltage, as well as nominal peak-f_T/f_{MAX}/MAG biasing currents of 1.5 mA/μm. Optimizing the design near P_{sat} and 2.3 V was not possible at the time due to VBIC model convergence issues, but even with the much improved HiCUM models, optimizing a design near the edges of the HBT model— close to voltage & current extremes— leads to significant modeling errors and misleading design optimizations since the transistor models are extracted and centered for more nominal operating conditions.

The large-signal PA design is optimized iteratively using source-pull and load-pull techniques on all three gain stages around the P_{1dB} point with $V_{CC} = 1.7$ V. Fig. 2.68 shows the simulated load-pull contours as well as the implemented output matching from 75–110 GHz

for the chosen 40 µm Q3 output HBT power-cell, which includes additional parasitics and other interconnect effects. The simulated load-pull attempts to find the required optimal load impedance presented to the transistor collector that results in maximum output power.

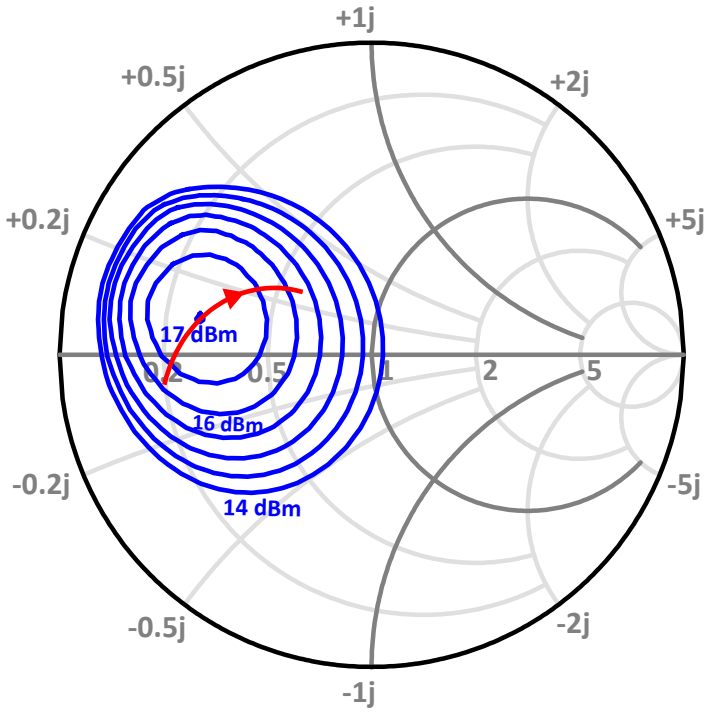


Figure 2.68: Q3 = 40 µm load pull contours: output power contours (blue) from 14 dBm to 17 dBm in 0.5 dB steps; realized load (red) from 75 GHz to 110 GHz with DC quiescent biasing at 60 mA and V_{CC} = 1.7 V.

The optimal load impedance is different than the conjugate match normally required for small-signal maximum power transfer conditions. As an HBT starts to compress, gain begins change from small-signal conditions, and the input and output impedances of the HBT begin to change with increasing RF input-power levels, moving the transistor away from linear-amplifier, small-signal matching conditions. Ideally, the load-pull contours of large power devices would be obtained experimentally using a dedicated load-pull setup in the lab environment, but such data is not available for the custom layout power cells and is also extremely difficult to obtain at W-band.

Fig. 2.69 illustrates the effects of increasing RF input powers on the Q3 40 µm output impedance[31] The Q3 aggregated power-cell HBT large-signal output impedance S_{22} (Z_{22})

[31]The plotted large-signal S_{22}/Z_{22} is the HBT output impedance looking *into* the HBT collector node. The optimal load impedance, as seen *from* the HBT collector node is the conjugate: $Z_{opt,load} = Z_{22}^{*}$ (for maximum

is plotted in Fig. 2.69 at peak-f_T biasing currents (1.5 mA/μm) and V_{CC} = 1.7 V. The Smith chart illustrates how the output impedance varies with RF-input-power conditions. The output impedance of the 40 μm Q3 HBT changes with RF input powers at 90 GHz from $Z_{22,\mathrm{SS}} = 13.3 - j10.2\ \Omega$ at small signals to $Z_{22,\mathrm{P1dB}} = 10.3 - j7.0\ \Omega$ at the 1dB-gain compression point to $Z_{22,\mathrm{Psat}} = 7.7 - j4.8\ \Omega$ at the onset of P_{sat}.

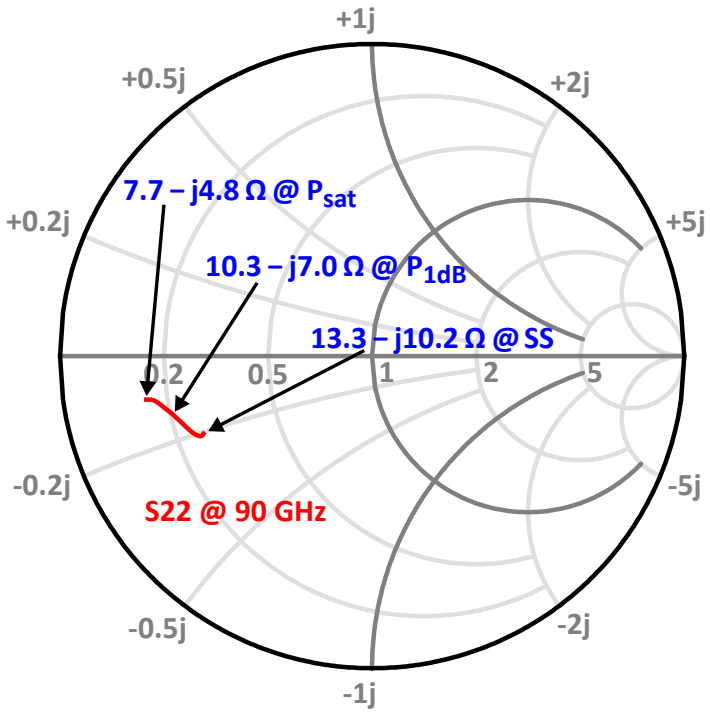


Figure 2.69: Large-signal S_{22} vs. input power at 90 GHz for Q3 = 40 μm, V_{CC} = 1.7 V.

- For the Q3 40 μm aggregated HBT power-cell, the large-signal optimal load impedance near P_{1dB} is $Z_{\mathrm{Q3,load,opt}} = 11.2 + j6.1\ \Omega$ while the optimal source impedance is $Z_{\mathrm{Q3,source,opt}} = 4 + j1.6\ \Omega$. The optimal power matches result in 17.5% PAE at 1-dB compression and 4.7 dB small-signal gain. (Peak PAE occurs farther into gain compression toward P_{sat}).

- For the Q2 20 μm aggregated HBT power-cell, the large-signal optimal load impedance near P_{1dB} is $Z_{\mathrm{Q2,load,opt}} = 25.3 + j13\ \Omega$ while the optimal source impedance is $Z_{\mathrm{Q2,source,opt}} = 7.5 + j2.4\ \Omega$. The optimal power matches result in 16% PAE at 1-dB compression and 5.2 dB small-signal gain.

power transfer with the given optimal load). Fig. 2.68 presents $Z_{opt,load}$ obtained from load-pull simulations while Fig. 2.69 presents S_{22}/Z_{22} characteristics looking into the HBT collector.

- For the Q1 10 μm aggregated HBT power-cell at 1.5 mA/μm peak-f_T/f_{MAX}/MAG biasing currents and $V_{CC} = 1.7$ V, the large-signal optimal load impedance near P_{1dB} is $Z_{\mathrm{Q1,\ load,\ opt}} = 46.8 + j26.3\ \Omega$ while the optimal source impedance is $Z_{\mathrm{Q1,\ source,opt}} = 14.5 + j3.9\ \Omega$. The optimal power matches result in 15% PAE at 1-dB compression and 5.7 dB small-signal gain.

Independent voltage and collector biasing of each gain stage allows selective adjustments to move stages into deeper Class AB operation for additional efficiency at higher collector voltages at the expense of some PA linearity. The independent biasing also allows adjustments arising from expected DC modeling errors and process variations. The independent bias provisions allows re-tuning the DC operating points of each gain stage, which can also adjust for thermal modeling errors and other process variations.

HBT impedances change with the DC quiescent collector currents and collector voltages, enabling some design tuning on the measurement bench. Since all three gain stages have independent current and voltage biasing provision, each stage can be fine-tuned independently. Fig. 2.70 shows the possible tuning range of the Q3 HBT output impedance under small-signal conditions (small-signal RF input powers). Fig. 2.71 shows the possible tuning range (by changing the DC quiescent bias currents and voltages) of the Q3 HBT output impedance under large-signal conditions close to P_{1dB} output powers. The simulations above 1.7 V are possible using the new HiCUM HBT model, but were not available during the time of the W-band design.

Fig. 2.72 illustrates the 40 μm Q3 HBT output impedance S_{22} @ 90 GHz under full gain compression (near P_{sat} with more than 3 dB gain compression) and its dependencies on the DC quiescent biasing currents and collector voltages. The DC quiescent biasing offers opportunities to fine-tune the output impedance response during measurements on the lab bench.

The Q1 and Q2 transistor stages have similar small-signal and large-signal impedance dependencies on quiescent DC bias currents, and the overall PA gain response can be fine-tuned by adjusting Q1 and Q2 biases independently without significantly affecting the Q3 output tuning for maximum PA output power. 15–25% adjustments can be made in HBT impedances simply with DC bias adjustments.

Harmonic Terminations

The 2nd harmonic component is the largest harmonic component in the PA followed by the 3rd harmonic. Presenting short-circuit (low impedance) terminations to the 2nd harmonic content at the HBT collectors controls and shapes the distortion introduced by Class AB and

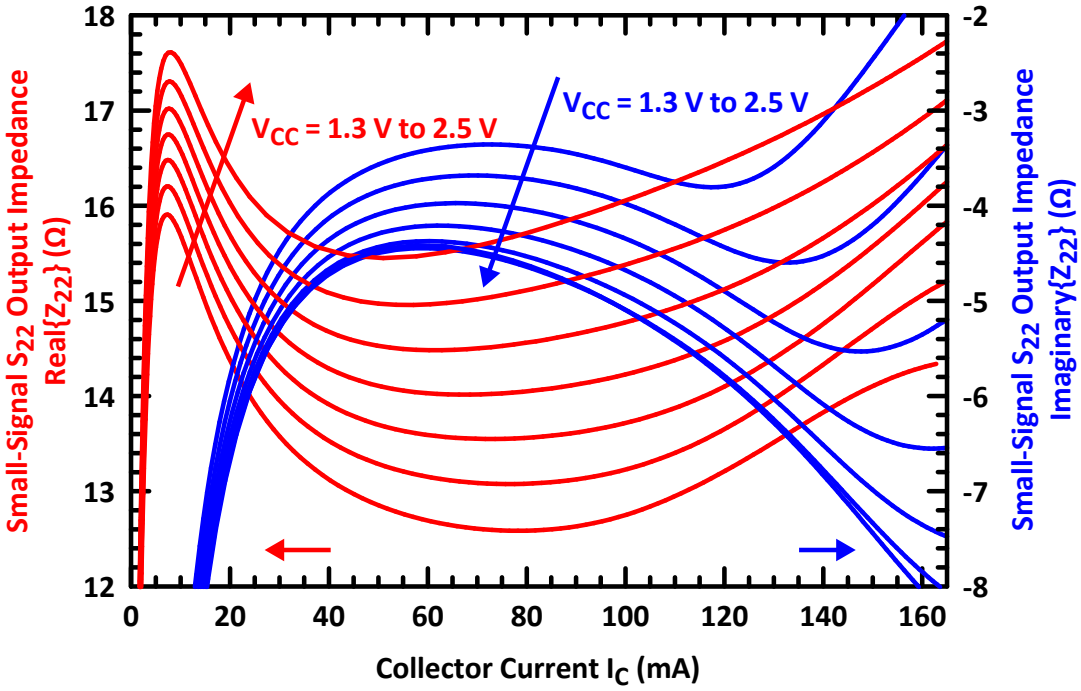


Figure 2.70: Q3 40 µm HBT small-signal S_{22}/Z_{22} vs. collector current I_C at 90 GHz over V_{CC} = 1.3 V to 2.5 V in 0.2 V steps.

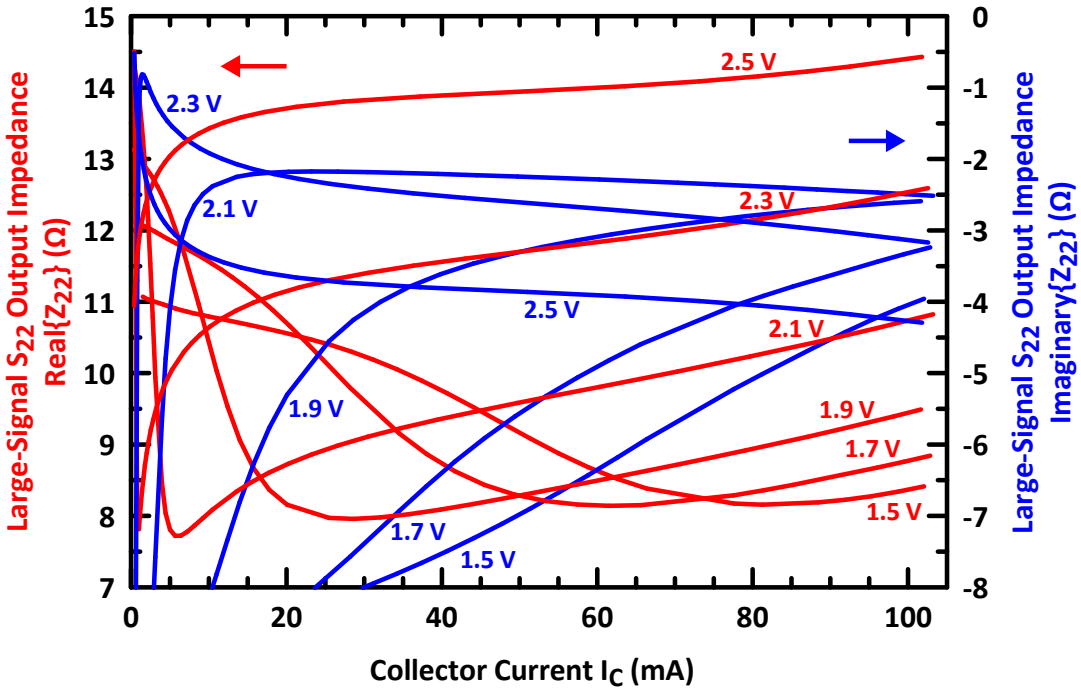


Figure 2.71: Q3 40 µm HBT large-signal S_{22}/Z_{22} vs. collector current I_C at 90 GHz over V_{CC} = 1.5 V to 2.5 V in 0.2 V steps.

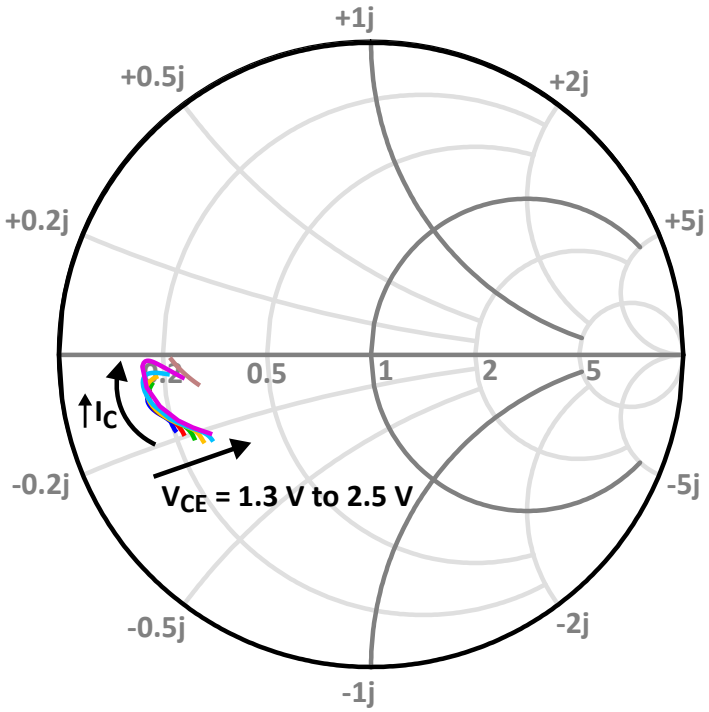


Figure 2.72: S_{22} of Q3 at P_{sat} at 90 GHz. V_{CE} = 1.3 V to 2.5 V in 0.2 V steps.

overdriven Class A operations, especially at high output powers where the PA is under gain compression. Terminating the 2nd harmonic in low impedance shapes the collector output waveform to be more sinusoidal in each gain stage. (The 3rd harmonic is terminated in an open circuit.)

Quarter-wavelength transmission-line resonators terminated in RF-short (at center frequency f_0) results in even harmonic terminations ($2f_0$) with low impedance, which is beneficial for harmonic terminations in Class AB operation [64]. The second harmonic content is managed at the output of each gain stage with the quarter-wavelength transmission-line resonators in the collector bias-insertion lines. The Q1–Q2 and Q2–Q3 interstage nodes also have quarter-wavelength base-feed transmission-line resonators that assist in shorting the 2nd harmonic content near 180 GHz.

The quarter-wavelength transmission lines at W-band frequencies (f_0) is a half-wavelength transmission-line at $2f_0$ frequencies, transforming the broadband MIM short (440 fF MIM capacitor and large bypass capacitor array) to a low impedance at HBT collector output at 180 GHz. The wideband bypass array, in conjunction with quarter-wavelength transmission lines, are critical in the providing a low-impedance termination of the second harmonic content. The smaller MIM capacitors ($<$ 440 fF) provide the necessary low-impedance references at 180 GHz.

EM simulations on the passive components and transmission lines have been done to 330 GHz to cover the third-harmonic response. The EM modeling accuracy at the higher-order harmonics (> 220 GHz) can be significantly relaxed in order to bring down simulation complexity. The absolute accuracy beyond 220 GHz is not important since internal HBT device capacitances shunt (short) the higher-order harmonics, but the EM models need to be continuous and well-behaved for large-signal harmonic balance simulations of the overall PA.

Fig. 2.73 shows the input impedances (seen by the HBT inputs/outputs) of the two types of quarter-wavelength transmission-line resonators, each terminated in 440 fF resonant MIM capacitor and wideband bypass capacitor array, and their combined responses. The transmission-line resonators present high impedances at f_0 and low impedances at $2f_0$. The 2nd harmonics are terminated in impedances $< 2\ \Omega$ near 180 GHz and provide, overall, a simulated > 18 dB of 2nd-harmonic suppression at the PA output.

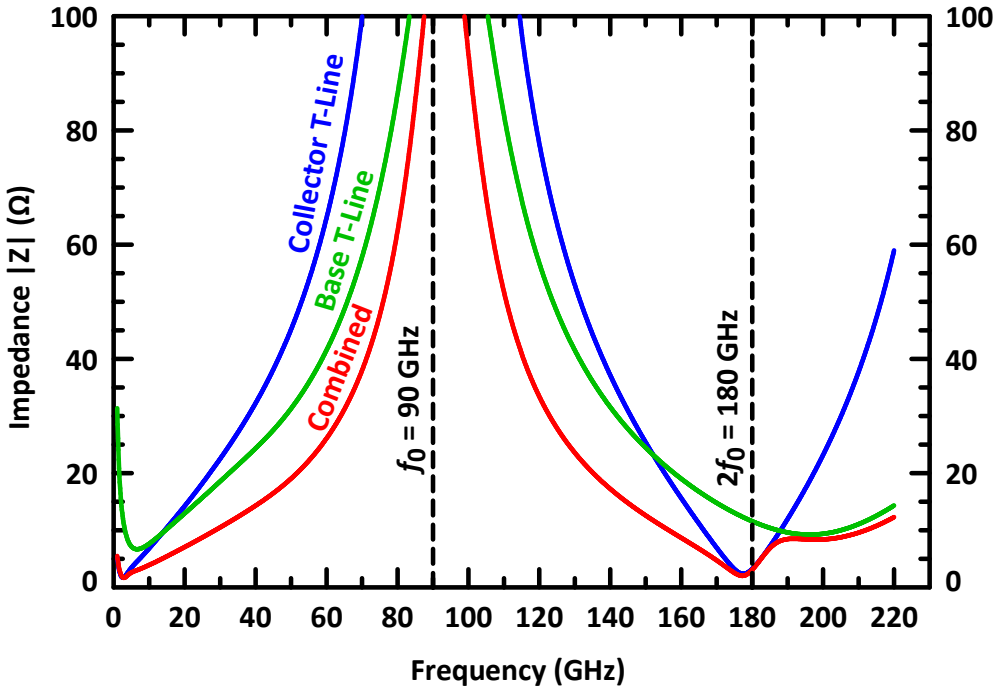


Figure 2.73: Input impedance of quarter-wave feed lines terminated in RF short by modeled MIM and bypass capacitor arrays. Results include the self-resonance modeling of the MIM caps and EM simulations of the quarter-wave transmission line feeds.

2.10 Design Variations

The baseline PA design is centered at 90 GHz with nominal $V_{CC} = 1.7$ V near peak-f_T biasing currents in Class A/AB operation. Each gain stage can be operated in purely Class A or more efficient Class AB, depending on the RF input powers and the collector voltage & quiescent current biasing in each stage.

The response of the W-band PA can be altered by selectively moving certain components away from their unit-cell values (without having to change the unit-cell itself and all associated instances used). In particular, changes in the output matching network and interstage matching shape the overall gain response of the PA that result in different operational midbands.

The baseline design is W1, which uses compact MIM capacitors for all lumped capacitors, centered for large-signal operation at 90 GHz and $V_{CC} = 1.7$–2.3 V. Design variations on the baseline W-band W1 design are made to push the frequency limits of the design, both at the higher and lower frequency limits of W-band (75–110 GHz). The output and interstage matching networks are tuned to lower and higher frequencies. High-Q custom metal-oxide-metal capacitors are used to push the operation frequency to the upper bandwidth limits of the quarter-wavelength transmission lines. These tunings are simple mask changes and do not affect re-working the entire layout.

Designs can be easily re-centered or re-tuned to lower or higher frequencies because of the many high-bandwidth quarter-wavelength bias-insertion networks. The quarter-wave biasing lines and resonant MIM-capacitors can stay fixed to their baseline values, and simply adjusting a few passive elements within the matching networks allows re-centering the PA design without major layout re-work. The frequency adjustments do not require modification of any active components, only metal line lengths (distributed inductors) and capacitor (series reactance) values. The baseline W-band design is successfully tuned throughout W-band in various design breakouts— down to 80 GHz and up to 105 GHz.

The design variations and breakouts also serve an important purpose of obtain better modeling data for the baseline W-band PA design. The shifting of certain components away from unit-cell values in highly-constrained design variations, as is done in Ka-band designs, allows in-situ study of the RF design and back-modeling of parasitics and process biases. The measured data strictly constrains the parameter space for determining passive model shifts and weaknesses in the active HBT models. Having additional data from built-in component skews helps fit the RF modeling across frequency in-situ without having to fully characterize each passive and active component by itself. The back-modeling of the results is constrained such that more accurate results can be obtained.

2.10.1 MOM Capacitors for > 100 GHz Designs

To push the baseline W-band design beyond 95 GHz in center frequency, smaller-valued capacitors are required. At the time of the initial W-band design, MIM capacitors below 90 fF did not satisfy the AC/RMS current-handling guidelines for reliability in the baseline W-band design because of the limited number of via contacts. For example, smaller MIMs in the 60 fF range only have one via contact [32].

The only choice with the 8HP limitations, at the time, was to use custom low-density metal-oxide-metal to create small parallel-plate capacitances within the lower copper layers. The lower-level copper metal (M1, M2, M3, and M4) have the highest parallel-plate capacitances between two adjacent vertical layers within the metal stackup, aside from the dedicated MIM capacitor. The copper-metal parallel-plate capacitors are lower in capacitance density and larger in area, which results in sufficient current-handling capabilities in the contact vias.

In two supplemental PA design variations center above 100 GHz, custom metal-oxide-metal (MOM) capacitors are used to create capacitance values much smaller than 90 fF. The topmost two copper metal layers, M3 and M4, are used to form MOM capacitors that sit 2.61 μm above the SiGe substrate. The M3 and M4 copper metal layers are separated by 0.35 μm, nominally, and sit within SiO_2 dielectric with an effective $\epsilon_r = 3.91$. The resulting M3–M4 MOM capacitor has a low capacitance density of only 0.1 fF/μm^2, 10% that of the MIM capacitor. In other words, for the same target capacitance, MOM capacitors are 10× the area of MIM capacitors, which makes them less suitable for large capacitors, but appropriate for synthesizing small reactances (< 90 fF) for > 100 GHz center frequencies.

The custom M3–M4 MOM capacitors have $Q > 20$ at 90 GHz, higher than the standard 8HP MIM capacitor with Q = 7–10 at 90 GHz. The higher capacitor Q helps to preserve the total PA system gain since maximum available HBT gains (MAG) above 100 GHz fall below fall below 4 dB per gain stage. The HBT MAG near 100 GHz is approximately 1 dB less gain per stage than at 90 GHz, resulting in a maximum system gain that is at least 3 dB less, even with higher Q MOM capacitors, and much lower PA efficiencies (< 5% peak PAE).

Fig. 2.74 illustrates two MOM capacitors used in the Q3 (output HBT) interstage and output matching networks. The M3–M4 MOM capacitor dimensions are $W \times L = 32\ \mu m \times 26\ \mu m$ and $W \times L = 32\ \mu m \times 18\ \mu m$ for 83 fF and 58 fF capacitance, respectively. Large openings are cut into the MQ RF-ground plane, which sits above the M3–M4 MOM capacitors, to reduce lateral and vertical parasitics— capacitive coupling. Other copper-metal structures and wiring are not used underneath the MOM capacitors, as well, and

[32] Later back-end-of-line process improvements have allowed many more vias for these small-valued MIMs (4 or more), but during the time of the design, MIM capacitors smaller than 90 fF began to violate current handling for electromigration for 100,000 power-on-hour operation at 100°C.

shallow-trench isolation (STI) is placed underneath the MOMs to reduce substrate losses.

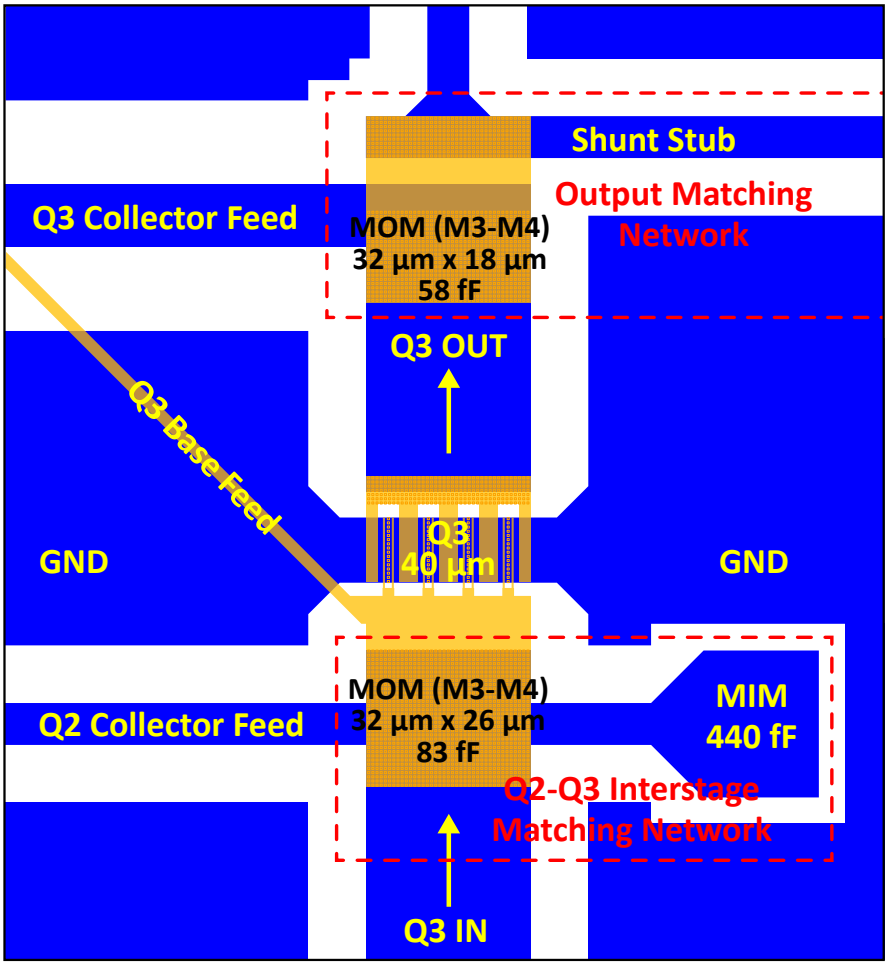


Figure 2.74: M3-M4 metal-oxide-metal (MOM) capacitor implementation in Design W5 & W6. The MOM capacitors are used in the Q2–Q3 interstage resonant matching network as well as the Q3 output matching network. The terminating short of the output matching-network's shunt stub is not shown.

MOM Tolerance Challenges

The copper-based BEOL process introduces significant variations in a M3–M4 MOM capacitor because of the copper damascene processing, which is an additive process that employs chemical-mechanical planarization processes. Metal-oxide-metal (MOM) capacitors created by using oxide between two copper metal layers suffers tolerance issues in the copper-plate thicknesses (±38% for each plate) as well as dielectric thicknesses (±23% between each pair of metals). In contrast, the high-tolerance MIM capacitor has aluminum plates with ±10% tolerances and nitride thicknesses within ±10% tolerances, as well.

The resulting M3–M4 MOM capacitors have $> \pm 23\%$ capacitance values with process variations compared to $\pm 10\%$ for MIM. While matching of identically-dimensioned, MIM capacitors is $< 0.1\%$, the MOM oxide and metal thickness are not as well controlled across the chip, and the expected matching has not been pre-characterized by the foundry.

An additional difficulty of using custom MOM capacitors is the modeling of the parallel copper plates when mandatory copper-metal cheesing is required by the foundry. The large MOM capacitor plates, drawn in CAD layout as solid rectangles, are automatically cheesed (punch-out holes in the large copper plates) during the foundry mask preparation in order to bring the local copper density down to $< 50\%$ within any 200 μm × 200 μm inspection window. This means that up to half of the metal within the copper plate could be punched out. The exact nature & placement of the cheesing and extent of the hole-punching is not known to designers during the chip layout process, so the exact structure of the custom MOM copper plate cannot be pre-determined or simulated precisely, and additional variations are expected on top of the $> \pm 23\%$ capacitance tolerances.

2.11 Measurements

This section presents the W-band PA measurement results. Small-signal, large-signal, and noise-figure measurements are taken on the baseline 90 GHz W-band PA as well as variations (breakouts) centered at different frequencies within W-band. All W-band PA measurements have been done at the University of California, San Diego: in the CalIT^2 labs and in the Rebeiz Group TICS lab (ECE Building EBU-1), also at UCSD. Key W-band high-powered sources and waveguide components have been borrowed from the EECS Radlab at the University of Michigan.

There is not enough available source power to drive the differential/balanced power amplifier into compression with differential inputs & baluns, so the parts are tested under single-ended operation, and +3 dB has been added to single-ended power measurements. The additional +3 dB (plus power saved from balun losses, ∼1 dB) is necessary to saturate the PA, and even this extra power is not enough to push the PA into 1-dB compression past 94 GHz at high collector voltages and is not enough power to saturate some of the design breakouts at 91 GHz. Expected differential output powers are made by adding +3 dB to the single-ended output power. Differential gain & PAE are expected to be the same as the single-ended case if the PA is used with truly differential inputs and outputs[33].

[33] A later work from the group, based on similar PA design approaches within W-band, verified the +3 dB output power boost after de-embedding the input and output balun losses [51]. Other research groups have had to apply the same +3 dB adjustments because of the limited source powers at W-band frequencies [14, 15, 17–19, 71–73].

The PA design limits and HBT device limits are pushed on the measurement bench since simulations of the PA with VBIC HBT device models are not able to predict the PA performance with $V_{CC} > 1.7$ V at high currents and high output powers (into full PA compression). Additional metrics are measured and characterized over frequency, input powers, and collector voltages including $P_{1\text{dB}}$, P_{sat}, PAE, gain, S-parameters, and noise figure[34].

The measured results do not de-embed the 0.2 dB losses of the on-chip RF input & output wirebond probe pads. When the PA is integrated into a larger transmitter system, these additional pad losses will not be present. Pad and pad-transition losses are included here because the PA has been designed to absorb the pad parasitics into the matching network designs. The large-signal measurements are much simpler in a matched 50 Ω impedance measurement setup.

Large-signal and small-signal gains quoted in the measurements are all based on transducer power gain, as described by Eq. (2.45).

$$G_T = \frac{P_{\text{load}}}{P_{\text{source,available}}} \tag{2.45}$$

The measurement results are presented for a SiGe die that represents typical, median performance over the parts harvested from the foundry 200 mm wafer. All measured W-band PA chips come from the same 8HP SiGe wafer, so no wafer-to-wafer characterization is possible. However, the foundry confirms that the delivered, measured parts represent typical process corners– the transistors and other devices are typical and not skewed "fast" or "slow".

The median-performance die is used for the data presented here. Multiple PA designs are contained within the single die, and the baseline PA design is measured on 10 bare (unpackaged and unbonded) SiGe dies via on-wafer micro-probing. The highest and lowest P_{sat} values measured across the 10 dies are approximately ±0.3 dB around the median-valued die. Small-signal gain is also within ±0.5 dB among the various dies. Some of the measurement variation can be attributed to process non-uniformity across the wafer, room-temperature variation[35], and measurement errors arising from small mechanical shifts in the mixed waveguide and coaxial probe-station setup with repeated lifting & landing of the on-wafer measurement probes.

The total power-on time for the design W1 die is > 500 hours with no measurable degradation in small-signal or large-signal performance. Devices have been tested at 2.3 V

[34]S-parameters and noise figure are done at small-signal powers only— they are not characterized over RF input powers

[35]The W-band waveguide probe station did not have thermal chuck control capabilities at the time, so all measurements are done at room temperature which logged at 20°–25°, depending on the time of day in the CalIT^2 building.

for over 50 hours, and then re-measured at lower collector voltages to determine if 2.3 V stresses changed device performance. The high-voltage does not impact the power amplifier, measurably, in RF performance. Even measurements at 2.5 V collector voltages, where P_{sat} is not increasing beyond the 2.3 V case, no permanent device damage or stress is observed in the RF performance.

2.11.1 Measurement Setups

The DC and small-signal measurements use a purely coaxial cable system for RF power transfer. The 1 mm coax cable passes DC–110 GHz and allows small-signal S-parameter measurements from 60–110 GHz. The S-parameters' upper and lower frequency ranges are limited by the network analyzer setup.

The PA measurements are taken directly on-die at room temperature (25°C) and are referenced to the differential G–S–G–G–S–G RF input/output pads. The large-signal measurements are made on the half-PA of the differential PA because of the limited output powers of the available W-band measurement sources. The additional +3 dB available power is crucial for pushing the W-band PA into compression since the overall PA gain is relatively low (< 15 dB) and input losses from the waveguide setup and 1 mm coaxial cable system is high . A +3 dB adjustment is made on the measured single-ended data to obtain the differential results, as commonly done at mmW frequencies where available RF input powers from measurement sources are limited [14, 15, 17–19, 71–73].

Because of limited available source power and other measurement complications, differential results are often obtained by adding 3 dB to single-ended measurements in differential mmW PAs. An integrated balun results in roughly 0.9–1.2 dB power loss (per balun), which impacts the available source power to drive the PAs into compression. In a separate later work in SiGe 9HP near 120 GHz [52], Ken H.-C. Lin, a design-group colleague, measured and de-embedded a similar differential PA from identically-mirrored halves and found that the differential output power is indeed +3 dB over the single-ended measurements within 0.2 dB measurement error. After de-embedding balun losses, the gain and PAE of the differential PA with integrated input & output baluns matched that of the measured single-ended PA with +3 dB data adjustments, as expected. Pfeiffer et al. in [14] also verified experimentally (with external input & output waveguide baluns) that adding +3 dB output power to single-ended measurements matched the results of fully differential measurements with external baluns de-embedded. A careful symmetric differential layout with isolated & well-matched halves allows any common-mode to differential-mode conversion to be ignored and makes adding +3 dB to single-ended output-power measurements fair when the PA is intended to drive

differential loads.

The expected differential PA output power is +3 dB that of the measured single-ended power. The differential PA gain and PAE are the same as the half-PA single-ended values— no adjustments are necessary. Measurement results include all pad transition losses on the chip— the RF pad and pad-transition losses are not de-embedded from the results. Many of the large-signal measurements are still source-power limited above 92 GHz, even with single-ended measurements and simplified (reduced loss) input-measurement chains.

Although the PA is not measured differentially, the full (both sides) of the PA is powered on and biased to verify that the two PA halves sides are thermally isolated in terms of DC & RF performance. Heat generated and dissipation from one half of the PA does not measurably affect the performance of the other PA half, and there is no thermal instability observed. Moreover, the thermal effects from power dissipated in one stage of the PA does not affect the other stages. (This is verified by monitoring any changes in DC quiescent currents of Q2 when selectively turning on & off the DC biasing of Q3.) The local heating effects in each aggregated power-cell HBT dominate the thermal effects.

Powering on the entire differential does not change the measured results of the single-ended measurements, indicating sufficient thermal isolated between the two PA halves. Both PA halves are tested independently with identical results within standard measurement errors (0.25 dB). No thermal runaway or instability is seen under zero RF input drive (DC quiescent point) when power dissipation within the HBT transistors are high since no DC power is being converted to RF output power. No thermal instability is observed under maximum, high RF-input powers, as well. Additional temperature effects are not tested because the W-band waveguide probe station does not have the proper thermal regulation equipment at the time.

The small-signal measurements are semi-automated using custom Agilent/Keysight VEE/HPIB scripts, but the large-signal measurements require manual, hand-tuning of the input power sources with various dial micrometers. Each Gunn diode oscillator source must be tuned with at least one manual micrometer screw gauge to shift the frequency. In the case of the wideband diode, two micrometers require simultaneous tuning with frequency changes. The output power delivered to the measurement setup is controlled by a manually-controlled variable attenuator with micrometer dial. Adjusting the manual attenuator sweeps the input power into the PA device under test (DUT).

The G–S–G on-chip RF micro-probes are calibrated for small-signal S-parameter measurements using TRL/LRL (thur-reflect-line, line-reflect-line) traceable to known NIST (National Institute of Standards and Technology) standards with the GGB Industries CS-5 calibration substrate. Appropriate adjustments are also taken to ensure that the RF and DC probes are

properly leveled. Calibration results in a reference plane at the input and output RF probe tips. The PA's on-chip G–S–G wirebond pad and transition losses (0.2 dB back-to-back) are included in the measurement data– pad losses and effects are not de-embedded. These additional pad losses and parasitics may be removed when the PA is integrated into a monolithic IC transmitter system.

Large Signal Setup

The W-band large-signal measurement setup is shown in Fig. 2.75. For more power in maximum P_{sat} measurements, non-essential waveguide components are removed from the measurement input-chain to increase the available RF input power to the PA chip.

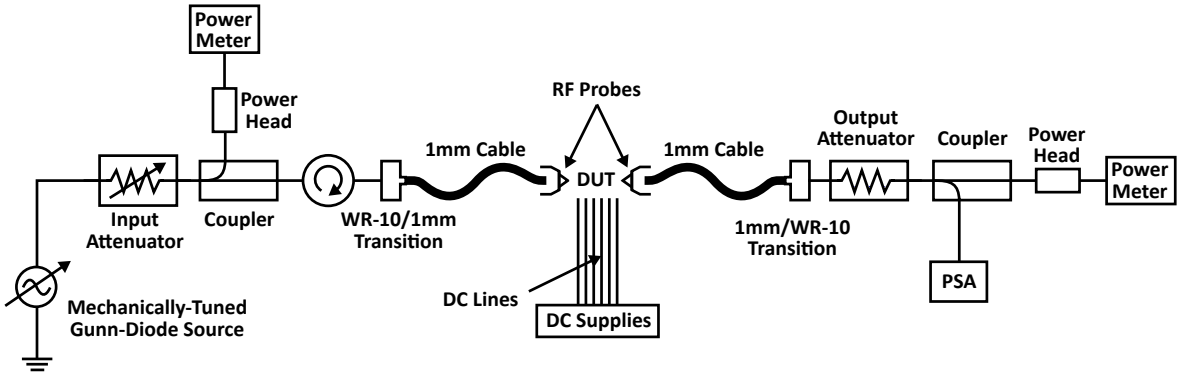


Figure 2.75: W-band large-signal measurement setup with WR-10 waveguide and 1 mm coaxial components for on-chip micro-probing.

Mechanically-tuned Gunn-diode oscillators with waveguide outputs and manually-adjusted variable waveguide attenuators provide the necessary input power sweeps for large-signal testing. Multiple Gunn-diode oscillator sources are required in the test setup since no single source, which is usually quite narrowband, can cover the entire 75–110 GHz W-band frequency range with sufficient output powers to push the W-band PA into full compression. Four separate sources are used:

1. J.E. Carlstrom Co. H101 with manual frequency tuning and manual back-short power tuning: 90–105 GHz with $\sim$ 10–75 mW (10.0–18.8 dBm) output powers (The design based is on [113].)

2. Millitech GDM-10-6-17 with manual frequency tuning from 90–94 GHz and $\sim$70 mW (18.5 dBm) output powers.

3. Millitech GDM-10-6-16 with manual frequency tuning from 79.5–83 GHz with ∼40 mW (16.0 dBm) output powers.

4. Millitech GDM-10-6-13 with manual frequency tuning from 74.5–78.5 GHz and ∼50 mW (17.0 dBm) output power.

Above 92 GHz, the large-signal measurements of P_{1dB}, P_{sat}, and peak-PAE are limited by the available source power, depending on the PA biasing conditions. A manually adjusted attenuator is necessary to control the output power after the Gunn-diode oscillator sources since there is no amplitude control in the source, only minor power tuning in the case of the Carlstrom wideband source.

The large-signal measurement is fully calibrated by carefully characterizing each component separately and also together as complete input/output chains using both small-signal S-parameters as well as power measurements with couplers and power detectors.

For measurements that require higher input powers such as P_{1dB}, P_{sat}, and peak-PAE, non-essential input components are gradually removed from the input chain to increase the available RF input power to the DUT. The input variable attenuator, coupler, and isolator are removed as needed. The available input power without these components is known precisely through individual component characterization & calibration.

For the maximum P_{sat} measurements, non-essential components in the input chain (like the variable attenuator, coupler, isolator, and extended waveguide bends) are removed to maximize the available measurement input power to drive the PA into compression as much as possible. The large-signal system is also re-calibrated to characterize the adjusted losses.

Low-loss WR-10 waveguide (0.100 inch × 0.050 inch waveguide-opening) components are used for the large-signal power measurements and noise figure measurements. 1 mm coaxial cables[36] provide a DC–110 GHz RF interface to the Cascade Microtech Infinity Probe micro-probes with 1.00 mm coaxial connectors. (The 1 mm coax measurement-grade cables are still lossy compared to WR-10 waveguide components.) The microprobes are configured in G–S–G (ground-signal-ground) with 100 μm pitch and have DC contact resistances $< 0.05\ \Omega$ when probing the on-chip aluminum RF-pads.

The waveguide attenuator pad sits at the at the output of the PA DUT. The attenuator prevents any stray reflections from small mechanical mismatches in the waveguides father down the measurement chain from interfering with the DUT since reflections are reduced by 2× the attenuation factor before reaching the DUT. The attenuator effectively presents a very stable, wideband 50 Ω load to the PA output. The attenuator also reduces the PA output powers to levels that the power detector/meter can safely handle.

[36] 1.0 mm describes the connector size.

At the input, a circulator is used instead of an attenuator to manage any stray reflections in the input chain. The low available power from the Gunn-diode sources prohibit using attenuator pads in the input chain because of the high insertion loss.

A spectrum analyzer at the output is used to tune the source frequency of the Gunn-diode oscillators and to check the wideband stability of the PA. Harmonic content cannot be observed directly because of the limited waveguide cutoff frequencies near W-band. The PA is closely examined for stability from DC–110 GHz using a purely coaxial cable measurement setup, which is lossier but wideband compared to waveguide setups. Out of band stability is checked on the spectrum analyzer, particularly at lower microwave frequencies (< 22 GHz) and sub-GHz frequencies for bias stability.

Custom DC Probes

DC biasing is done using a custom, made-to-order DC multi-contact wedge-probe board from GGB Industries (MCW-26) with separate force and sense DC lines built into the microprobes, 1 mil[37] tungsten needle points, 180 pF capacitors very close to the needle tips, and additional 0.1 μF capacitors in series with 5 Ω de-Q resistors (for wideband response) on the probe board for additional bypass filtering of each DC-force needle. The custom DC needle and board are illustrated in Fig. 2.76.

There are 3 dedicated force & sense probe pairs on the high-current collector pads to adjust the V_{CC} collector voltages on the power-HBTs precisely. The force & sense needles contact each other electrically at only two places: at the on-chip collector DC pad and within the external DC power supply. The force and sense lines, along with DC power supplies that support the sense & source feature, ensure that the voltage delivered to the PA chip pad is precisely what is desired. Tungsten DC probes, when oxidized, can have contact resistance on the order of 1 Ω when contacting aluminum pads, and the voltage drop over the DC bias lines can be significant at high PA current draws (resulting in I-R drops of hundreds of mV or more).

Separate ground pads with their own dedicated DC-ground needles provide additional external-bias stability at low frequencies (MHz range). Separating all the bias lines provides extra isolation to prevent any low-frequency oscillations loops in the test setup. Additional very-large capacitors (μF to mF range) to are placed around the external DC-power supply terminals for extra bias filtering of extremely low frequency (KHz-MHz range).

The custom DC needle probes with on-needle capacitive filtering provide the necessary off-chip filtering for PA test bench since the such large capacitances cannot be integrated on

[37] 1 mil = 0.001" = 25.4 μm.

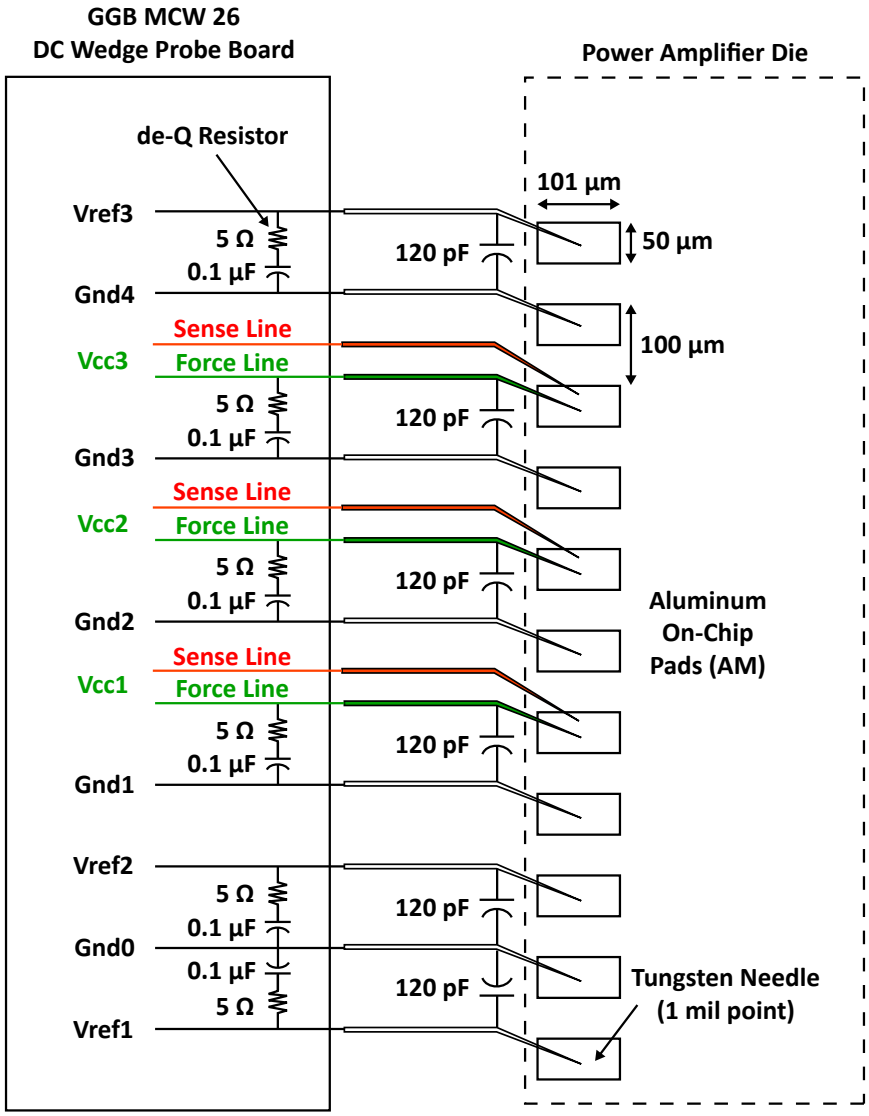


Figure 2.76: Custom DC probe contact wedge with force & sense collector lines.

chip and are normally added at the board level with a dedicated printed circuit board (PCB) measurement provision with 0201 & 0204 SMT external bypass-capacitor components.

2.11.2 Measurement Results for Baseline Design W1

The baseline W1 design operates nominally at 1.7 V collector voltage and 1.5 mA/μm peak-f_T/f_{MAX}/MAG biasing currents. For the nominal-biasing current density case (1.5 mA/μm), the reference voltage for each PA stage's active bias is individually adjusted such that in the half-PA, Q1, Q2, and Q3 operate at 15 mA, 30 mA, and 60 mA, respectively. Differences in expected/simulated bias currents (when the active-bias unit-cells are operating at a fixed reference voltage) arise from different mutual-thermal coupling coefficients for each transistor-packing configuration, as discussed in Section 2.5.6. The manual bias adjustments compensate for process variations in the transistors (e.g., β_F) and resistors; and for modeling errors in the temperature effects of self-heating and mutual self-heating within the aggregated power transistors.

Optimized current densities explore biasing points that lead to increased output powers (that are not modeled in the simulations beyond $V_{CC} > 1.7$ V). The PA can be biased down in current at higher collector voltages to put the operation into deeper Class AB for more efficiency and more output powers (since the PA contains proper the harmonic terminations). Optimizing the currents and reference voltages also allow some limited tuning of the output-power match and the interstage gain matching, which affects the gain shape across frequency and power. Unless specified otherwise, all measurement results are made at room temperature at 25°C.

The RF input and output pad (G–S–G) parasitic effects are included and absorbed into the input and output matching network designs. The 50-Ω input and output pads have 0.1 dB loss per pad (0.2 dB back-to-back total) with > 22 dB input return loss within W-band. The input and output RF pad losses are included in all measurements– not de-embedded. In designs with additional system integration, these pads, associated parasitics, and loss of 0.2 dB can be eliminated.

At nominal peak-f_T quiescent biasing currents (1.5 mA/μm) for all gain stages, the baseline W1 differential PA (full PA) consumes a total quiescent power of 415 mW (244 mA) from a 1.7 V supply at nominal current densities. For the baseline W-band design (W1), the measured peak small-signal gain (transducer gain) is 14.5 dB at 86 GHz, with a 3dB-gain bandwidth of 79.3–96.8 GHz and a fractional bandwidth of 20%. The peak large-signal performance occurs at 90 GHz with a saturated output power of $P_{\mathrm{sat}} = 17.6$ dBm (58 mW) and a peak PAE of 9.9%. The output P_{1dB} at 90 GHz is 14.8 dBm (30 mW).

The graphs of the measured results for the W1 baseline design and other experimental breakouts are presented in the following results sections.

Back-Modeled Simulations

Simulations results in the following measurement graphs are based on some back-modeling of measured S-parameter data. Design breakouts from the prior Ka-band work and from the W-band measurement data allow well-constrained back-modeling because of the unit-cell design in the baseline PAs. Small shifts in many passive-component expected values, either arising from process bias or from modeling errors, can be back-modeled. For example, a PA design breakout with a known 20% reduction of transmission-line length (or MIM capacitor value) can be directly compared to the baseline control case and other breakouts. Modeling the known difference in expected component values tunes out the process-variation bias in the back-models.

The unit-cell approach to the baseline PA design and its breakouts allows well-constrained back-modeling and simplification of a large multi-variable parameter space since relative matching of the passive components, all done in the back-end-of-line metal, is better than 1%. The W-band re-uses many similar unit-cell structures and circuit-blocks from the Ka-band designs, and the Ka-band measurement data provides additional parameter-space reduction/constraints when analyzing simulation correspondence to measured data across many PA breakouts.

Only small-signal results and DC-bias matching, which includes updated mutual thermal-coupling coefficients for each aggregated power-cell HBT, are back-modeled. Large-signal results are not back-modeled into the simulation results. All PA breakouts share the same back-modeled simulation parameters. Each PA is not back-modeled separately, but all of them are used together, as a group, to strictly constrain the number of possible variables in the back-model extraction process.

W1 Small-Signal Results

Fig. 2.77 shows the small-signal S-parameters[38] at Class A biasing at collector voltages $V_{CC} = 1.7$ V and nominal peak-f_T biasing of 1.5 mA/μm current density.

The small-signal gain 3-dB bandwidth spans 79–97 GHz, a fractional bandwidth of 20%, with the midband frequency near 88–90 GHz. The small-signal gain peaks at 14.6 dB at 86–87 GHz and is greater than 5 dB throughout the entire 75–110 GHz W-band.

[38]In this text, charted S-parameters such S_{21} are plotted as power: $10\log_{10}|S_{21}|^2 = 20\log_{10}|S_{21}|$ (in dB).

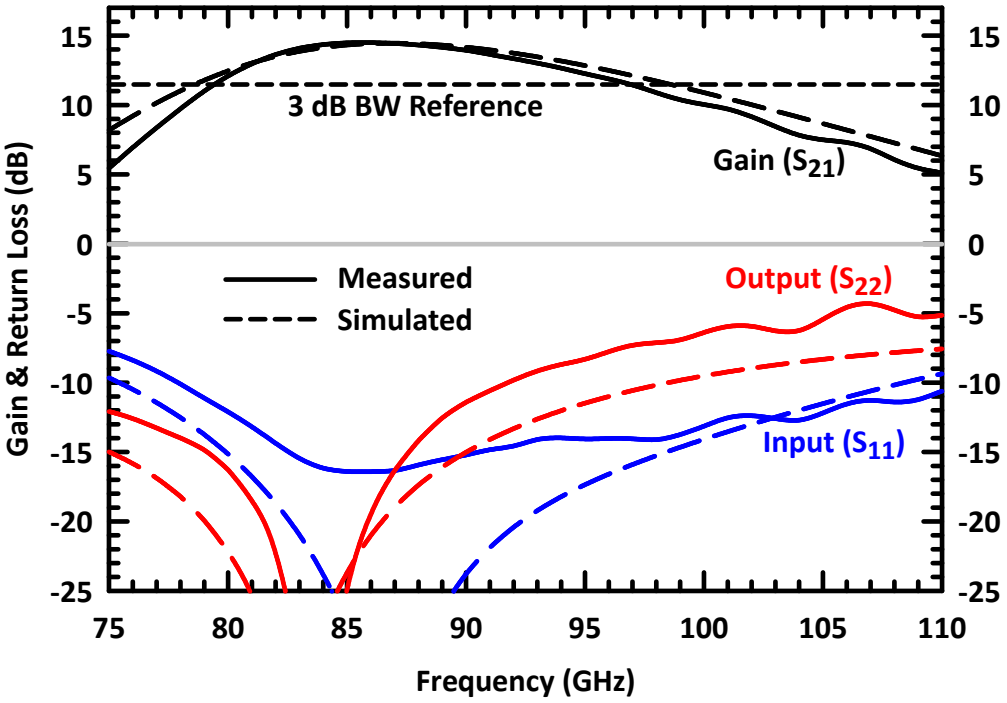


Figure 2.77: Design W1 (90 GHz) measured (solid) and simulated (dashed) S-parameters at nominal biasing currents with $V_{CC} = 1.7$ V.

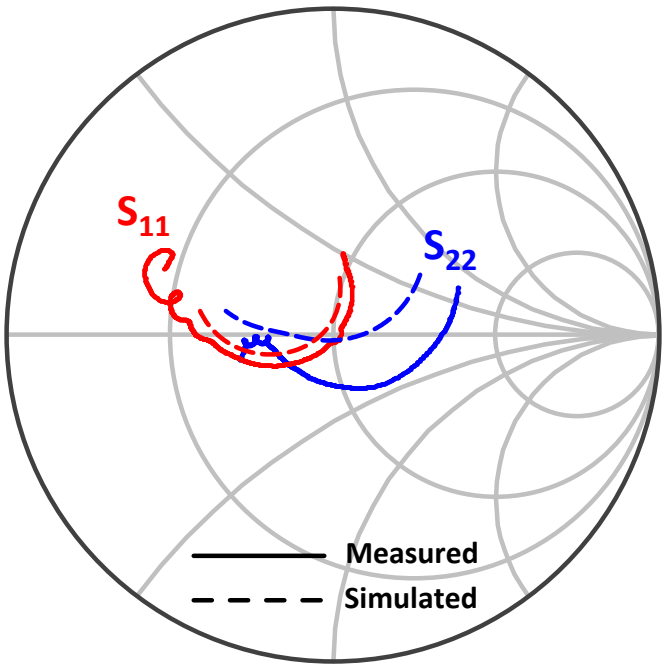


Figure 2.78: Measured (solid) versus simulated (dotted) small-signal input S_{11} and output S_{22} for the baseline W-band PA design W1 at nominal biasing currents and $V_{CC} = 1.7$ V.

In these PA designs, the peak small-signal gain tends to be about 5% lower in frequency than that for peak output power P_{sat}. This is normal for Class AB and even purely Class A designs because of the shift in HBT impedances with increasing RF powers— the two frequencies for peak gain and peak power rarely coincide exactly. The PA has a high enough gain and power bandwidth such that the small offsets in peak performance frequency do not affect its overall wideband performance.

W1 Group Delay Results

The group delay of a PA is an important metric of phase distortion over frequency as well as a measure of transit times of signals (at different frequencies) through the PA. Group delay effects can cause non-uniform distortions (with frequency) in wideband signals. Signal distortions arising from group delay are especially important in systems such as wideband communications systems and monopulse radars. Group delay (in seconds) is given by Eq. (2.46).

$$\tau_g = -\frac{d\phi}{d\omega} \tag{2.46}$$

where ϕ is the phase response of the amplifier (in radians) and ω is the frequency (in rad/sec).

The phase slope of the PA's measured S_{21} is used to find the corresponding group delay versus frequency in Fig. 2.79. The PA's high-pass, 2-element compact matching networks results in low absolute group delays and little group-delay variation across the PA operating bandwidth. The group delay is relatively flat over a wide bandwidth. The measured absolute group delay is < 46 psec throughout W-band, has very little in-band ripple, and varies 26 psec within W-band from its 46 psec peak at 82 GHz to its 20 psec low at 110 GHz.

The measured group delay is considered very short for a PA at W-band since series combiners are not required at the transistor level for power combining/splitting between unit cells within large power devices. Low absolute group delay is achieved by designing all quarter-wavelength transmission lines and other distributed components (short-circuited stubs) to be placed in shunt of the primary signal path and not inserted in series. Low group-delay variation and ripple is achieved by using similar high-pass structures in the input, output, and interstage matching networks as well as wideband quarter-wavelength resonators for high-impedance bias insertion over large bandwidths.

W1 Large-Signal: Nominal Biasing at 1.7 V

The baseline W1 design has large-signal operation centered at 90 GHz. Fig. 2.81 plots the large-signal input-power sweep for the W-band baseline PA, design W1, operating at 90 GHz

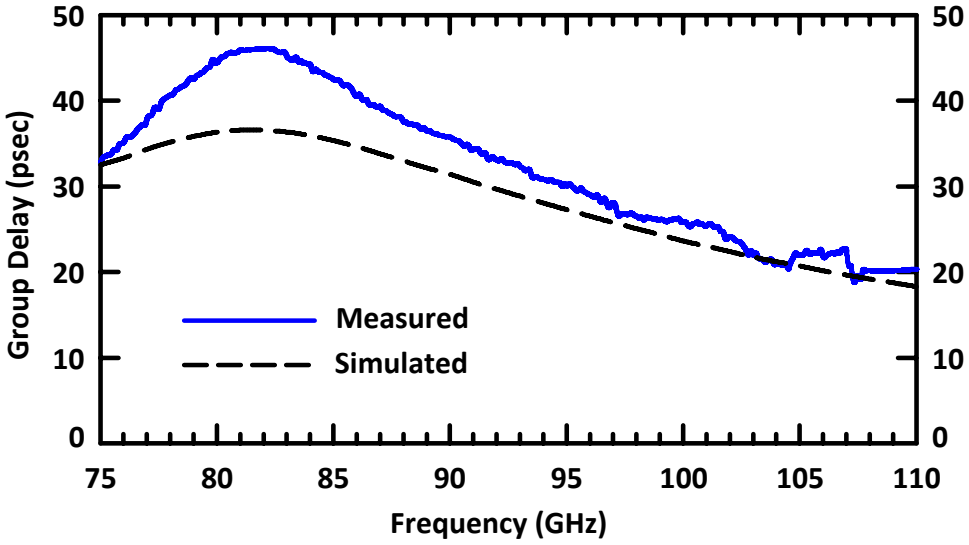


Figure 2.79: Measured vs. simulated W-band PA group delay.

with nominal 1.5 mA/μm biasing currents and V_{CC} = 1.7 V. Since baluns could not be used because of limited available source power in the measurement setup, the plotted output powers are adjusted +3 dB from measured single-ended values to reflect expected powers under differential operation. At nominal biasing currents and V_{CC} = 1.7 V, the differential output power is 17.6 dBm at 90 GHz, and the peak PAE is 9.9%. The small-signal gain is 14.5 dB.

Fig. 2.81 plots the measured and simulated large-signal performance over W-band frequencies for baseline design W1 at nominal biasing currents and V_{CC} = 1.7 V. The quiescent DC power consumption is 414 mW for the full differential PA from a 1.7 V supply. P_{1dB} is 14.8 dBm at 90 GHz. Above 103 GHz, the measurement setup does not have enough available source power to drive the PA into full compression to achieve P_{sat}. The PA maintains its large-signal performance over its 3 dB-gain bandwidth.

W1 Large-Signal: Optimized Biasing at 2.3 V

Since large-signal simulation, using the initial VBIC HBT models, have severe large-signal convergence issues for $V_{CC} > 1.7$ V, optimizing biasing currents in each gain stage for 1.7–2.3 V Class AB operation must be done experimentally during the measurements in the lab. At higher collector voltages, peak-f_T and gain occurs at slightly lower than expected DC current densities (around 10%), changing the HBT operational impedances and matching-network effects. Reducing the collector currents I_C at high V_{CC} collector voltages biases the PA deeper into Class AB operation and increases the peak PAE efficiency at higher collector voltages.

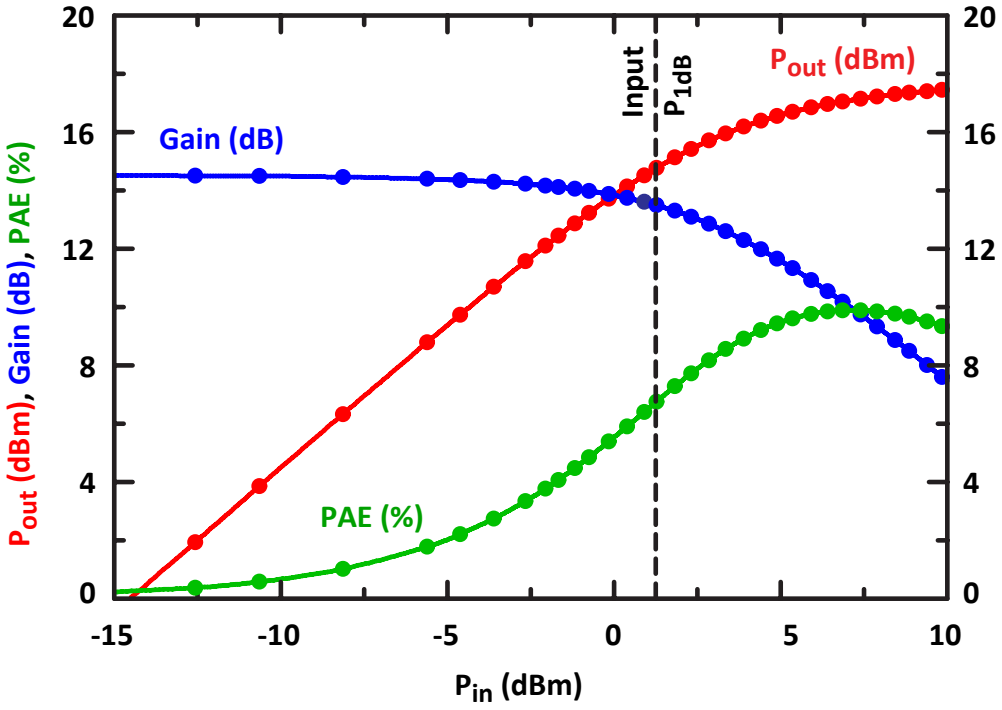


Figure 2.80: Large-signal input-power sweep at 90 GHz at nominal biasing with $V_{CC} = 1.7$ V for W-band PA baseline design W1.

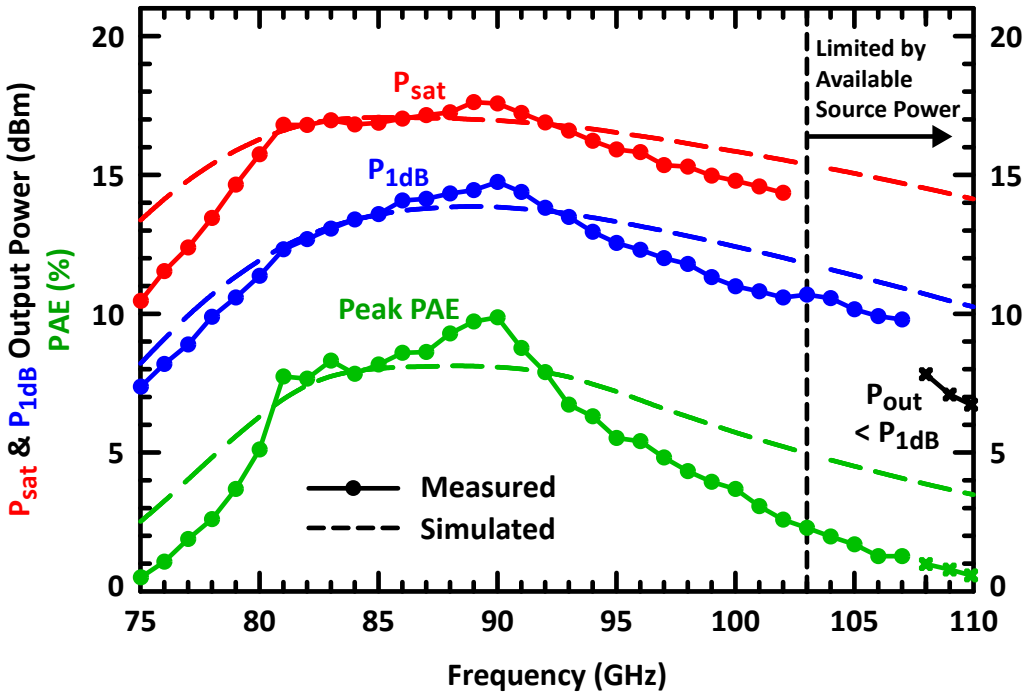


Figure 2.81: Measured (solid) and simulated (dashed) large signal results for nominal W1 design with $V_{CC} = 1.7$ V.

The PA design allows the collector voltage to be increased beyond 1.7 V because the quiescent currents in each transistor stage can be adjusted independently. This allows optimal tuning of the amplifier's performance. The optimized cases use lower DC quiescent-current densities in the PA gain stages with values ranging 0.3–1.5 mA/μm (2–12.5 mA/μm^2), resulting in a PA which is biased in Class AB operation for higher PAE and output powers.

Fig. 2.82 presents the large-signal performance at 90 GHz as a function of collector voltage for the W-band baseline PA design W1. The collector voltage for the W-band PA is raised from 1.7 V to 2.3 V in 0.2 V steps The dotted lines show the measured data for nominal current densities— transistors biased at peak-f_T quiescent currents for Class A/AB operation; and the solid lines show measured data for optimized current densities. Each of the three gain stages has DC quiescent currents fine-tuned individually to maximize overall P_{out} and PAE for the optimized-biasing cases.

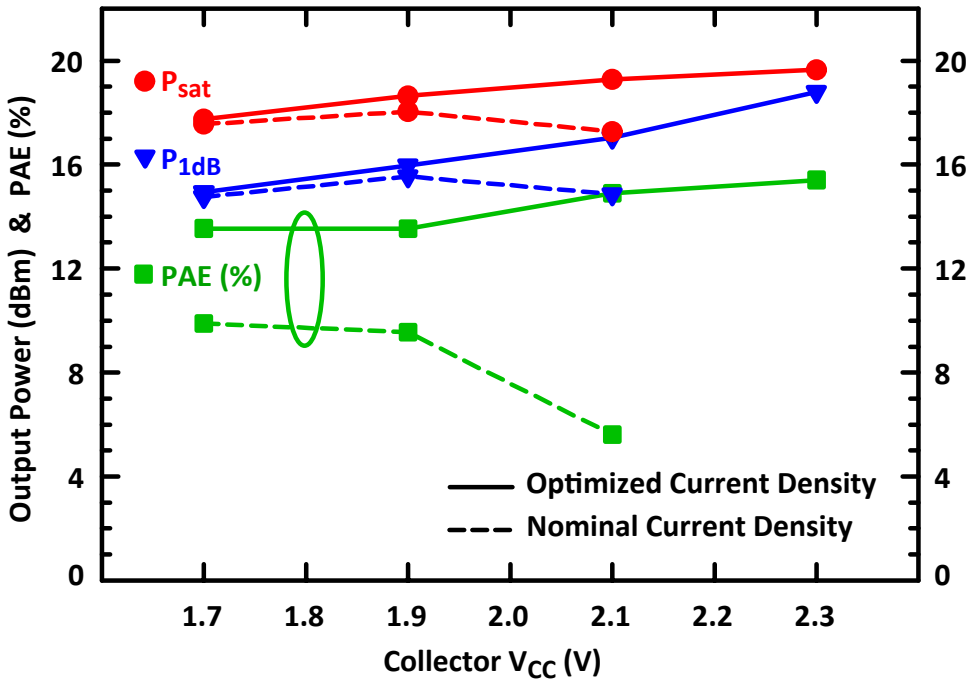


Figure 2.82: Measured large-signal characteristics vs. the collector voltage at 90 GHz (mid-band). Dotted lines show results with biasing at nominal current densities. Solid lines show results with optimized bias currents for each stage (see text).

Increasing the collector voltage beyond 2.3 V does not result in any appreciable increase in maximum output power or efficiency. Beyond 2.5 V, impact ionization effects begin to limit the PA, and the maximum P_{sat} for the design plateaus— no additional power can be successfully pushed from HBTs.

At 90 GHz and under optimized-bias current conditions, a maximum output power

P_{sat} = 19.6 dBm (91 mW) is achieved at 2.3 V with a maximum PAE of 15.4%. The corresponding output power at the 1 dB gain-compression point is P_{1dB} = 18.8 dBm (76 mW). There is no noticeable degradation in large-signal or small-signal performance after > 500 hours of testing from 1.7 V to 2.3 V on the same chip.

Fig. 2.83 plots the large-signal responses over frequency for the baseline W1 90 GHz design at V_{CC} = 2.3 V with optimized bias currents. Measurements above 93 GHz are limited by the available source power. The maximum measured P_{sat} is 19.9 dBm at 90 GHz, and the maximum P_{1dB} is 18.8 dBm at 90 GHz. Peak PAE is 15.4%.

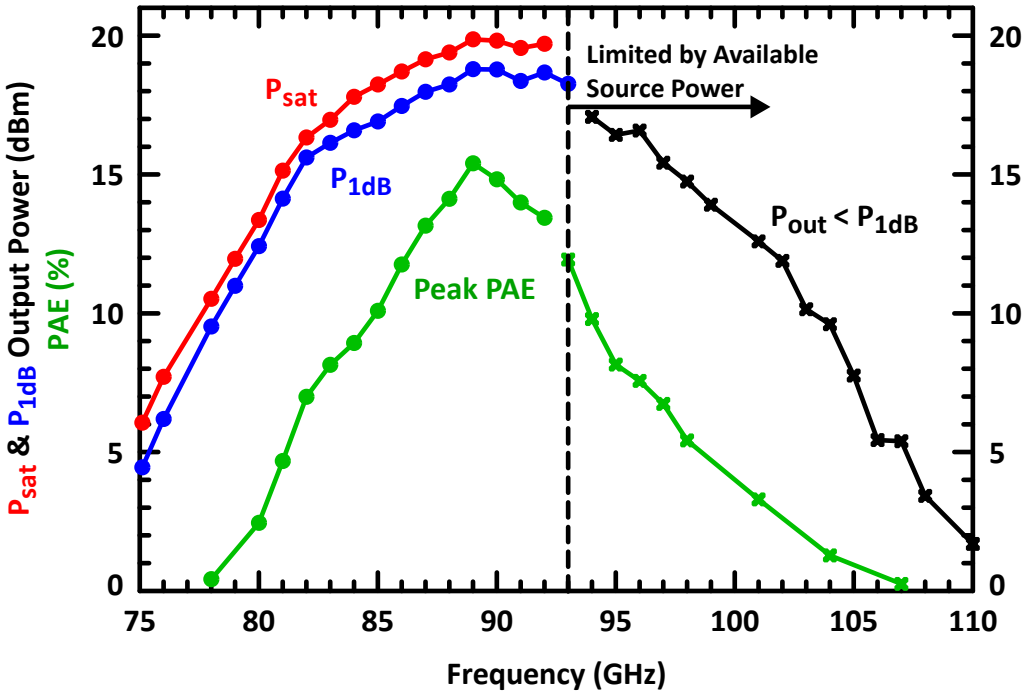


Figure 2.83: Design W1 (90 GHz, baseline) large signal results with optimized biasing at V_{CC} = 2.3 V.

Summary of maximum/peak values for baseline design W1 at 90 GHz with V_{CC} = 2.3 V: maximum P_{sat} = 19.9 dBm, P_{1dB} = 18.8 dBm, peak-PAE = 15.4%, 11.0 dB small signal gain, 3-dB small-signal bandwidth[39] from 79–97 GHz (18 GHz or 20%).

W1 Large-Signal vs. V_{CC} = 1.7–2.3 V

Fig. 2.84 summarizes the measured large-signal P_{sat}, P_{1dB}, peak-PAE, and output collector efficiency $\eta_{C,\text{Q3}}$ versus collector voltages V_{CC} = 1.7–2.3 V at 90 GHz. Results at nominal

[39]The 3-dB small-signal bandwidth is determined from the upper & lower corner frequencies where the small-signal gain, S_{21}, drops 3 dB from its peak value.

biasing currents are marked with crosses while results with optimized biasing currents are plotted as lines.

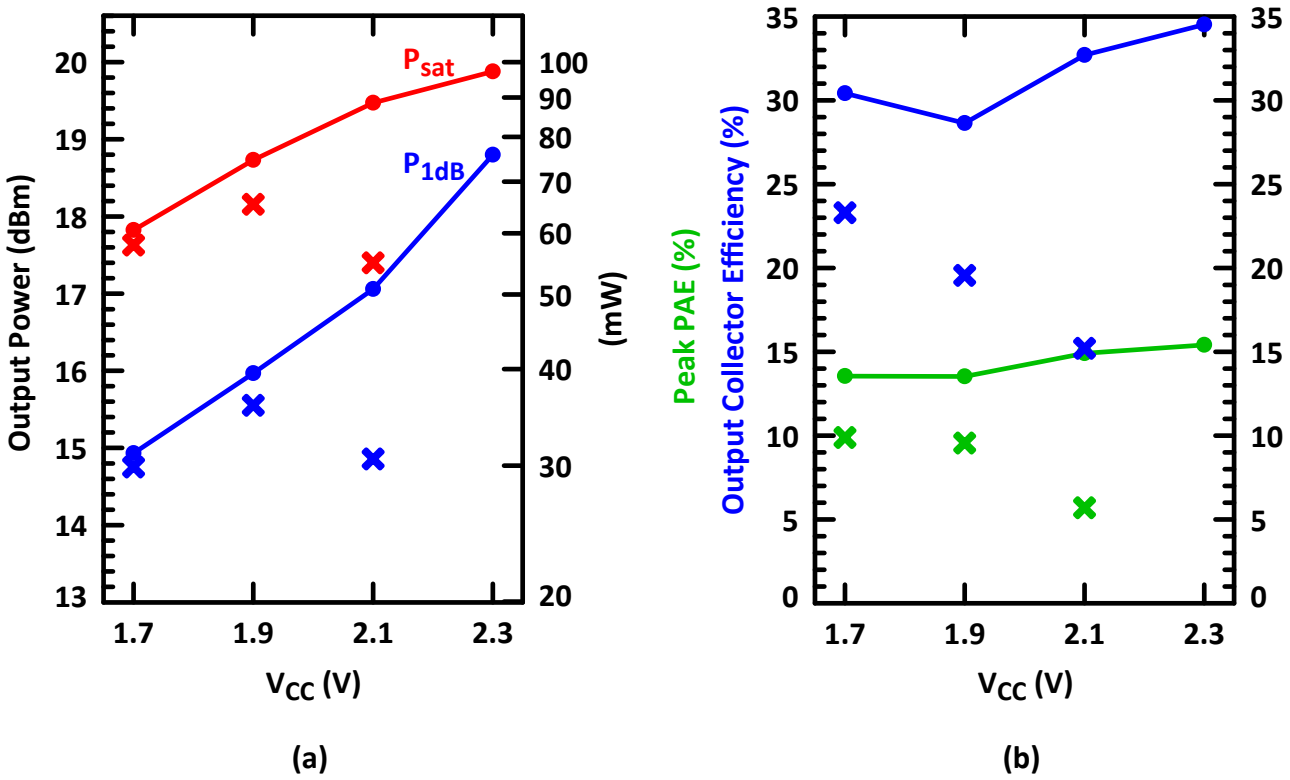


Figure 2.84: Design W1 (90 GHz, baseline) measured large-signal data with nominal (crosses) and optimized (lines) biasing currents. (a) Maximum output powers vs. V_{CC} and (b) Peak efficiencies vs. V_{CC}.

The output-collector efficiency in Eq. (2.47) is calculated for each W-band PA and provides valuable insight into the isolated efficiency of the Q3 output transistor and its output-matching network.

$$\eta_{C,\mathrm{Q3}} = \frac{P_{\mathrm{out}}}{P_{DC,\mathrm{Q3}}} \tag{2.47}$$

$\eta_{C,\mathrm{Q3}}$, provides a metric of power efficiency of the final output stage, ignoring the requirements of input source power to the output Q3 HBT, the effects of low transistor gain in the output stage, and the active bias-reference network associated with Q3, and all other gain stages & bias reference power dissipations. η_C is a useful indication of the overall efficiency of the output stage: the matching network, chosen load line, collector bias-insertion network, and HBT operation point. The theoretical η_C of an ideal Class A output transistor with inductive biasing is 50% and higher for Class AB and higher-order classes of operation.

2.11.3 Measurement Results for Breakouts W2, W3, and W4

W2, W3, and W4 Small Signal Results

Design W2 applies a -15% shrink in lengths to the two transmission-line stubs in 2-element interstage matching networks to move the design to 94 GHz where there are many W-band applications. The shift moves the 90 GHz PA to 94 GHz while maintaining the output transistor matching. Fig. 2.85 plots the resulting measured small-signal response of design W2.

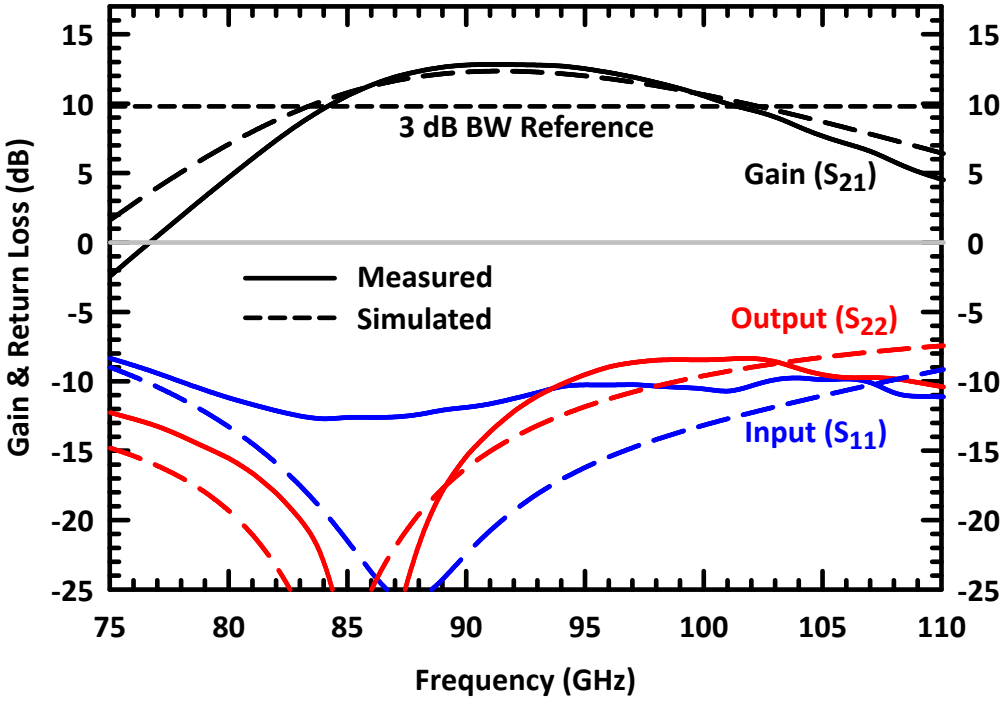


Figure 2.85: Design W2 (94 GHz) measured (solid) and simulated (dashed) S-parameters at nominal biasing currents with $V_{CC} = 1.7$ V.

Design W3 applies a -15% shrink in lengths to all unit-cell transmission lines used in the matching networks: the two interstage transmission-line stubs as well as the input- and output-matching network stubs. The shift moves the 90 GHz PA to 94 GHz while shifting the output power matching, as well. Fig. 2.85 plots the resulting measured small-signal response of design W3.

Design W4 attempts to tune down the frequency and peak the large signal response closer to the low-end of W-band where there are also many mmW applications. Design W4 applies a +15% increase in lengths to all unit-cell transmission lines used in the matching networks: the two interstage transmission-line stubs as well as the input- and output-matching network

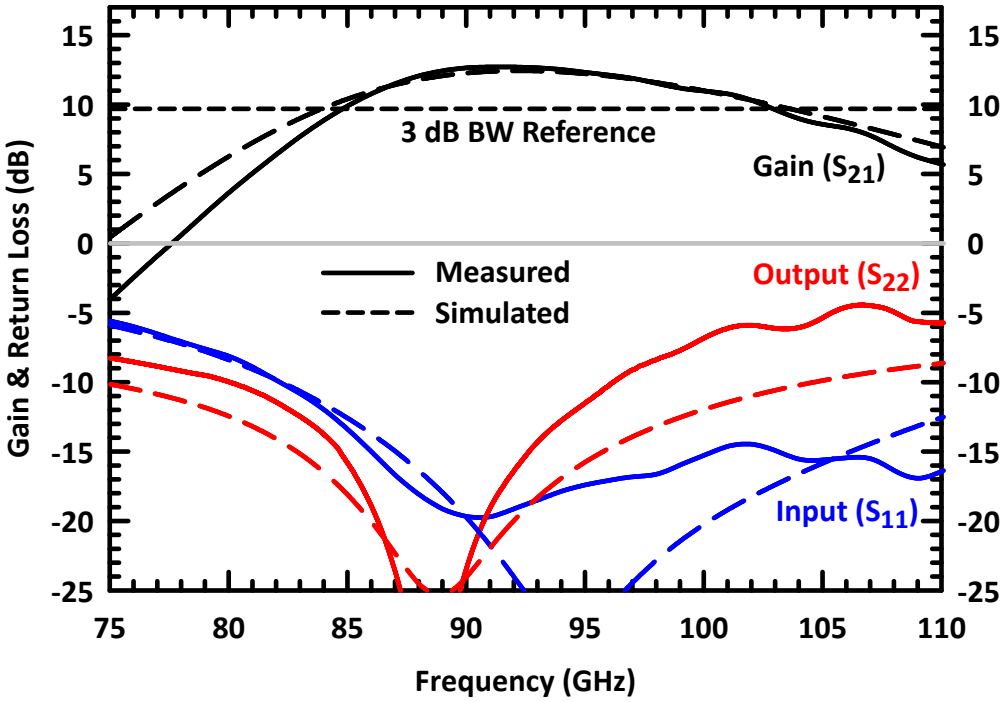


Figure 2.86: Design W3 (92 GHz) measured (solid) and simulated (dashed) S-parameters at nominal biasing currents with $V_{CC} = 1.7$ V.

stubs. The shift moves the 90 GHz PA to 83 GHz while shifting the output power matching, as well. Fig. 2.85 plots the resulting measured small-signal response of design W4.

W2, W3, W4 Large-Signal: Optimized Biasing at 2.3 V

As done with the baseline W1 90 GHz design, the other PA variations are optimized from 1.7–2.3 V operation. The results of the measurements for $V_{CC} = 2.3$ V are presented here. See Section 2.11.2 for descriptions of the bias-optimization process at 2.3 V and Section 2.11.3 for descriptions of the design changes in each breakout variation.

Fig. 2.88 plots the large-signal results for design W2 (94 GHz) under optimized biasing for $V_{CC} = 2.3$ V.

Summary of maximum/peak values for design W2 at 94 GHz with $V_{CC} = 2.3$ V: maximum $P_{\text{sat}} = 19.4$ dBm, $P_{\text{1dB}} = 18.1$ dBm, peak-PAE = 12.4%, 13.0 dB small-signal gain, and 3-dB small-signal bandwidth from 84–101 GHz (17 GHz or 18%),

Fig. 2.89 plots the large-signal results for design W3 (92 GHz) under optimized biasing for $V_{CC} = 2.3$ V.

Summary of maximum/peak values for design W3 at 92 GHz with $V_{CC} = 2.3$ V: maximum $P_{\text{sat}} = 18.7$ dBm, $P_{\text{1dB}} = 17.3$ dBm, peak-PAE = 11.5%, 12.8 dB small-signal gain, and 3-dB

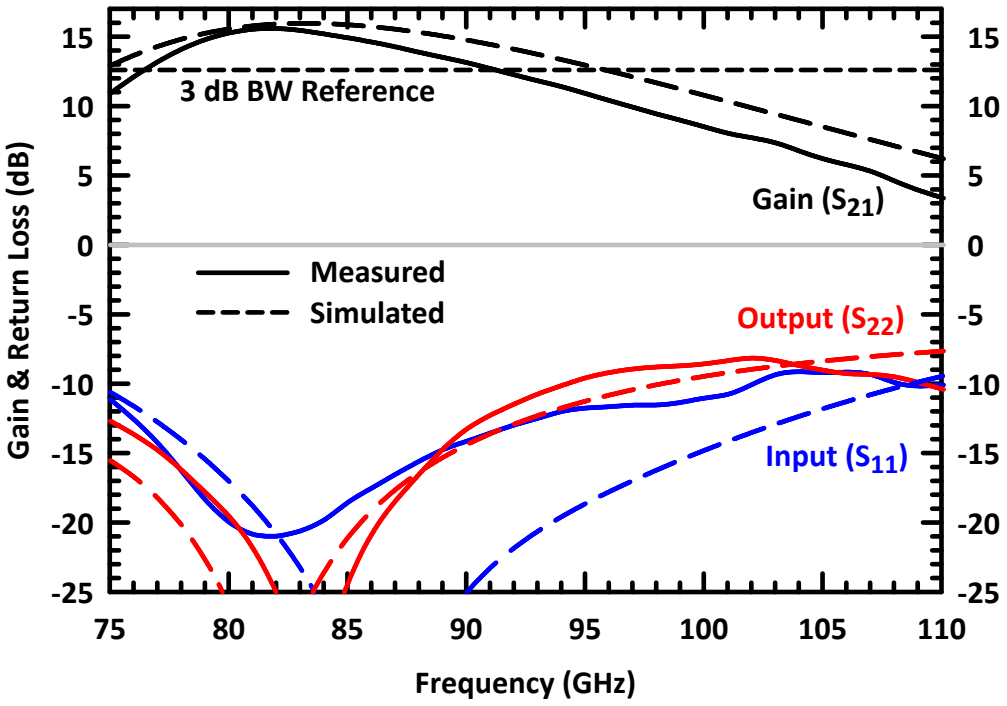


Figure 2.87: Design W4 (83 GHz) measured (solid) and simulated (dashed) S-parameters at nominal biasing currents with $V_{CC} = 1.7$ V.

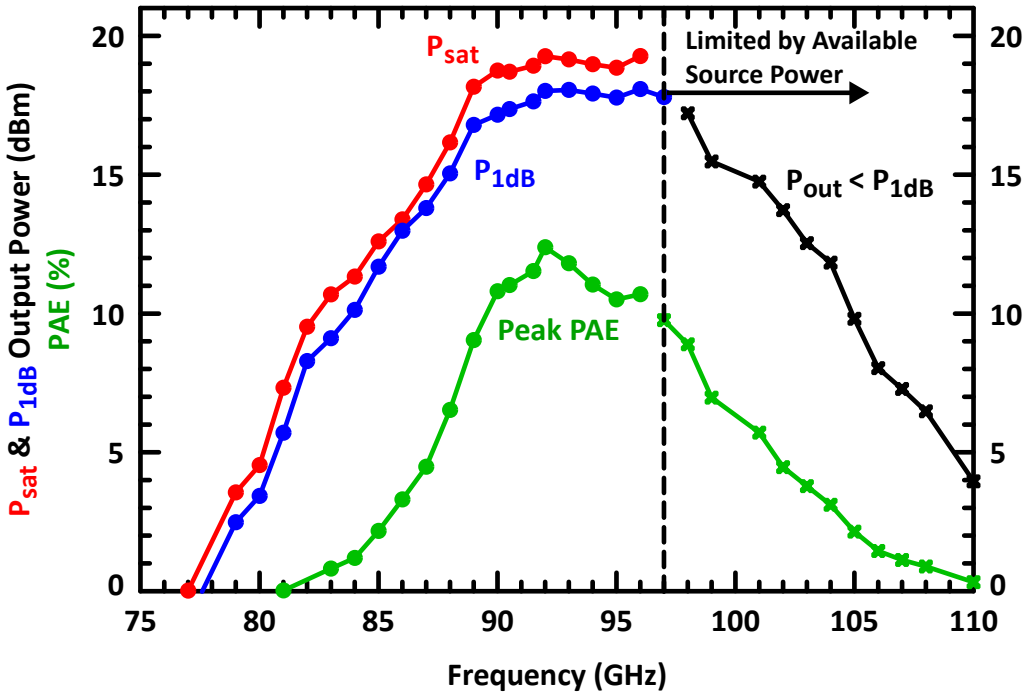


Figure 2.88: Design W2 (94 GHz) large signal results with optimized biasing at $V_{CC} = 2.3$ V.

small-signal bandwidth from 85–103 GHz (18 GHz or 19%).

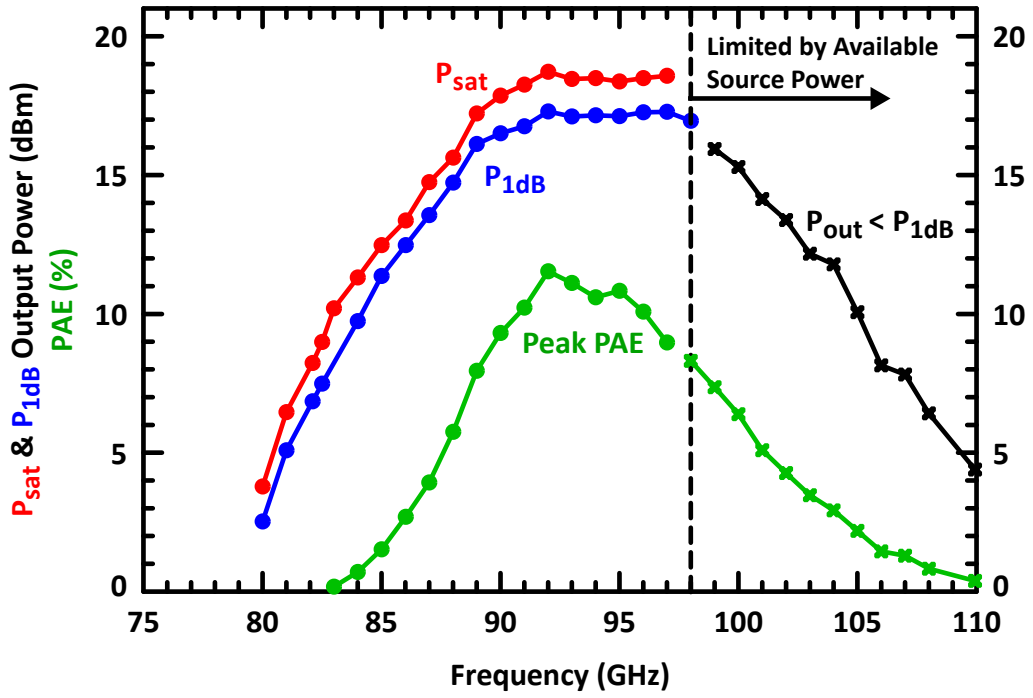


Figure 2.89: Design W3 (92 GHz) large signal results with optimized biasing at V_{CC} = 2.3 V.

Fig. 2.90 plots the large-signal results for design W4 (83 GHz) under optimized biasing for V_{CC} = 2.3 V.

Summary of maximum/peak values for design W4 at 83 GHz with V_{CC} = 2.3 V: maximum $P_{\rm sat}$ = 20.3 dBm, $P_{\rm 1dB}$ = 18.7 dBm, peak-PAE = 15.1%, 15.7 dB small-signal gain, and 3-dB small-signal bandwidth from 76–91 GHz (15 GHz or 18%),

W2, W3, W4 Large-Signal vs. V_{CC} = 1.7–2.3 V

Figs. 2.91, 2.92, and 2.93 summarize the measured large-signal $P_{\rm sat}$, $P_{\rm 1dB}$, peak-PAE, and output collector efficiency $\eta_{C,\rm Q3}$ versus collector voltages V_{CC} = 1.7–2.3 V for the W2, W3, and W4 design variations, respectively. Results at nominal biasing currents are marked with crosses while results with optimized biasing currents are plotted as lines.

2.11.4 Measurement Results for Breakouts W5 & W6

> 100 GHz PAs with MOM Capacitors

Designs W5 and W6 are experimental breakouts for the purpose of exploring the upper frequency limits of the W-band PA design without complete re-design and re-work of the

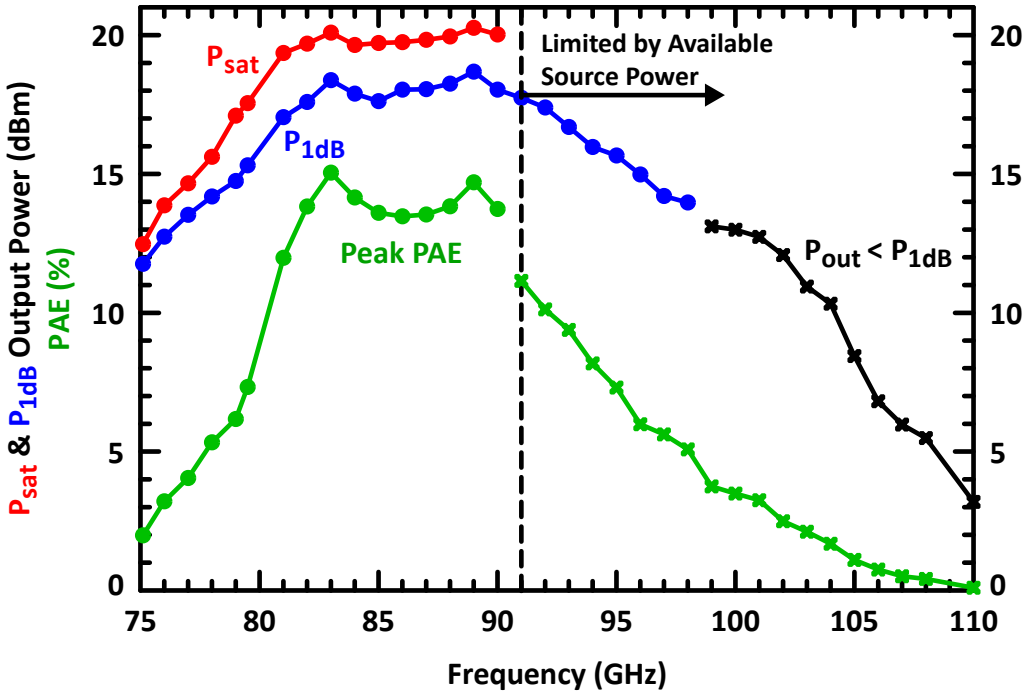


Figure 2.90: Design W4 (83 GHz) large signal results with optimized biasing at $V_{CC} = 2.3$ V.

entire chip layout. Only the matching networks are altered to shift the baseline design to higher frequencies. The quarter-wavelength collector-biasing and base-biasing transmission lines and 440 fF RF-shorting MIM capacitors & wideband bypass capacitor arrays remain the same as the baseline W1 design.

Designs W5 and W6 use M3-to-M4 custom metal-oxide-metal (M3–M4 MOM) capacitors in the resonant interstage matching networks as well as the output matching network. The custom low-density SiO_2 capacitors are formed between copper metal layer M3 and M4. The low available HBT gain (MAG $<$ 4 dB @ 100 GHz) results in difficulties driving transistors fully into compression and results in significantly reduced peak-PAE values compared to the baseline 90 GHz designs.

MOM designs pushing the W-band design closer to 100 GHz suffer from very low gain conditions as MAG continues to drop. Ideally, another gain stage and re-optimization of all quarter-wave feed lines would recover significant PA performance. The not enough available source power from the measurement setup past 95 GHz to fully characterize the large-signal performance.

In future design iterations of the $>$ 100 GHz PA designs, some re-tuning of the W5 and W6 designs is necessary to recover gain and efficiency. In particular, small adjustment should be made to all the quarter-wavelength transmission-line unit-cells (both in the collector

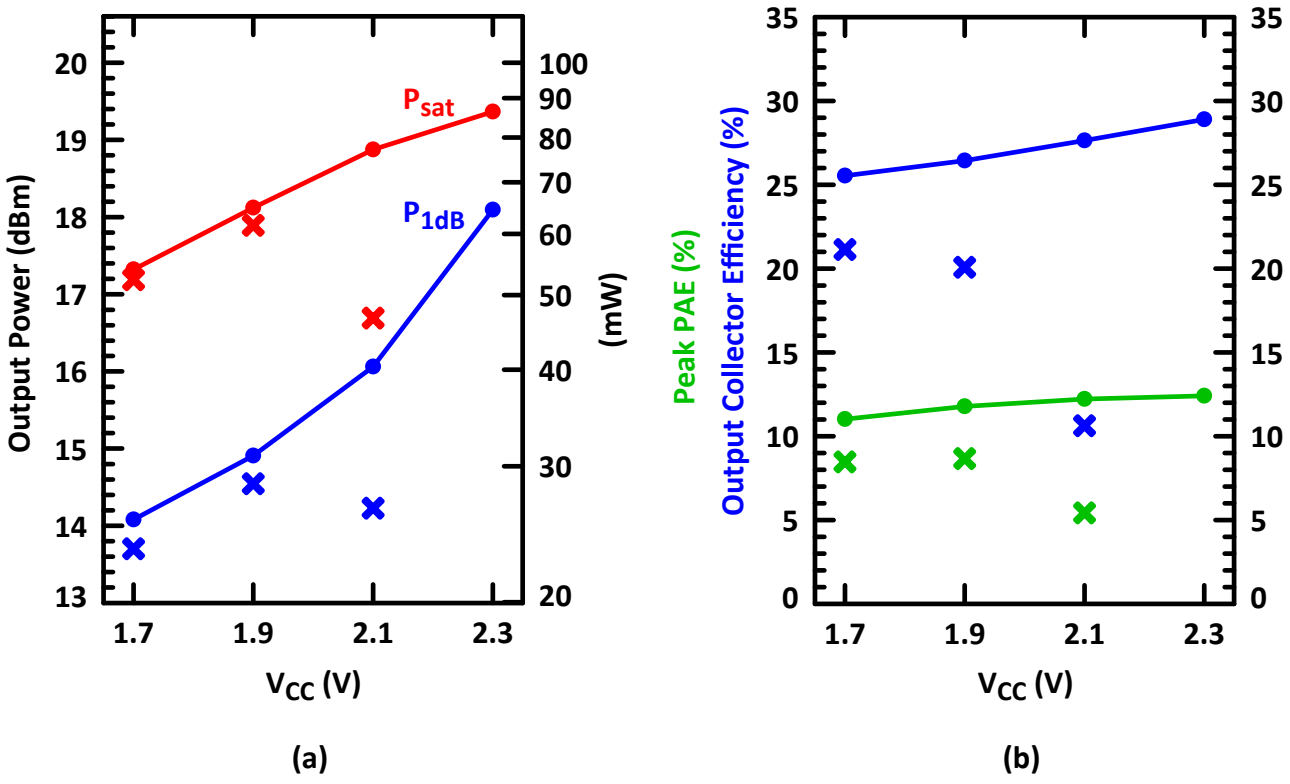


Figure 2.91: Design W2 (94 GHz) measured large-signal data with nominal (crosses) and optimized (lines) biasing currents. (a) Maximum output powers vs. V_{CC} and (b) Peak efficiencies vs. V_{CC}.

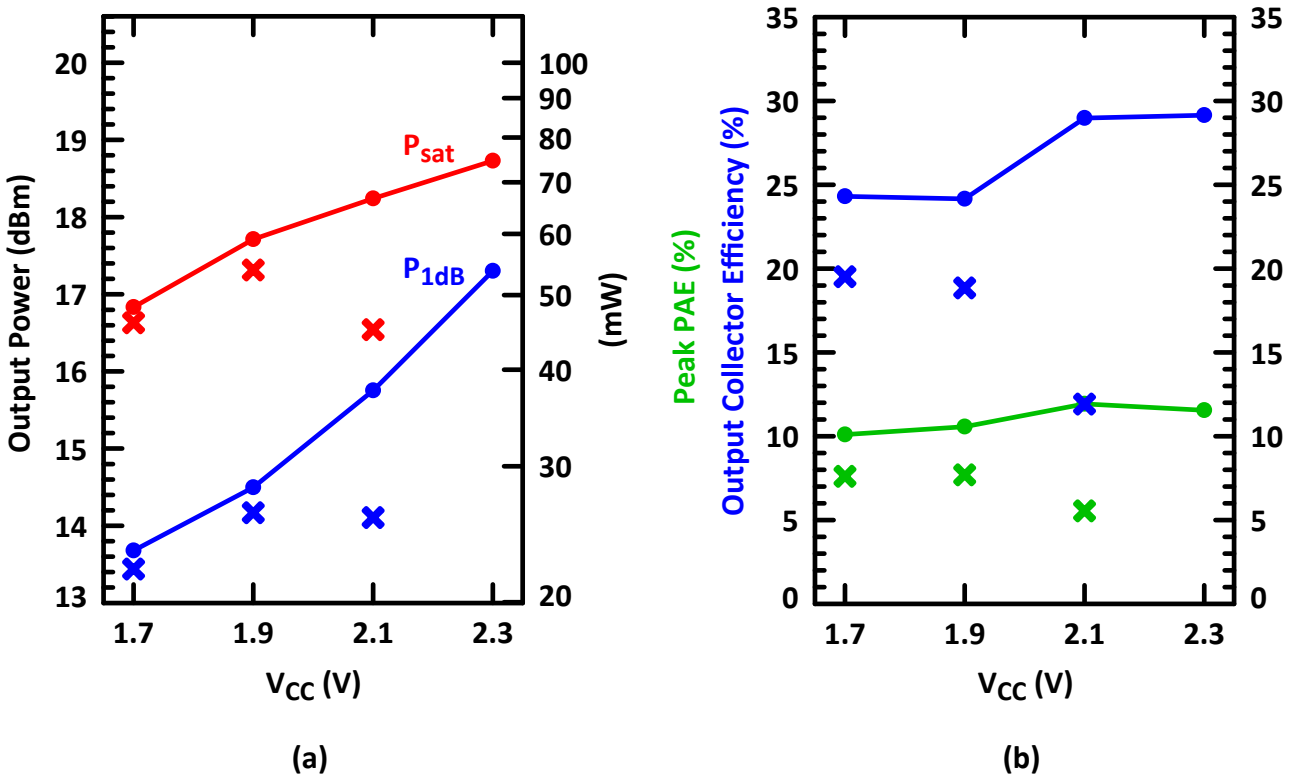


Figure 2.92: Design W3 (92 GHz) measured large-signal data with nominal (crosses) and optimized (lines) biasing currents. (a) Maximum output powers vs. V_{CC} and (b) Peak efficiencies vs. V_{CC}.

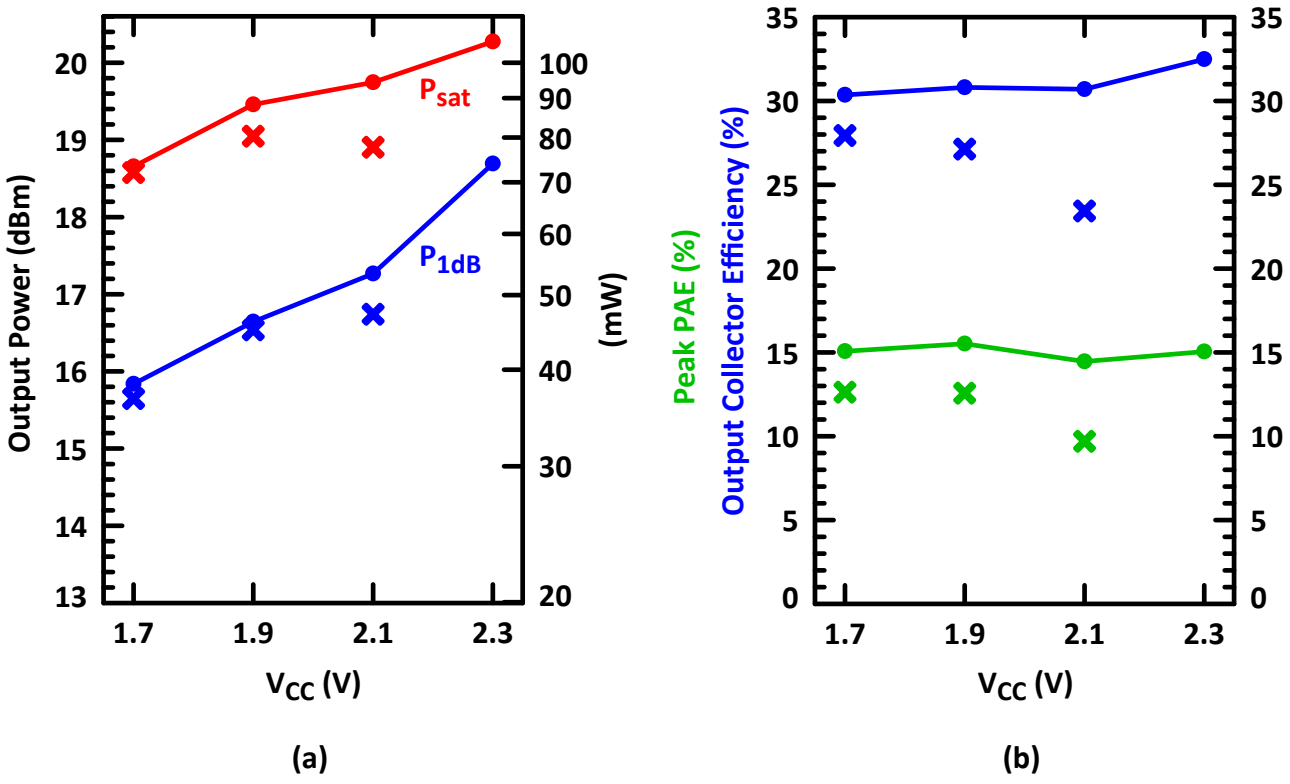


Figure 2.93: Design W4 (81 GHz) measured large-signal data with nominal (crosses) and optimized (lines) biasing currents. (a) Maximum output powers vs. V_{CC} and (b) Peak efficiencies vs. V_{CC}.

and base lines) as well as the self-resonant 440 fF MIM capacitor unit cell. Since available per-stage gain is close to only 3 dB, 105 GHz is near the upper limits of the common-emitter topology for PA design in 8HP technology.

W5 & W6 Small Signal Results

In design W5, centered at 101 GHz, the matching-network capacitors, using custom M3–M4 MOMs, are approximately 26–37% smaller in value (but 6–7× larger in area) than the baseline W-band design. Shunt-stub transmission-line inductors in the input, interstage, and output matching networks are approximately 11–27% shorter. All quarter-wavelength transmission lines and 440 fF RF-shorting capacitors remain at their same values as the baseline W1 design.

The small-signal gain for design W5 operating at nominal biasing currents and V_{CC} = 1.7 V is plotted in Fig. 2.94. The small-signal gain S_{21} peaks at 8.5 dB @ 101 GHz, and the 3-dB bandwidth is 91–108 GHz (17 GHz or 17% fractional bandwidth).

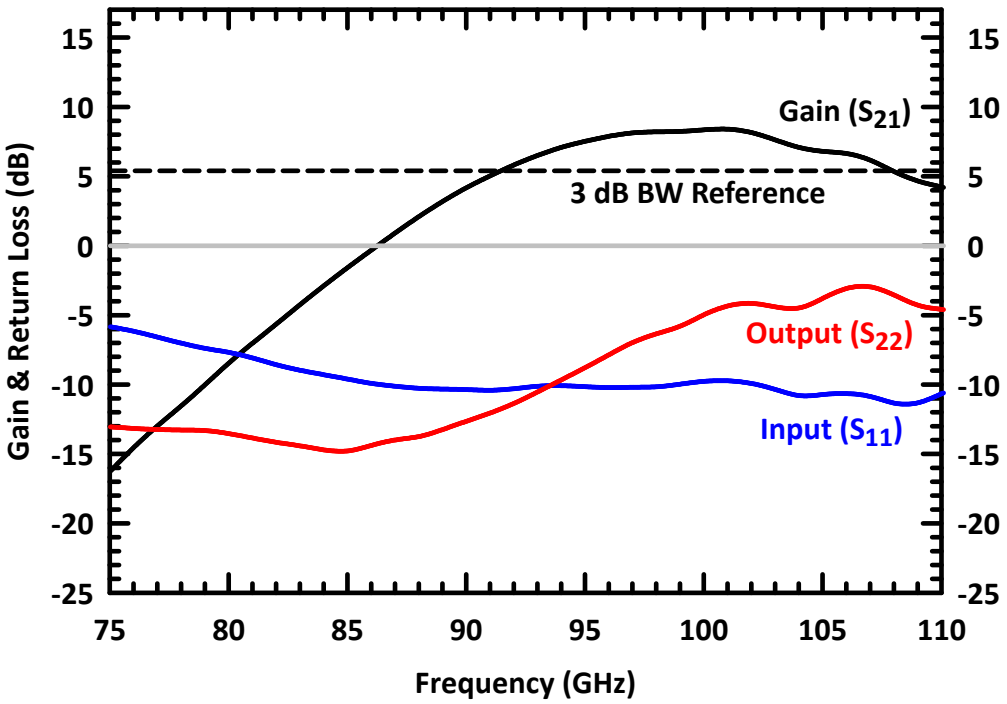


Figure 2.94: Design W5 (101 GHz) measured S-parameters at nominal biasing currents and V_{CC} = 1.7 V.

In design W6, centered at 105 GHz, the matching-network capacitors, using custom M3–M4 MOMs, are approximately 38% smaller in value than the baseline W-band design. Shunt-stub transmission-line inductors in the input, interstage, and output matching networks

are approximately 40% shorter. All quarter-wavelength transmission lines and 440 fF RF-shorting capacitors remain at their same values as the baseline W1 design.

The small-signal gain for design W6 operating at nominal biasing currents and V_{CC} = 1.7 V is plotted in 2.95. The small-signal gain S_{21} peaks at 5.5 dB @ 105 GHz, and the 3-dB bandwidth is 97 GHz to >110 GHz, limited by the available source power in the measurement setup (> 13 GHz or > 12% fractional bandwidth, limited by measurements).

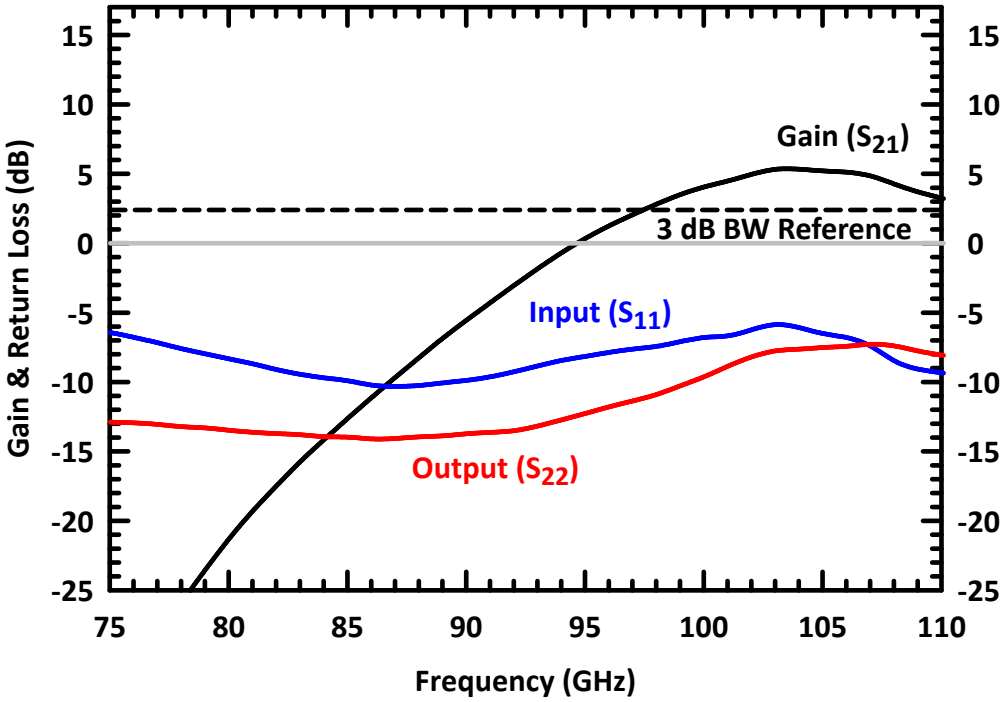


Figure 2.95: Design W6 (105 GHz) measured S-parameters at nominal biasing currents and V_{CC} = 1.7 V.

W5 & W6 Large-Signal: Optimized Biasing at 2.3 V

Fig. 2.96 plots the large-signal results for design W5 (101 GHz) under optimized biasing for V_{CC} = 2.3 V.

Summary of maximum/peak values for design W5 at 101 GHz with V_{CC} = 2.3 V: maximum P_{sat} = 16.7 dBm, P_{1dB} = 13.8 dBm, peak-PAE = 3.2%, 8.6 dB small-signal gain, and 3-dB small-signal bandwidth from 91–108 GHz (17 GHz or 17%).

Fig. 2.97 plots the large-signal results for design W6 (105 GHz) under optimized biasing for V_{CC} = 2.3 V.

Summary of maximum/peak values for design W6 at 105 GHz with V_{CC} = 2.3 V: maximum P_{sat} = 14.1 dBm, P_{1dB} = 10.6 dBm, peak-PAE = 1.3%, 5.4 dB small-signal gain, and 3-dB

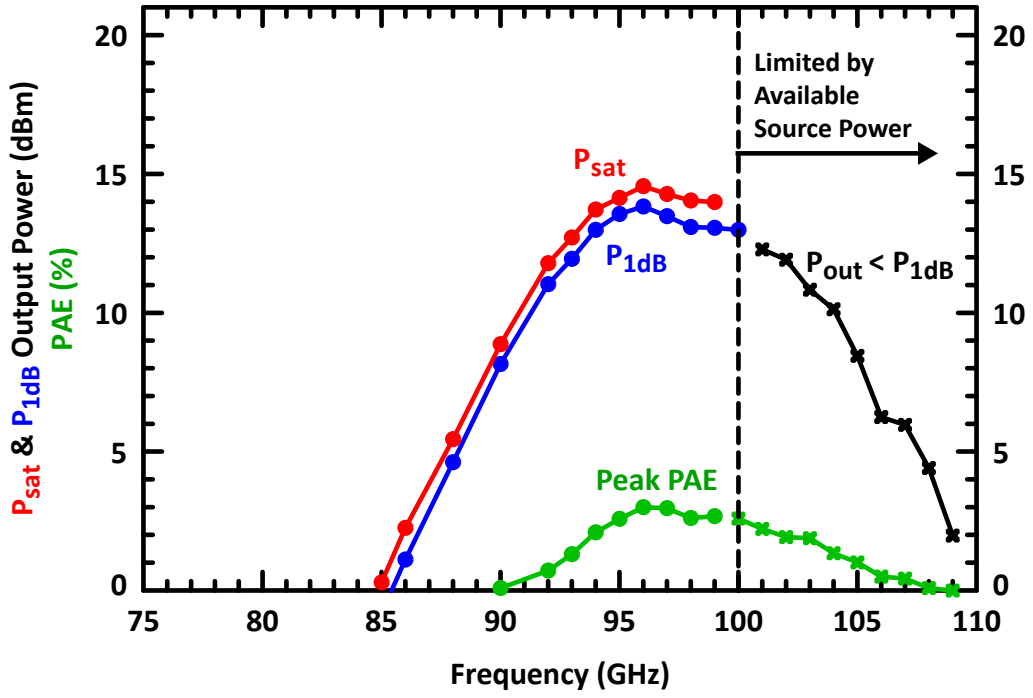


Figure 2.96: Design W5 (101 GHz) large signal results with optimized biasing at $V_{CC} = 2.3$ V.

small-signal bandwidth from 97 GHz to > 107 GHz (> 13 GHz or > 13%), which is limited by the available source power in the measurement setup.

W5 & W6 Large-Signal vs. V_{CC} = 1.7–2.3 V

Figs. 2.98 and 2.99 summarize the measured large-signal $P_{\rm sat}$, $P_{\rm 1dB}$, peak-PAE, and output collector efficiency $\eta_{C,\rm Q3}$ versus collector voltages $V_{CC} = 1.7$–2.3 V for the W5 and W6 design variations with MOM capacitors, respectively. Results at nominal biasing currents are marked with crosses while results with optimized biasing currents are plotted as lines.

2.11.5 Noise Figure Measurements

The PA noise figure (NF) is measured on the baseline W-band PA design (90 GHz design) at nominal biasing currents and $V_{CC} = 1.7$ V. The W-band noise figure measurement uses low-loss WR-10 waveguides in the setup shown in Fig. 2.100. Noise figure measurements are taken using two different methods for verification: the gain method and the Y-factor method. (See Appendix A for accompanying equations.) The results are the first known published noise-figure measurements for W-band SiGe power amplifiers.

The quarter-wave transmission-line terminated in MIM RF-short and wideband bypass

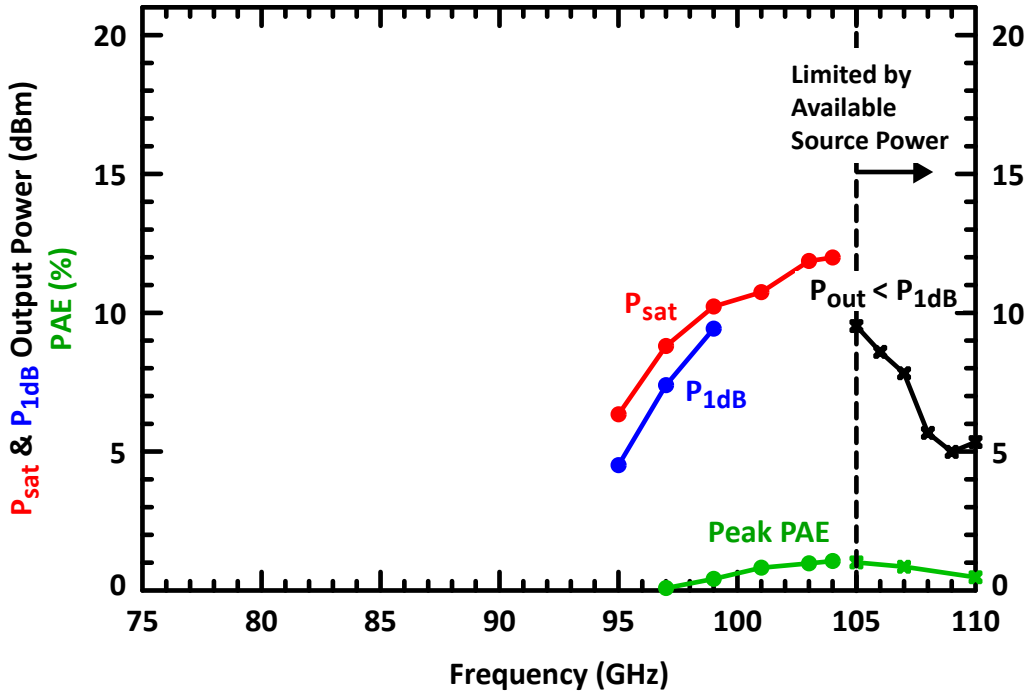


Figure 2.97: Design W6 (105 GHz) large signal results with optimized biasing at $V_{CC} = 2.3$ V.

MIM capacitors prevents bias-circuit thermal noise and shot noise from reaching the power-transistor HBT base terminal at 90 GHz. The bias-insertion technique improves the overall noise figure of the PA by > 0.6 dB compared to purely resistive biasing techniques. No other resistors are used outside of the noise-isolated bias circuits.

At collector $V_{CC} = 1.7$ V operation, the HBT minimum noise figure $\mathrm{NF_{min}} = 5.8$ dB at 90 GHz (with no additional emitter inductive degeneration) near peak-f_T biasing current densities. The $\mathrm{NF_{min}}$ value holds true for each gain stage Q1, Q2, and Q3, although practical achievable values are limited by matching-network constraints since optimal impedances becomes lower as the transistor sizes increase. Associated power gain[40] $\mathrm{NF_{min}}$ is 6.0 dB. At nominal collector operating voltages $V_{CC} = 1.7$ V, HBT avalanche noise from impact ionization is low and does not impact the transistor noise figure.

The $\mathrm{NF_{min}}$ is approximately 6.0 dB for each gain stage with 6.0 dB associated gain when the HBT input is matched for noise and the HBT output is conjugately matched (maximum power transfer). The mutual thermal coupling does increase the $\mathrm{NF_{min}}$ by approximately 0.3 dB for Q3 under nominal operating conditions at $V_{CC} = 1.7$ V.

The system-level noise-figure analysis of three cascaded gain stages with each gain at

[40] Associated gain is the small-signal gain under the condition that the HBT input is matched for $\mathrm{NF_{min}}$ while the output is conjugately matched.

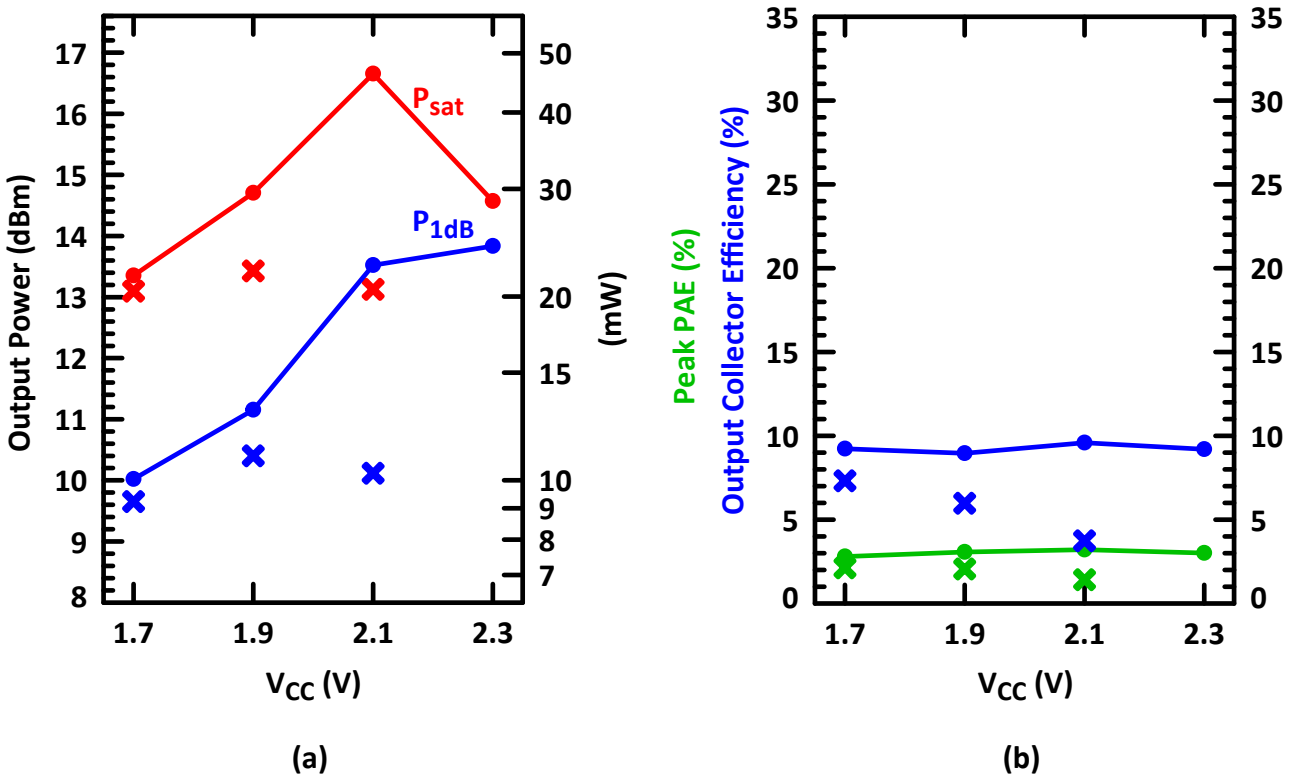


Figure 2.98: Design W5 (101 GHz) measured data with nominal (crosses) and optimized (lines) biasing currents. (a) Maximum output powers vs. V_{CC} and (b) Peak efficiencies vs. V_{CC}.

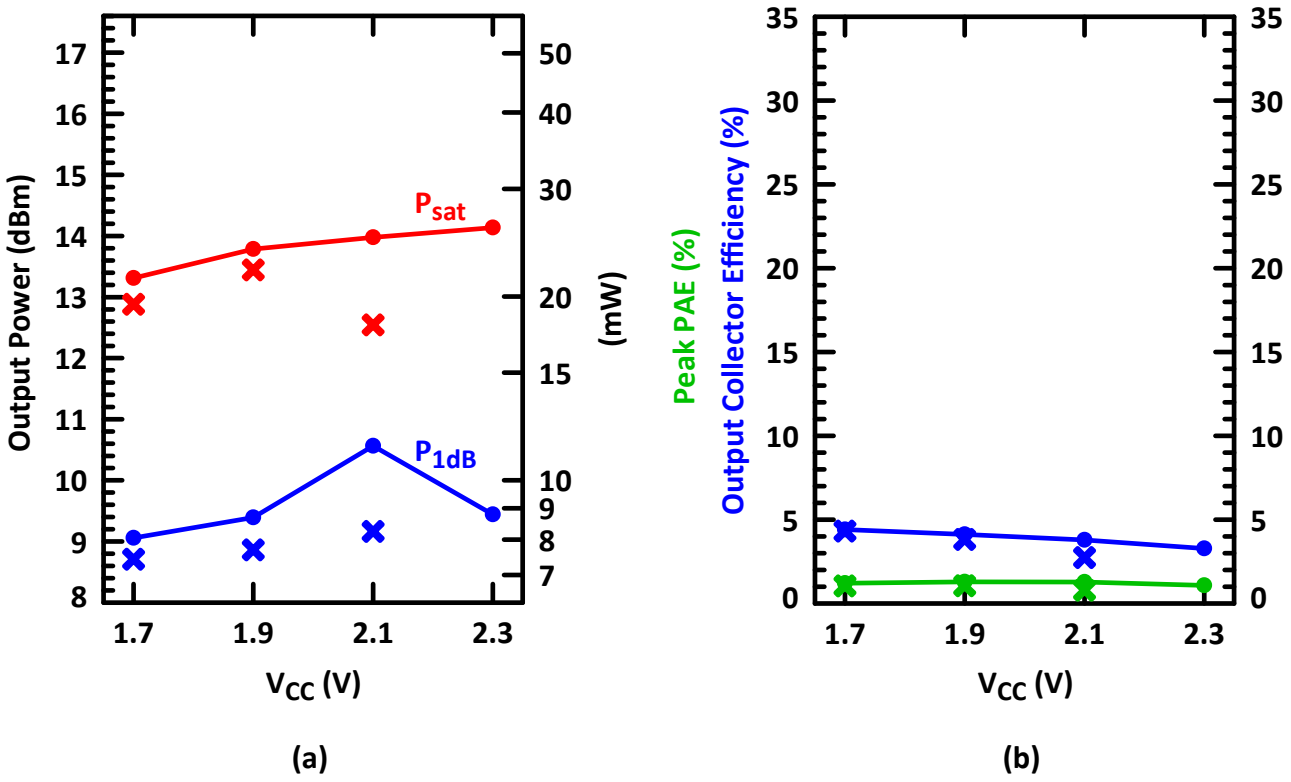


Figure 2.99: Design W6 (105 GHz) measured data with nominal (crosses) and optimized (lines) biasing currents. (a) Maximum output powers vs. V_{CC} and (b) Peak efficiencies vs. V_{CC}.

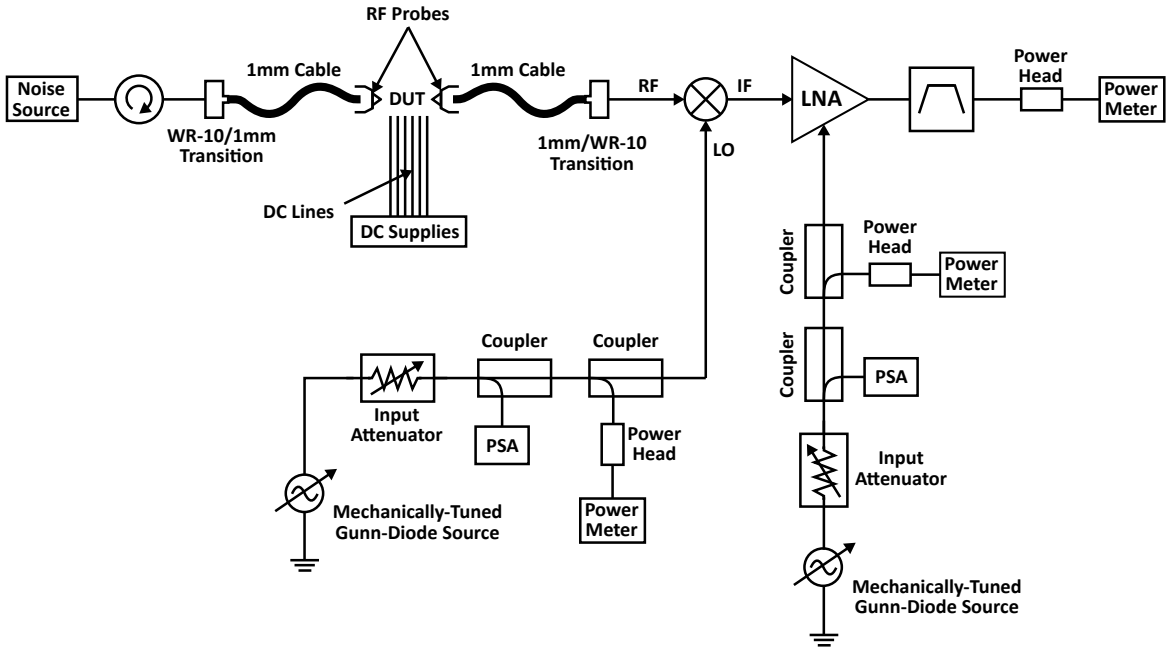


Figure 2.100: Noise figure measurement setup.

matched ideally to NF_{min} at the inputs shows that the ideal noise floor for the PA matched ideally for optimal noise performance is $NF_{input,min} = 6.8$ dB with maximum ideal cascaded gain of 17.8 dB.

The measured noise figure for the baseline W1 design at nominal biasing currents and 1.7 V is shown in Fig. 2.101. The noise figure near 90 GHz is 8.0 dB.

The W-band PA compares favorably to low-noise amplifiers (LNAs) designed in the same 8HP process. Noise figure (NF) of dedicated LNA designs within the same technology is also 8–10 dB noise figure [114]. The PA draws considerably more power to be considered a useful standalone LNA, but the low noise figure aids in the eventual transmitter design, which must take into account in-band and out-of-band transmitter noise.

2.11.6 Results Summary and Comparisons

Summary

Table 2.10 summarizes the measured W-band large-signal results[41] and peak small-signal gain for each design breakout.

[41]Quoted output powers are differential values extrapolated from single-ended measurements by adjusting +3 dB since there is not enough available source power in the measurement setup to use internal or external baluns.

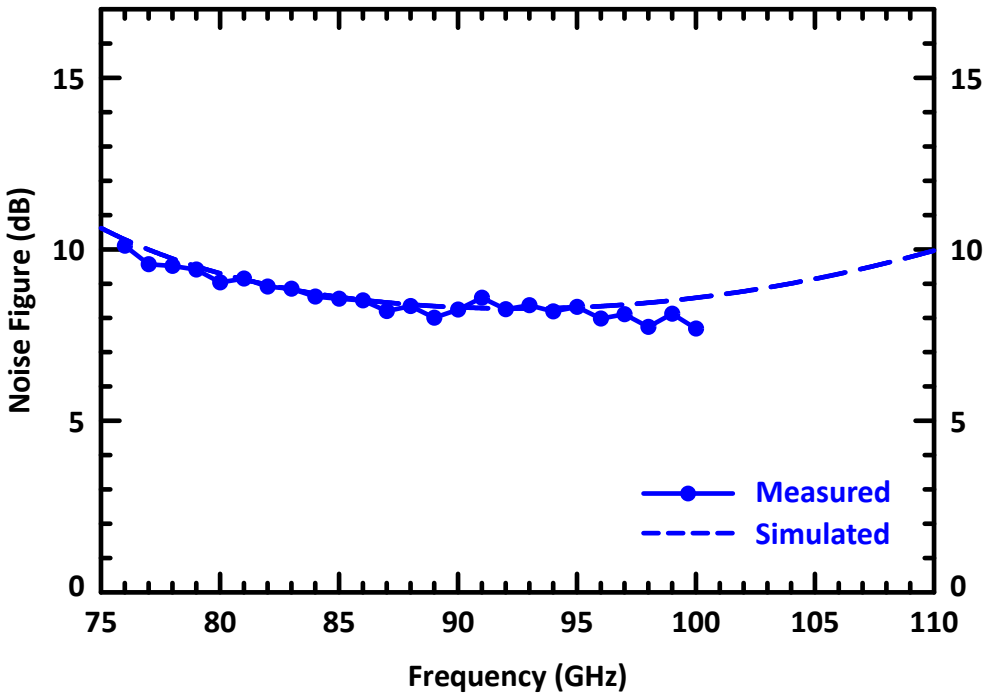


Figure 2.101: Measured (solid) and simulated (dashed) noise figure at 90 GHz for the baseline W1 design at nominal biasing currents and $V_{CC} = 1.7$ V.

Table 2.10: W-Band PA large-signal summary with optimized biasing. Gain is quoted for the peak small-signal (SS) values.

Design	**Center Freq (GHz)**	$P_{\mathbf{sat}}$ **(dBm)**	$P_{\mathbf{sat}}$ **(mW)**	$P_{\mathbf{1dB}}$ **(dBm)**	**Peak PAE (%)**	**Peak** η_C **(%)**	**Max SS Gain (dB)**
W1	90	19.9	98	18.8	15.4	34.5	14.6
W2	94	19.4	87	18.1	12.4	28.9	13.0
W3	92	18.7	74	17.3	11.9	29.2	12.8
W4	83	20.3	107	18.7	15.5	32.5	15.7
W5	101	16.7	47	13.8	3.2	9.6	8.6
W6	105	14.1	26	10.6	1.3	4.4	5.4

Table 2.11: W-Band PA large-signal bandwidths summary at nominal biasing currents.

Design	**3dB-$P_{\mathbf{sat}}$ BW (GHz)**	**3dB-$P_{\mathbf{1dB}}$ BW (GHz)**	**3dB-Gain BW (GHz)**	**Half-PAE BW (GHz)**
W1	79–101 (24%)	80–98 (20%)	79–97 (20%)	80–97 (19%)
W2	83–[>102] (>21%)	83–106 (24%)	84–101 (18%)	86–101 (16%)
W3	84–105 (22%)	85–104 (20%)	85–103 (19%)	86–100 (15%)
W4	76–99 (26%)	77–96 (22%)	76–91 (18%)	77–95 (21%)
W5	89–[>105] (>16%)	89–[>105] (>16%)	91–108 (17%)	92–[>105] (>13%)
W6	97–[>105] (>8%)	97–[>105] (>8%)	97–[>110] (>13%)	99–[>105] (>8%)

Large-signal bandwidths are listed in Table 2.11. Large-signal bandwidths are generally in the ∼20% range. Bandwidths are given for PAs operating at nominal biasing currents at $V_{CC} = 1.7$ V since there is not enough available source power in the measurement system to obtain the 3dB-bandwidths at higher collector voltages when more input power is required to compress the PAs. Designs W5 and W6, which operate above 100 GHz, do not have representative bandwidth values available since no large-signal waveguide source was available beyond 105 GHz at the time, and the network analyzer reached its frequency limits at 110 GHz for small-signal S-parameters.

The 3dB-P_{sat} bandwidth (BW) is the frequency range in which P_{sat} is within 3 dB of its maximum value. The 3dB-P_{1dB} bandwidth is the frequency range in which P_{1dB} is within 3 dB of its maximum value. The half-PAE bandwidth is the frequency range in which the peak-PAE is within 50% of its maximum value.

Result Comparisons

Table 2.12 summarizes the measured results and compares this work with prior works (before & during the time the results from this design were initially published) in silicon-based power amplifiers within W-band, including CMOS/SOI.

Operational frequency bandwidth is difficult to standardize among the comparisons since not all report small-signal gains across bandwidth, especially if the PA is biasing deep into Class AB or B, for example, since power gain will depend on the RF input power. Besides

Table 2.12: W-band PA comparison at the time of original publication

Authors	Year	Technology	Lith (nm)	f_T (GHz)	Freq (GHz)	3-dB BW (GHz)	P_{sat} (dBm)	P_{1dB} (dBm)	Peak PAE (%)	S.S. Gain (dB)	Supply (V)	Topology	Stages	PA Mode	Input SE/Diff	Output SE/Diff	Baluns (if diff)	Power Combining	Class	Comments
This Work @ Optimized Bias @ 2.3 V	2008	SiGe BiCMOS	130	200	**90**	79–97	**19.9**	18.8	**15.4**	11.0	2.3	CE	3	Diff	Diff	Diff	None	-	AB	3 dB added to SE measurements
This Work @ Optimized Bias @ 2.1 V							**19.3**	17.0	**14.9**	13.2	2.1									
This Work @ Optimized Bias @ 1.9 V							**18.6**	16.0	**13.5**	14.3	1.9									
This Work @ Optimized Bias @ 1.7 V							**17.7**	14.9	**13.5**	14.6	1.7									
This Work @ Nominal Bias					**90**	79–97	**17.6**	14.8	**9.9**	14.0	1.7								A	
This Work (Variation W2)					**94**	85–103	**19.4**	18.1	**12.4**	13.0	2.3								AB	
This Work (W5 with MOM caps)					**101**	92–108	**16.7**	13.8	**3.2**	8.6	2.3								AB	
Afshari [26]	2006	SiGe BiCMOS	130	200	**85**	NR	**21.0**	NR	η_C = **4**	8.0	-2.5 & 0.8	Cascode	8	SE	SE	SE	-	4	NR	Distributed amplifier
Komijani [21]	2005	SiGe BiCMOS	120	200	**77**	65–80	**17.5 (@1.8V)**	14.5	**12.8**	17.0	1.5	CE	4	SE	SE	SE	-	2	AB	
Nicolson [23]	2008	SiGe	130	230	**77**	NR	**14.5**	12.0	**15.7**	19.0	1.8	CE	3	SE	SE	SE	-	-	B	
Pfeiffer [14]	2004	SiGe	120	207	**77**	NR	**12.5**	11.6	**2.5**	5.6	2.5	CE	2	Diff	Diff	Diff	None	-	AB	3 dB added to SE measurements
Glisic [18]	2008	SiGe:C BiCMOS	250	200	**65**	59–62[a]	**18.9**	17.2	**9.8**	24.8	4	Cascode	3	Diff	Diff	Diff	None	-	A	3 dB added to SE measurements
					61.5		**17.9**	13.6	**7.5**	31.7										
Pfeiffer [19]	2007	SiGe BiCMOS	130	200	**64**	NR	**23.0**	NR	**6.3**	20	4	Cascode	2	Diff	Diff	Diff	External input balun	4		3 dB added to SE measurements
Pfeiffer [16]	2005	SiGe BiCMOS	120	207	**64**	NR	**13.5**	7.3	**3.1**	11.0	4	Cascode	1	Diff	Diff	Diff	External	-	AB	
					61.5		**14.0**	8.5	**4.2**	12.0										
					59		**14.5**	9.3	**4.6**	14.0										
Floyd [15]	2005	SiGe	120	200	**61.5**	NR	**16.2**	11.2	**4.3**	10.8	2.5	CE	2	Diff	Diff	Diff	None	-	AB	3 dB added to SE measurements
Reynolds [13]	2004	SiGe	120	200	**61.5**	NR	**11.8**	8.7	**NR**	10.7	1.1	CE	2	Diff	Diff	Diff	None	-	AB	
Pfeiffer [17]	2007	SiGe BiCMOS	130	200	**60**	58–65	**20.0**	13.1	**12.7**	17.0	4	Cascode	1	Diff	Diff	Diff	External input balun	-	AB	3 dB added to SE measurements

[a] Estimated from graph

3dB-gain small-signal or large-signal bandwidths, quoted operational bandwidths are often based on other non-standard conditions (e.g. large-signal gain is > 8 dB or > 5 dB; P_{sat} is greater than a certain user-chosen value, etc.).

W-Band One-Way Linear PA Comparisons: SiGe

Table 2.13 compares this work with all one-way, linear SiGe power amplifiers, current and prior, within W-band. The PAs in this table are not power combined (i.e., one-way) beyond operating differentially, if applicable. Only linear-mode, Class A/AB, Class B, and Class C SiGe power amplifiers are included– switch-mode SiGe PAs are not included.

W-Band One-Way Switch-Mode PA Comparisons: SiGe

Table 2.14 compares recent research into non-linear, switch-mode PAs (Class E) within W-band. Only SiGe-based PAs are included in the table— no CMOS/SOI. Power-combined PAs, aside from differential operation, are not included in the table.

W-Band One-Way Linear PA Comparisons: CMOS/SOI

Table 2.15 summarizes recent CMOS/SOI one-way linear PAs within W-band. Non-linear switch-mode CMOS/SOI PAs and power-combined PAs, aside from differential operation, are not included in the table.

W-Band One-Way PA Comparisons

Figure 2.102 is a dot plot of reported saturated output power P_{sat} for various one-way[42] power amplifiers in silicon technology, which includes SiGe, SiGe:C, bulk CMOS, and SOI CMOS. The presented W-band PA compares favorably with other approaches, even with some designs in the next-generation node, 90 nm SiGe ($f_T > 280$ GHz).

Figure 2.103 is a dot plot of reported PAE for various one-way, non-power-combined silicon power amplifiers, which includes SiGe, SiGe:C, bulk CMOS, and SOI CMOS.

The values for the W-band SiGe linear one-way power amplifiers in the W-band dot plots are quoted from [14, 23, 42, 43, 51, 58, 71, 115, 116, 126–128].

The values for the W-band SiGe switch-mode power amplifiers in the W-band dot plots are quoted from [59, 117, 118].

The values for the W-band SiGe power-combined power amplifiers in the W-band dot plots are quoted from [9, 21, 26, 47, 48, 51, 58, 129–132]

[42] "One-way" refers to non-power-combined outputs. Differential-output power amplifiers are considered one-way as long as they have no additional power-combining structures in the outputs.

Table 2.13: W-band one-way linear SiGe PA comparison

Authors	Year	Technology	Lith (nm)	f_T (GHz)	Freq (GHz)	3-dB BW (GHz)	P_{sat} (dBm)	P_{1dB} (dBm)	Peak PAE (%)	S.S. Gain (dB)	Topology	Stages	PA Mode	Input SE/Diff	Output SE/Diff	Baluns (if diff)	Class	Comments
This Work @ Optimized Bias @ 2.3 V	2008	SiGe BiCMOS	130	200	**90**	79–97	**19.9**	18.8	**15.4**	11.0	CE	3	Diff	Diff	Diff	None	AB	3 dB added to SE measurements
This Work @ Optimized Bias @ 2.1 V							**19.3**	17.0	**14.9**	13.2								
This Work @ Optimized Bias @ 1.9 V							**18.6**	16.0	**13.5**	14.3								
This Work @ Optimized Bias @ 1.7 V							**17.7**	14.9	**13.5**	14.6								
This Work @ Nominal Bias					**90**	79–97	**17.6**	14.8	**9.9**	14.0							A	
This Work (Variation W2)					**94**	85–103	**19.4**	18.1	**12.4**	13.0							AB	
This Work (W5 with MOM caps)					**101**	92–108	**16.7**	13.8	**3.2**	8.6							AB	
Kalantari [115]	2010	SiGe BiCMOS	130	200	**98**	91–110	**5.9**	4.9	**7.2**	12.5	Traveling Wave	8	SE	SE	SE	-	Linear Traveling Wave	
Song [58]	2014	SiGe BiCMOS	90	300	**93**	NR	**18.0**	NR	**17.2**	0.0	Cascode	2	SE	SE	SE	-	Deep AB	
Fuqan [42]	2013	SiGe:C	NR	200	**85**	NR	**16.6**	14.6	**11.8**	32.0	Cascode	3	Diff	SE	SE	Integrated	A	
Wu [43]	2013	SiGe BiCMOS	130	200	**82**	NR	**8.0**	NR	**2.3**	5.0	CE	3	Diff/SE	SE	SE	Integrated input balun	A	Input balun de-embedded; Dual-band
Lin [51]	2016	SiGe BiCMOS	90	310	**80**	68–105	**16.2**	12.2	**10.7**	22.0	CE	3	SE	SE	SE	-	A	
Demirel [116]	2010	SiGe BiCMOS	130	230	**79**	72–85	**18.0**	13.5	**8.2**	21.5	CE	4	Diff	SE	SE	Integrated	A/AB	Baluns de-embedded
Nicolson [23]	2008	SiGe	130	230	**77**	NR	**14.5**	12.0	**15.7**	19.0	CE	3	SE	SE	SE	-	B	
Pfeiffer [14]	2004	SiGe	120	207	**77**	NR	**12.5**	11.6	**2.5**	5.6	CE	2	Diff	Diff	Diff	None	AB	3 dB added to SE measurements
Demirel [71]	2009	SiGe BiCMOS	130	230	**77**	59–71	**13.0**	6.2	**2.2**	13.5	CE	4	Diff	Diff	Diff	None	AB	3 dB added to SE measurements
					65		**17.5**	13.5	**7.5**	19.0								
					60		**18.0**	13.0	**8.6**	17.3								
Glisic [18]	2008	SiGe:C BiCMOS	250	200	**65**	59–62[a]	**18.9**	17.2	**9.8**	24.8	Cascode	3	Diff	Diff	Diff	None	A	3 dB added to SE measurements
					61.5		**17.9**	13.6	**7.5**	31.7								
Demirel [72]	2009	SiGe BiCMOS	130	230	**65**	NR	**18.1**	15.4	**7.0**	21.0	CE	4	Diff	Diff	Diff	None	AB	3 dB added to SE measurements
					65	59–71	**18.0**	14.5	**7.8**	20.5								
					60		**18.8**	13.5	**9.8**	18.3								
Pfeiffer [16]	2005	SiGe BiCMOS	120	207	**64**	NR	**13.5**	7.3	**3.1**	11.0	Cascode	1	Diff	Diff	Diff	External	AB	
					61.5		**14.0**	8.5	**4.2**	12.0								
					59		**14.5**	9.3	**4.6**	14.0								
Floyd [15]	2005	SiGe	120	200	**61.5**	NR	**16.2**	11.2	**4.3**	10.8	CE	2	Diff	Diff	Diff	None	AB	3 dB added to SE measurements
Reynolds [13]	2004	SiGe	120	200	**61.5**	NR	**11.8**	8.7	**NR**	10.7	CE	2	Diff	Diff	Diff	None	AB	
Pfeiffer [17]	2007	SiGe BiCMOS	130	200	**60**	58–65	**20.0**	13.1	**12.7**	17.0	Cascode	1	Diff	Diff	Diff	External input balun	AB	3 dB added to SE measurements

[a] Estimated from graph

Table 2.14: W-band non-linear, switch-mode SiGe PA comparisons

Authors	**Year**	**Technology**	**Lith (nm)**	f_T **(GHz)**	**Freq (GHz)**	**3-dB BW (GHz)**	P_{sat} **(dBm)**	P_{1dB} **(dBm)**	**Peak PAE (%)**	**Topology**	**Stages**	**PA Mode**	**Class**
Song [59]	2015	SiGe BiCMOS	130	300	**93**	**17.7**	NR	**40.4**	14.0	Cascode	2	SE	E
						19.3	NR	**37.9**	11.4				
Datta [117]	2017	SiGe BiCMOS	90	300	**88**	**19.5**	NR	**15.9**	15.0	CE	5	SE	E
					85	**22.0**		**19.1**	17.0	2-Stacked	5	SE	
					83	**23.3**		**17.1**	18.7	3-Stacked	6	SE	
Datta [118]	2016	SiGe BiCMOS	90	NR	**88**	**19.5**	NR	**15.9**	15.0	CE	NR	SE	E
					85	**22.0**		**19.1**	17.0	2-Stacked	5	SE	
					83	**23.3**		**17.1**	18.7	3-Stacked	6	SE	

Table 2.15: W-band one-way linear CMOS/SOI PA comparison

Authors	**Year**	**Technology**	**Lith (nm)**	**Freq (GHz)**	P_{sat} **(dBm)**	P_{1dB} **(dBm)**	**Peak PAE (%)**	**S.S. Gain (dB)**	**Topology**	**Stages**	**PA Mode**	**Input SE/Diff**	**Output SE/Diff**	**Baluns (if diff)**	**Class**
Xu [119]	2011	CMOS	65	**100**	**13.9**	10.4	**8.5**	13.0	CS	3	Diff	Diff	Diff	Integrated	A
					13.8	11.2	**9.4**	13.4							
					12.0	9.0	**8.9**	10.1							
					11.8	10.1	**10.0**	10.3							
Agah [34]	2014	CMOS SOI	45	**91**	**19.2**	NR	**14.0**	13.0	3-Stacked	3	SE	SE	SE	-	A
Tsai [120]	2012	CMOS	65	**90**	**14.8**	12.5	**8.7**	12.0	CS	3	SE	SE	SE	-	A/AB
Jayamon [121]	2013	CMOS SOI	45	**89**	**17.0**	11.5	**9.0**	8.0	3-Stacked	1	SE	SE	SE	-	A/AB
Agah [122]	2013	CMOS SOI	45	**89**	**15.8**	NR	**11.0**	10.2	2-Stacked	2	SE	SE	SE	-	NR
Chan [123]	2011	CMOS	90	**80**	**13.3**	10.8	**11.8**	18.0	Cascode	2	SE	SE	SE	-	A
Kim [33]	2012	CMOS SOI	45	**80**	**12.4**	12.0	**14.2**	11.0	Cascode	2	SE	SE	SE	-	AB
Jayamon [124]	2016	CMOS SOI	32	**78**	**19.6**	NR	**18.0**	12.0	3-Stacked	1	SE	SE	SE	-	A
					18.7		**24.0**	11.0				SE	SE		AB
Rohani [125]	2017	CMOS	28	**78**	**12.0**	7.5	**13.0**	17.0	CS	3	SE	SE	SE	-	A/AB

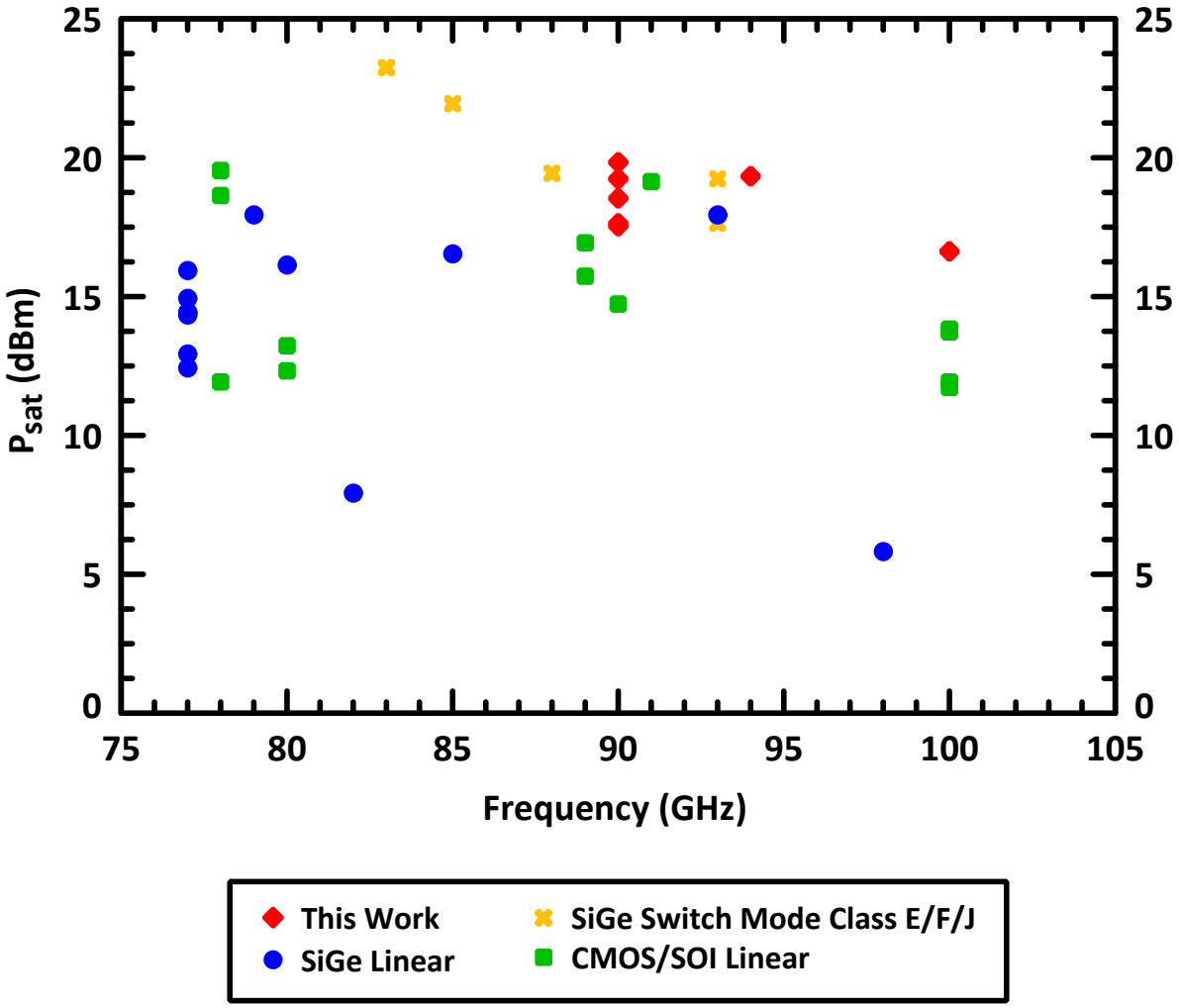


Figure 2.102: Survey of silicon power amplifiers and their saturated output powers P_{sat} for W-band one-way (non-power-combined) PAs. The W-band PAs in this work (and breakout variants) are in red.

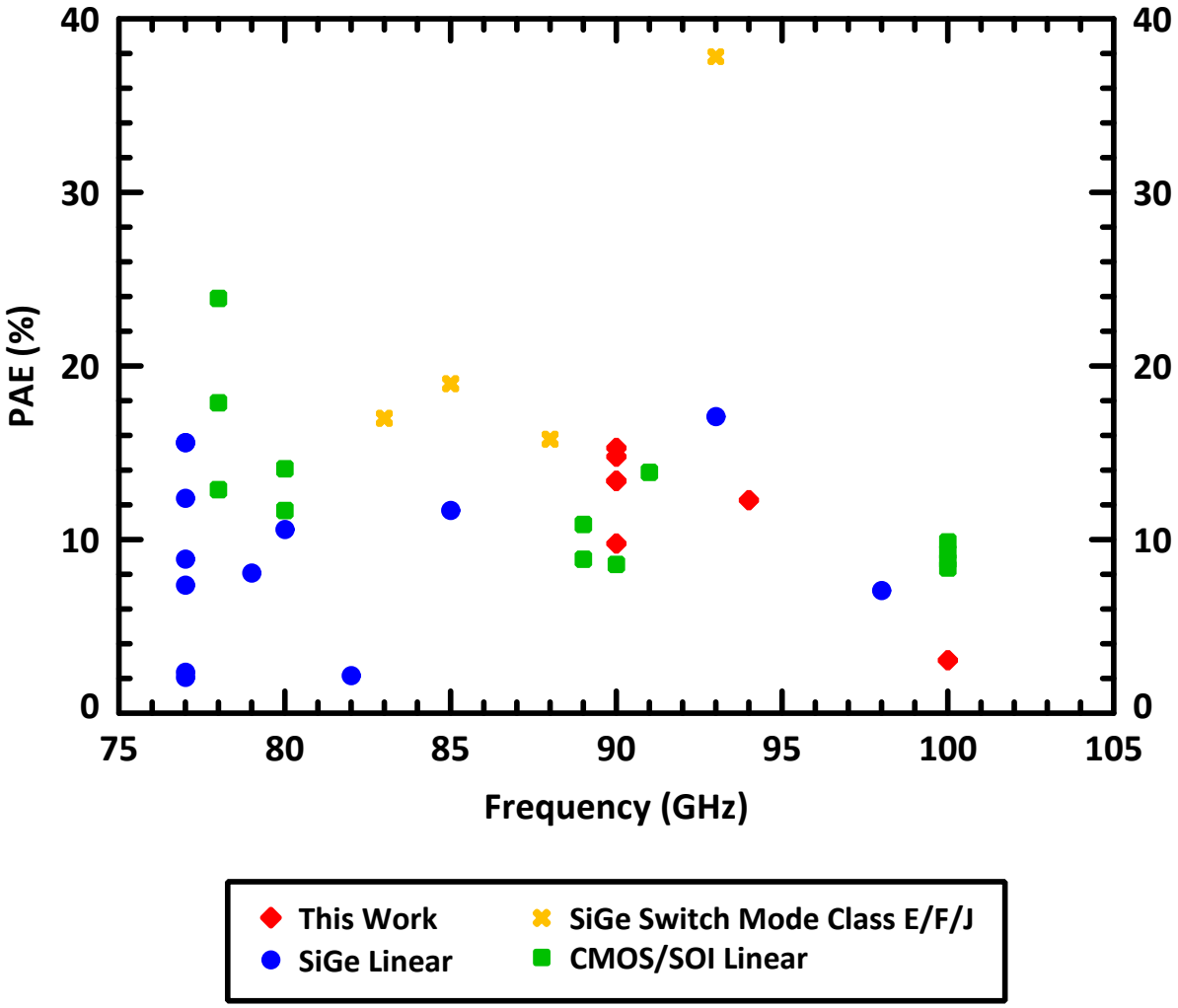


Figure 2.103: Survey of silicon power amplifiers and their power added efficiency (PAE) for W-band one-way (non-power-combined) PAs. The W-band PAs in this work (and breakout variants) are in red.

The values for the W-band CMOS/SOI linear, one-way power amplifiers in the W-band dot plots are quoted from [33, 34, 119–125].

Ka-Band to W-Band PA Comparisons

Fig. 2.104is a dot plot of reported saturated output powers P_{sat} for SiGe linear one-way PAs, SiGe power-combined PAs, SiGe switch-mode one-way (non-power combined aside from differential operation) PAs, CMOS/SOI linear one-way PAs, and the PAs presented in this dissertation: W-band 90 GHz PA & variations, Ka-band 30 GHz PA & variations, and the quasi-optically-combined 3 × 3 PA array. Frequencies outside of the primary bands of interest— Ka-band (26.5–40 GHz) and W-band (75–110 GHz)— may not include all reported PAs.

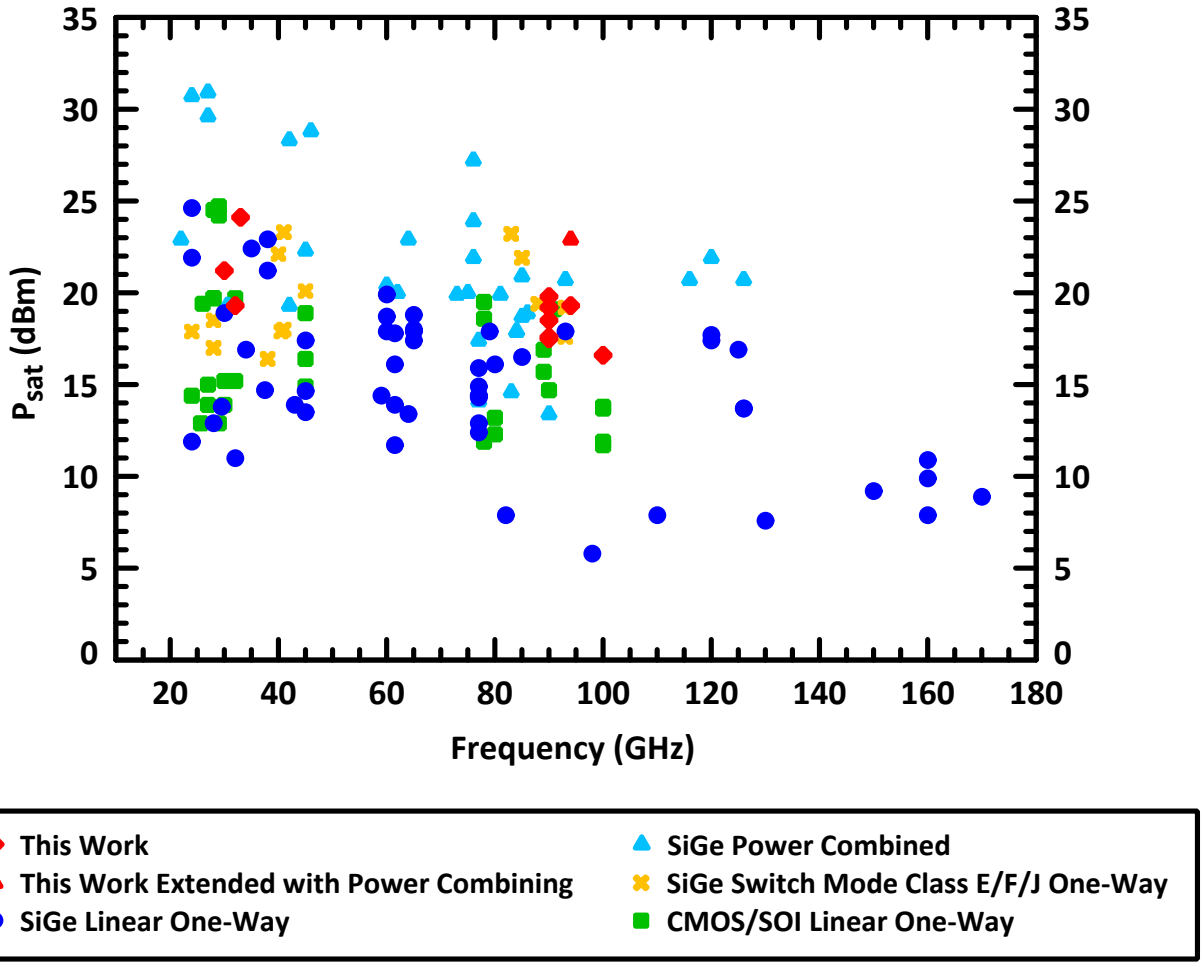


Figure 2.104: Dot plot of report saturated output powers P_{sat} for Ka-band to W-band PAs.

Fig. 2.105 is a dot plot of reported PAEs for the same PAs in Fig. 2.104. Frequencies

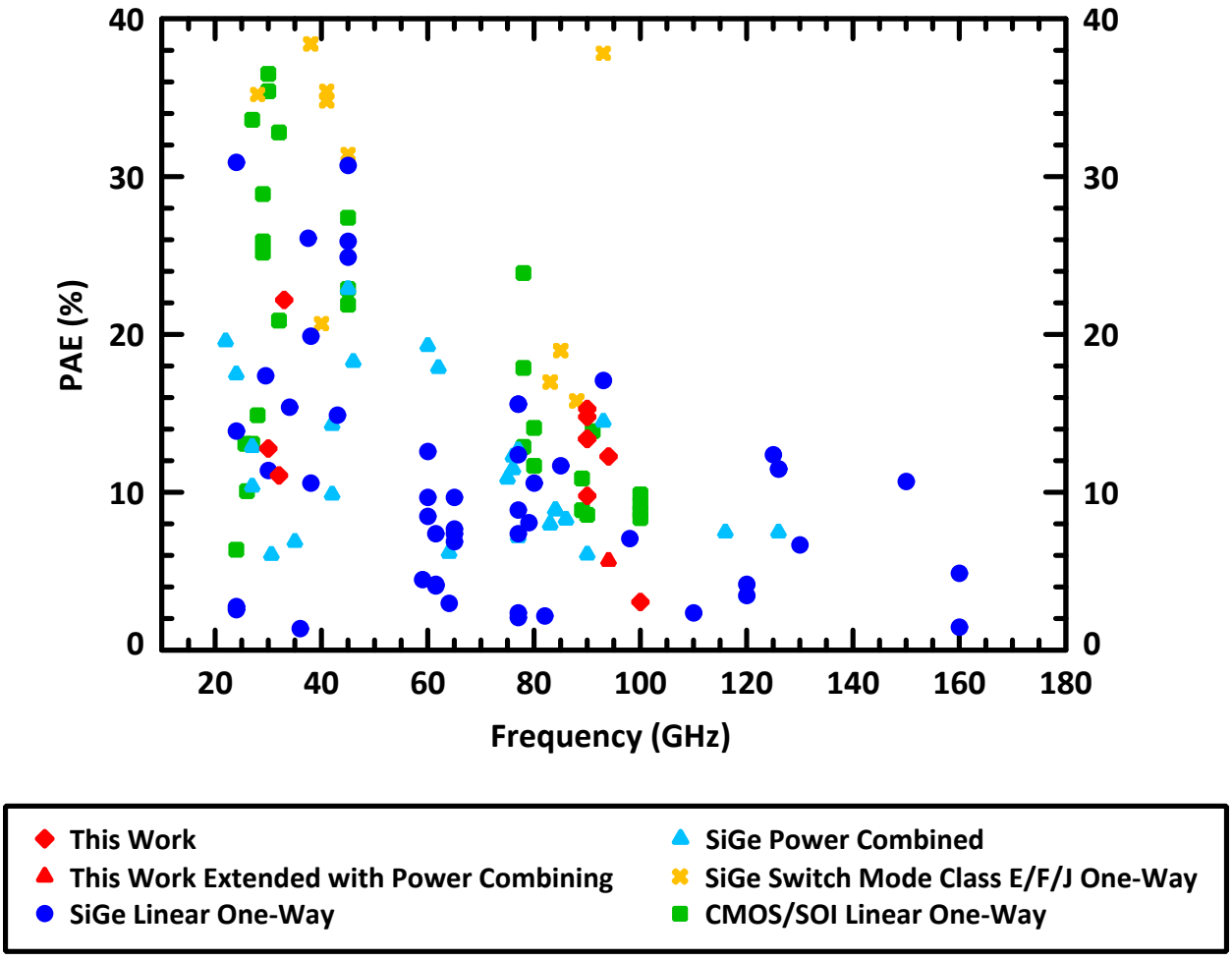


Figure 2.105: Dot plot of report power added efficiency (PAE) for Ka-band to W-band PAs.

outside of the primary bands of interest— Ka-band (26.5–40 GHz) and W-band (75–110 GHz)— may not include all reported PAs.

The values for SiGe linear, one-way power amplifiers in the dot plots are quoted from [13–18, 23–25, 25, 27, 41–43, 43, 51, 52, 58, 71–73, 115, 116, 126–129, 133–148].

The values for SiGe switch-mode power amplifiers in the dot plots are quoted from [44–46, 50, 59, 117, 118, 149–152].

The values for SiGe power-combined power amplifiers in the dot plots are quoted from [9, 19–22, 26, 41, 47–52, 58, 129–132, 139, 147, 153–157]

The values for CMOS/SOI linear, one-way power amplifiers in the dot plots are quoted from [28, 33–35, 119–125, 158–169].

2.12 Conclusion

This work has presented a W-band 79–97 GHz fully-integrated SiGe linear power amplifier which achieves high gain and high output power over the 20% 3dB-gain bandwidth. Several key design strategies allowed enabled this performance with first-pass silicon success: judicious utilization of unit-cell lumped-element MIM capacitors and wide unit-cell RF transmission lines in the input, output, and inter-stage matching networks; unit-cell self-resonant MIM capacitors; low 20 Ω external base impedances at DC and RF; geometric increases in transistor areas; compact & dense layout of transistor cells based on high-f_T npn cells; lower-Q bias chokes implemented below the RF ground plane; and large bias decoupling capacitor arrays around the chip (with and without resistors) for broadband frequency stability. The design can be easily scaled to other W-band frequencies by simply changing the shunt-stub transmission-line lengths in the matching-networks.

By choosing design values and implementations of components that can be easily re-used throughout the design, the PA maintains high performance and wideband characteristics in spite of expected process shifts and modeling errors. The effects of the npn HBT modeling errors are reduced by using small, well-modeled unit-cell transistors to build larger transistors. The mutual thermal coupling of the aggregated unit-cells are exacted from DC measurements and modeled for each power cell, allowing more accurate DC & RF design optimization. The design methodology results in wideband performance in both the small-signal and large-signal characteristics. The resulting PA has a fractional bandwidth of 20% in both small-signal parameters (small-signal gain, input/output matching) as well as large signal parameters (gain, output saturated power, output 1 dB-compression point). The presented design approach achieves 19.9 dBm or 98mW of output power at 90 GHz. The design is successful in its first silicon iteration without prior measured large signal load-pull data on the custom-designed

power transistors.

The implemented W-band design closely approaches the power limits of the common-emitter configuration at 90–100 GHz. As demonstrated, the presented design approach is useful for generating significant linear mmW output powers up to a frequency of approximately $0.45 \times f_T$, after which HBT gain limitations prevail.

The linear W-band power amplifier presented here is usable stand-alone or integrated into an array or other power-combining techniques. The PA's gain, efficiency, and output power support further PA integration into larger transmitter systems and arrays, as later demonstrated in Chapter 4.

Chapter 3

Ka-Band SiGe HBT Power Amplifiers

3.1 Introduction

An fully integrated silicon germanium power amplifier is presented for wideband applications covering 26 to 40 GHz [7]. No distributed amplifier or negative feedback techniques are employed to achieve wideband performance. The 42% fractional-bandwidth PA has a gain of 13 dB and a saturated output power of 19.4 dBm with 11.2% PAE from 32 to 33 GHz. The output P_{1dB} and P_{sat} are greater than 15.5 dBm and 17 dBm, respectively, from 26 to 40 GHz. The 1.83 mm^2 chip consumes 525 mW (375 mA) from a 1.4 V supply.

3.1.1 Balanced Amplifiers

The PA design presented here uses a balanced amplifier topology to boost the output power by +3 dB relative to a single amplifier. The balanced topology utilizes two identical and symmetric PAs by driving them in a balanced fashion (180 degrees out of phase). The resulting output is also balanced, as illustrated in Fig. 3.1. Taking a single PA and using it as a building block for a balanced PA results in aggregate PA with the same gain and PAE but +3 dB higher output power, as illustrated in Fig. 3.2.

3.2 Circuit Implementation

Power amplifier designs are notorious for requiring design iterations to achieve target circuit performance. Many challenges arise because large-signal designs push device capabilities to their limit and often push models past their useful bounds. Distributed thermo-electrical interactions further complicate the predictability of power amplifier designs. Since no foundry data or models based on pulsed, large-signal load-pull measurements are available, the design approach utilizes simple and robust design topologies and layout techniques in order to maximize the success of a first-turn design.

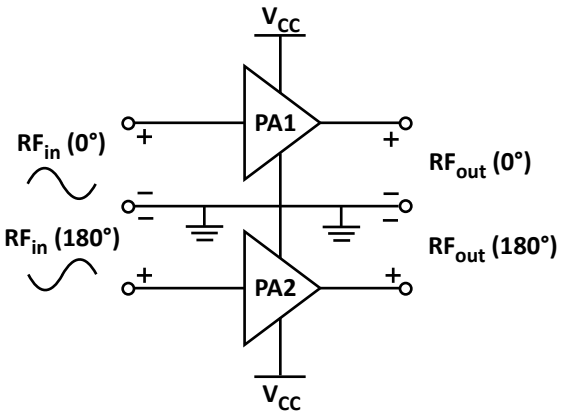


Figure 3.1: Balanced PA consisting of two identical PA units (PA1, PA2). The two PAs are driven out of phase, creating a balanced system.

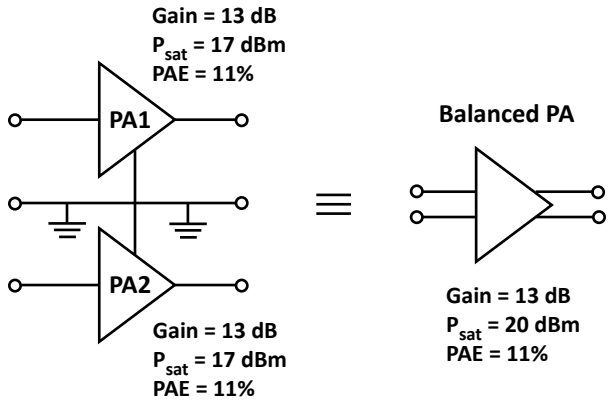


Figure 3.2: PA1 and PA2, when driven in a balanced fashion, result in an aggregate PA with the same gain and PAE but 3dB higher output power.

The balanced PA is designed in a commercially available 0.12 μm SiGe BiCMOS technology with a bipolar transistor f_T of 200 GHz. Fig. 3.3 and Fig. 3.4 present the PA design schematic and layout, respectively. To attain wideband performance, the design takes advantage of the high f_T HBT, broadband matching techniques, and a judicious use of distributed and lumped elements. The design achieves wideband performance without employing distributed amplifier or negative feedback techniques.

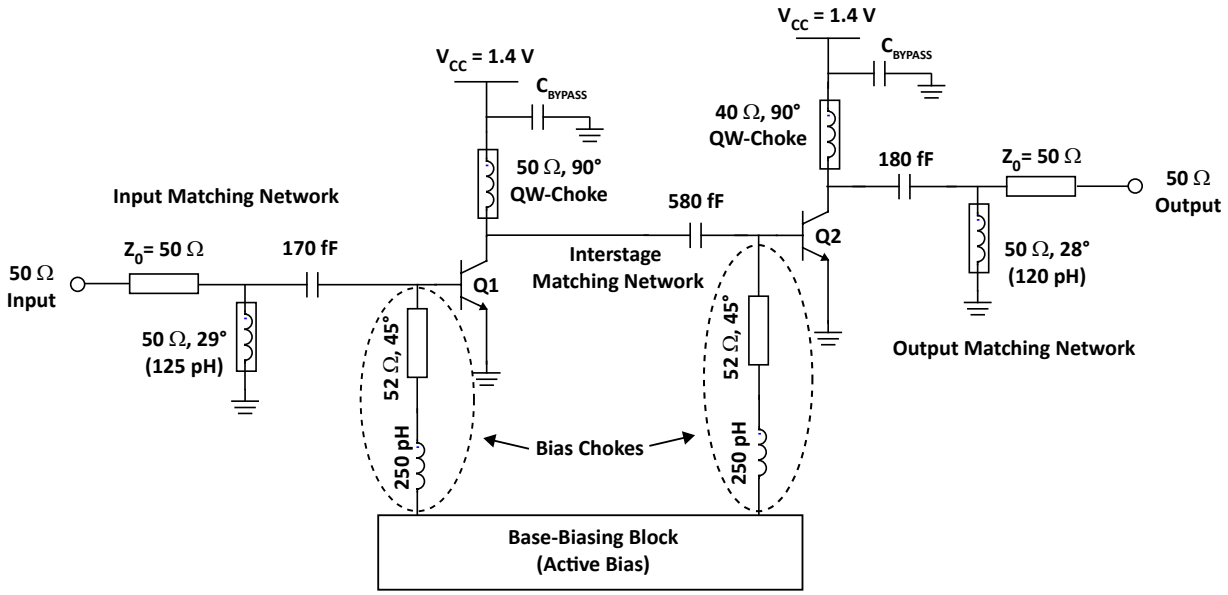


Figure 3.3: Simplified half-schematic of the balanced Ka-band PA. Transmission line values are given for the midband frequency.

The amplifier operates in the very linear Class A region. Class A operation allows the PA to be used in systems that use AM modulation in the envelope of the carrier. Class A amplifiers offer very high linearity, but their performance comes at the cost of high quiescent, DC current dissipation (and thus, has a lower efficiency).

3.2.1 Design Methodology

The implemented design preserves the high f_T performance of the HBT devices while giving careful design consideration to the layout, RF coupling, low-frequency stability, and reliability of the power amplifier.

The wideband design does not employ distributed amplifier (traveling-wave amplifier), feedback, or power combining techniques. Rather, the wideband performance is achieved by compensating the matching networks to offset the native gain (f_T) rolloff of the power transistors. Unfortunately, a broadband "lumped amplifier" design has no simple analytical

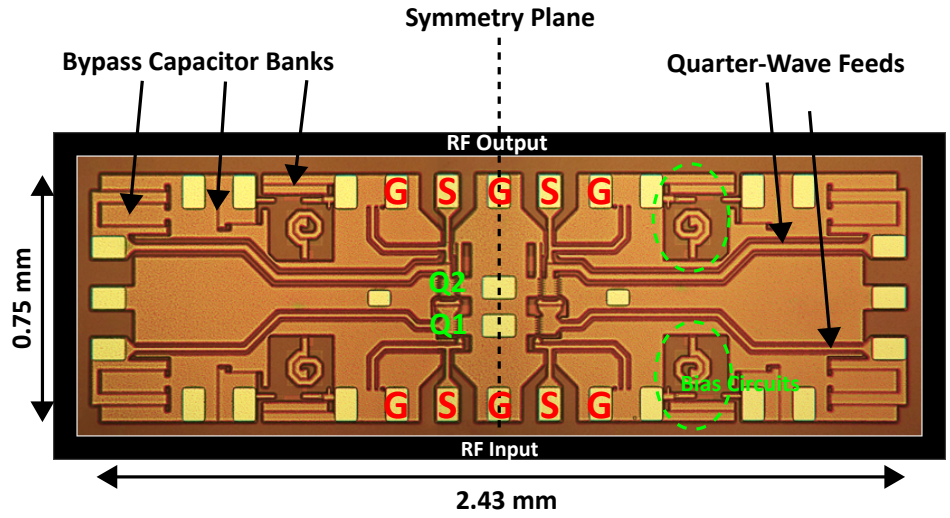


Figure 3.4: Chip micrograph. The total area, including pads and integrated bypass capacitors, is 1.83 mm^2.

or direct synthesis approach for circuit design and relies on computer-aided techniques to iteratively refine a design [104]. Balancing performance metrics such as gain, output power, and PAE across a wide operating frequency are important in a successful design.

The approach taken in this design is to begin optimizations at the midband frequency (32–35 GHz), determine typical source/load-pull contours for each PA stage (Q1 and Q2) at typical power levels, and then expand the design to cover a wide bandwidth. Keeping matching networks, biasing, and feeds as simple as possible simplifies the design process, minimizes losses, and keeps the design more predictable while maintaining wideband performance.

A "reliability-aware" design is constrained by electromigration limits, and reliability is always taken into account throughout the design process of this PA. Where possible, no minimum-sized lines or devices are used, vias are placed in ample redundancy, and foundry-specified reliability guidelines are followed to obtain a target reliability of 100,000 power-on-hours at 100°C.

3.2.2 Simulation Tools

Two primary simulation tools are used in the design of the balanced power amplifier: Agilent ADS Suite and Cadence Design Framework II. The Agilent ADS tools include the native, harmonic-balance circuit simulators, RFDE (RF Design Environment), Dynamic Link, and Momemtum. Integrated 2.5D electromagnetic simulations and harmonic balance prove extremely useful in the analysis, design, and optimization of large-signal circuits. Cadence

is used for final layout, design-rule check, parasitic extraction, and verification (LVS). Both Agilent ADS and Cadence are supported by the foundry models.

3.2.3 General Layout Considerations

Fig. 3.4 presents the PA-chip photomicrograph. The chip, including pads, consumes an area of 1.83 mm^2. On-chip DC bypass capacitors are included for integrated supply and bias filtering.

Special layout constraints exist for the PA to be driven differentially on-chip. While it is true that any two identical PA's driven differentially can deliver 3 dB more power than a single PA, creating such a balanced PA from a standalone, non-balanced, single-ended design is not trivial.

While one may argue that any single-ended power amplifier can be converted to a balanced amplifier by adding a second amplifier, integrated circuit solutions place additional restrictions on the floorplan of the amplifier. Creating a useable and probe-able layout restricts the location of the RF I/O pads and DC feeds further– the I/O's of each PA half are restricted to three sides. In the presented design, the floorplan of each PA unit– each half of the balanced amplifier– is laid out such that the unit PA can be mirrored across a plane of symmetry while maintaining the I/O's at the periphery of the chip.

Fig. 3.5 demonstrates the floorplan constraints in any fully-integrated balanced PA design. Fig. 3.5a. illustrates the integration issue if a standalone, single-ended PA (designed without balanced drive in mind) is used in an integrated fashion. The resulting I/Os in the center of the combined chip present a practical limitation on measuring and using the balanced amplifier. Fig. 3.5b. shows the necessary layout constraints in order to mirror each PA half across a plane of symmetry. Note that the I/O's only exist on periphery of the combined chip.

The total die area is dominated by passives: transmission lines, matching networks, power/ground distribution, and interconnects. The RF signal path is concentrated in the center of the chip while biasing is pushed to the chip peripheries. The power transistors for each symmetric half of the balanced PA are located in the center of the chip, assuring that each half of the PA have transistors that remain well-matched to the other half. Large, redundant ground busses and ground planes wick away thermal energy from the power transistors, mitigating local temperature effects. The RF pads are designed with 150 μm pitch to accommodate the available measurement probes.

Additionally, the use of deep-trench isolation grids, high-resistivity regions, and large, redundant substrate contacts isolate the transistors from each other and isolate RF from

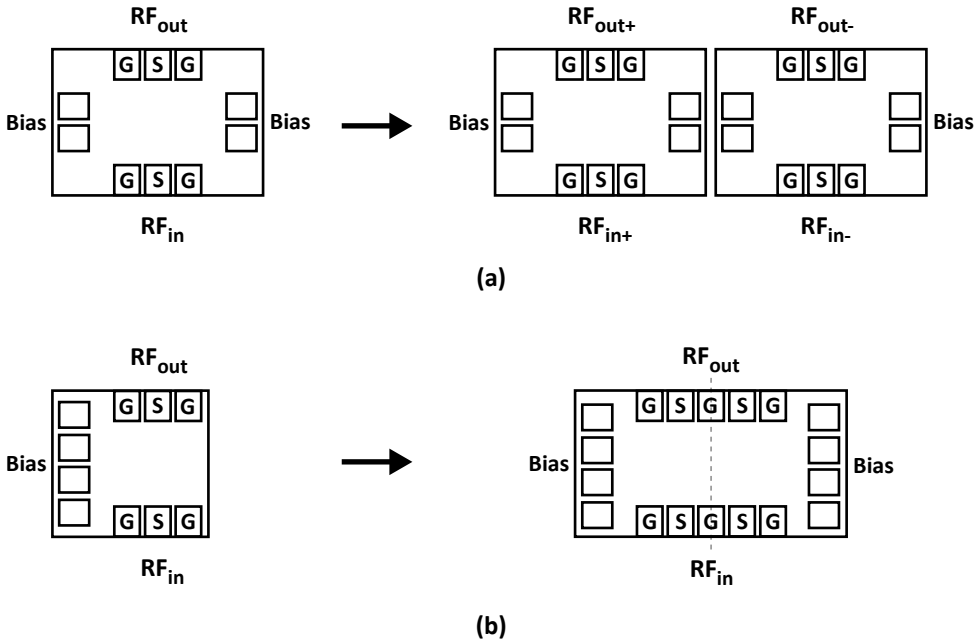


Figure 3.5: The issue of creating an integrated balanced PA from a unit PA that has not been designed with balanced drive in mind is illustrated in (a). A floorplan-aware solution is provided in (b).

DC regions. The isolation of the power transistor area avoids any unintended affects from substrate injection currents arising from the highly energetic carriers in the HBT devices.

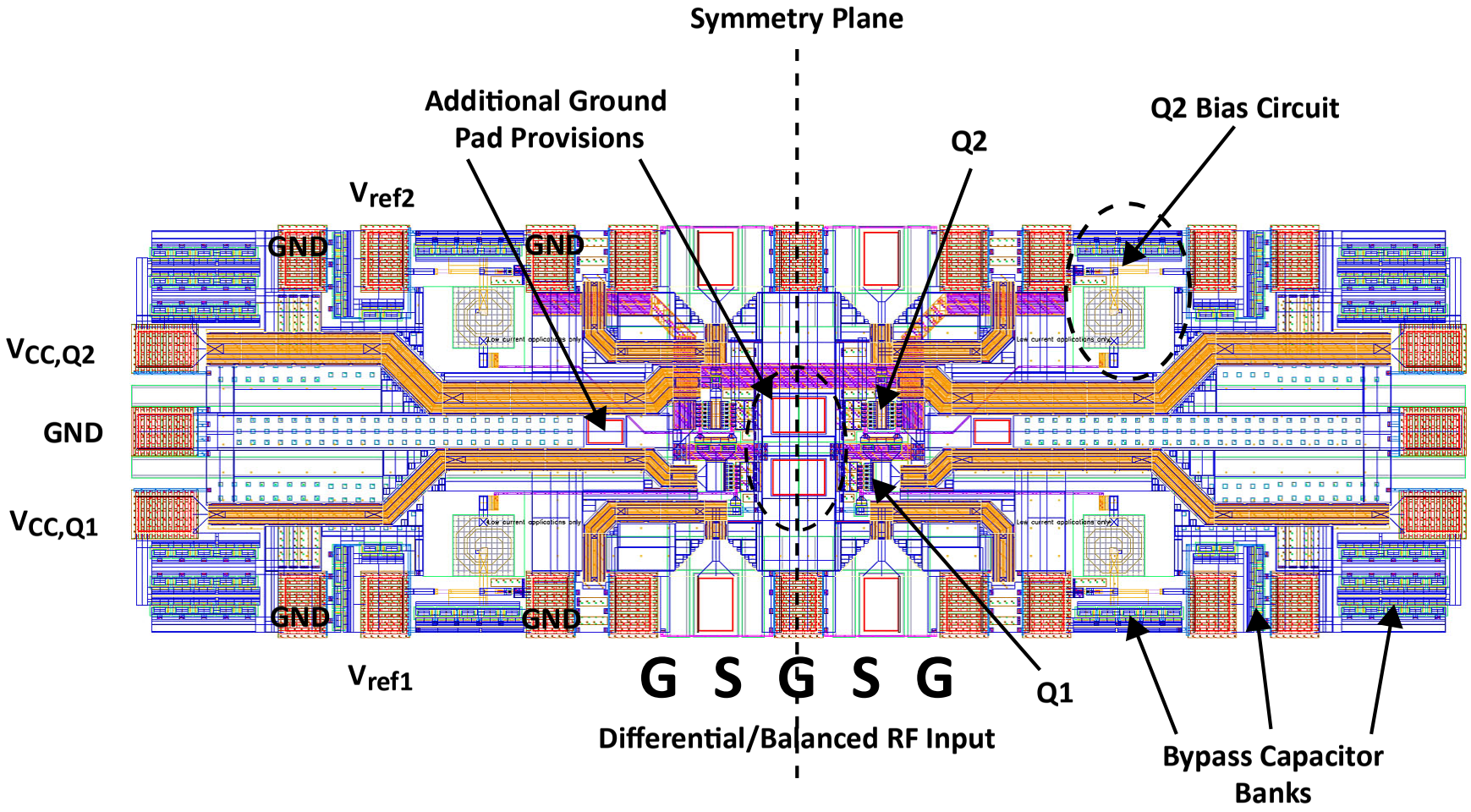


Figure 3.6: Markup of the Ka-band PA differential layout.

3.2.4 Topology Overview

The PA consists of two common-emitter gain stages biased in linear, Class A operation with integrated input, output and interstage matching networks. Wideband performance is achieved in a compact manner by utilizing high-pass matching networks in the input, output and interstage network as well as low-pass filters in the collector feeds.

RF inductances are implemented via high-Q, shorted transmission-line stubs, and MIM capacitors are used because of their compactness (high capacitance density).

Both Q1 and Q2 stages are matched to provide wideband performance and suitable output power. From an integration standpoint, if required, additional system gain can be implemented in a series of PA driver stages where the lower output power requirements ease wideband, high-gain designs, and push the design space into the small-signal region.

Fig. 3.7 presents the high-level block diagram of the two-stage Ka-band half-PA. The other half of the differential PA is identical and mirrored about the symmetry plane. Independent collector voltage biases and base biases, which control the HBT currents, allow independent collector and current tuning of each gain stage.

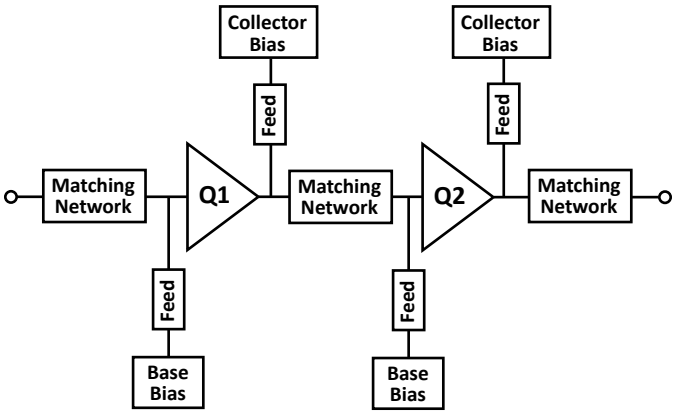


Figure 3.7: Half-PA block schematic.

3.2.5 Active Devices

A two-stage, common-emitter power amplifier is implemented using standard parameterized cells in the technology design kit. Although customized power-transistor layouts are typically preferred for power amplifier designs, no such power cell layouts have been developed by the foundry, and stringent design rules and lack of carrier-level device modeling data prevent user-developed power cells. The foundry npn HBT layouts are optimized for high-f_T,

small-signal operation. This design utilizes the optimized f_T device to achieve wideband performance using straightforward matching/compensation techniques.

Design kit models and foundry-measured data show that a 5 µm HBT, dual-base device achieves the highest possible f_T in the process (200 GHz). This 5 µm HBT is used as the unit cell to build aggregate power devices. The aggregate power devices simply place these 5 µm cells in parallel while sharing collector regions to keep the layout compact. A high-reliability layout option is chosen to widen the collector contacts, providing redundant vias and wider metals for improved power handling.

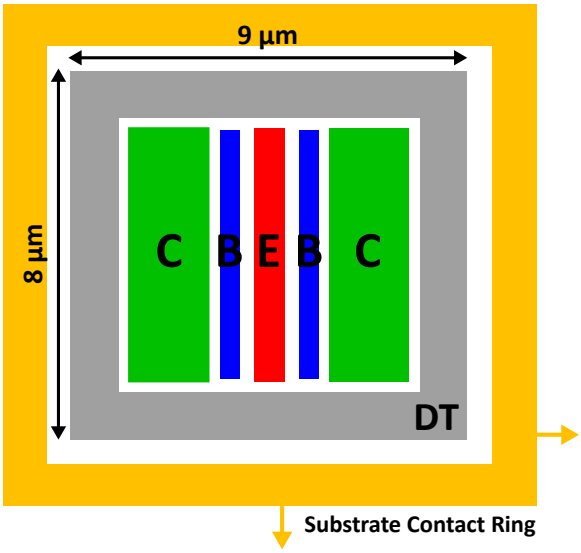


Figure 3.8: A standard 5µm dual-base, dual-collector npn HBT cell with high-reliability switch enabled.

A 5 µm C–B–E–B–C npn layout (dual-base, dual-collector contacts), shown in Fig. 3.8, is chosen over a simpler, more compact C–B–E configuration because the symmetric transistor achieves a higher f_T/f_{MAX}. The increase in performance results from the reduction of collector access resistance and a more symmetric spread of injected electrons into the collector, thus delaying the onset of the Kirk effect and f_T collapse at high current densities [75, 170].

The first common-emitter stage uses a 30 µm aggregate device (Q1) while the second common-emitter stage uses a 60 µm device (Q2) shown in Fig. 3.9.

The transistors are biased at peak f_T current density. Peak f_T occurs at a quiescent bias point of roughly 1.5 mA per µm of emitter finger length. This corresponds to a quiescent current of 45 mA and 90 mA in the Q1 (first stage) and Q2 (second stage) power transistors, respectively.

Geometric scaling between the two stages (Q2 begin twice the size of Q1) ensures that the output transistor limits the compression characteristics– the first stage, Q1, will be

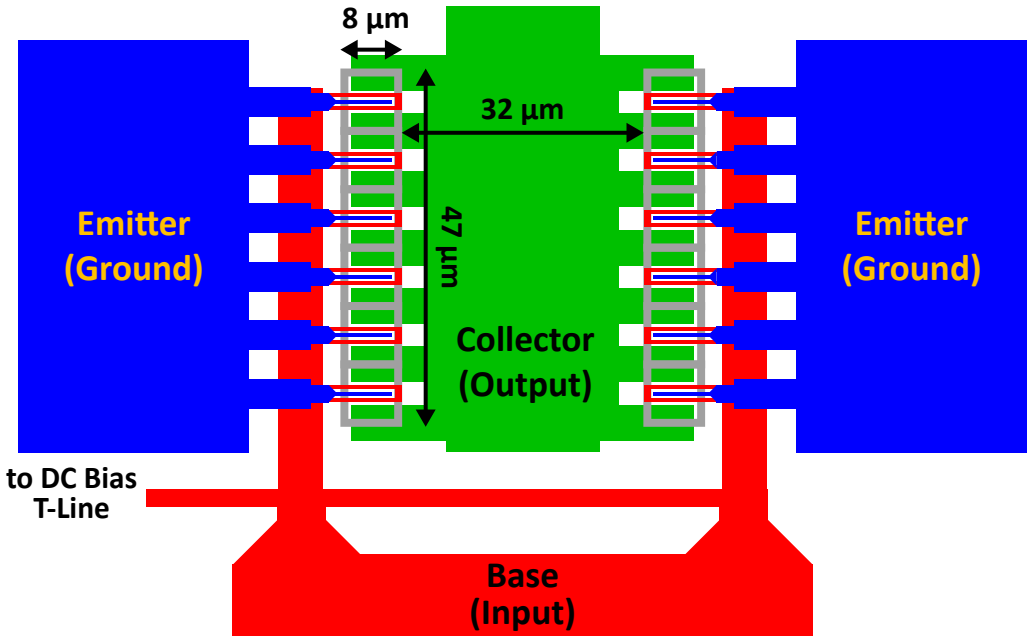


Figure 3.9: Layout of common-emitter output transistor Q2 = 60µm

able to deliver enough input power to the output Q2 transistor such that the compression characteristics are dominated by the output stage.

Simulations based on a cascode topology show that the cascode topology does not offer P_{sat} improvements over common-emitter topologies, and this finding is supported by recent SiGe work– the benefits mostly cancel out, and the matching the cascode leads to a more narrowband match [16, 171].

3.2.6 Biasing

Active biasing is used, rather than resistive biasing, because it gives fine control over the power amplifier quiescent currents and maintains more consistent DC current under RF drive and across temperature [172].

High-f_T SiGe HBT devices suffer from low BV_{CEO}, limiting the maximum collector voltage swing and deliverable output power. A design may overcome BV_{CEO} limitations by presenting a low external base impedance to the power transistors, thereby increasing the effective collector breakdown voltage well beyond BV_{CEO} [74]. Using this bias-design technique, SiGe PA designs avoid an early onset of collector impact ionization, and resulting PAs are able to push higher voltage swings and output powers. This wideband PA design overcomes the technology's 1.7 V BV_{CEO} by presenting low impedances to the base terminal of the power transistor at DC and the operating frequencies. The input and interstage matching networks present 10 to 20 Ω impedances from 26 to 40 GHz while the base biasing, consisting of a high-impedance transmission line and 250 pH spiral inductor, presents a 40 Ω DC impedance. The resulting effective collector-emitter breakdown voltage, BV_{CER}, allows the output voltage at the collector to peak above 2.5 V.

Precision back-end-of-line, integrated tantalum nitride (TaN) thin-film resistors are used in the bias generation. These TaN resistors provide excellent absolute accuracy (3σ tolerance better than 10%) and relative matching across the chip (better than 0.1%), ensuring consistent bias currents from each half of the balanced PA and from die to die. The TaN resistors also have a negative temperature coefficient, offsetting some of the PA temperature dependence by compensating with higher quiescent current draw at higher temperatures.

3.2.7 Output Transistor Characteristics

Fig. 3.10 presents the DC I-V simulation of the output 60 μm device using the implemented bias scheme. Note that the collector breakdown voltage is greater than $BV_{\text{CEO}} = 1.7$ V at low current levels. The maximum current swing is limited by electromigration constraints. The load line plotted in this particular example is roughly 15 Ω.

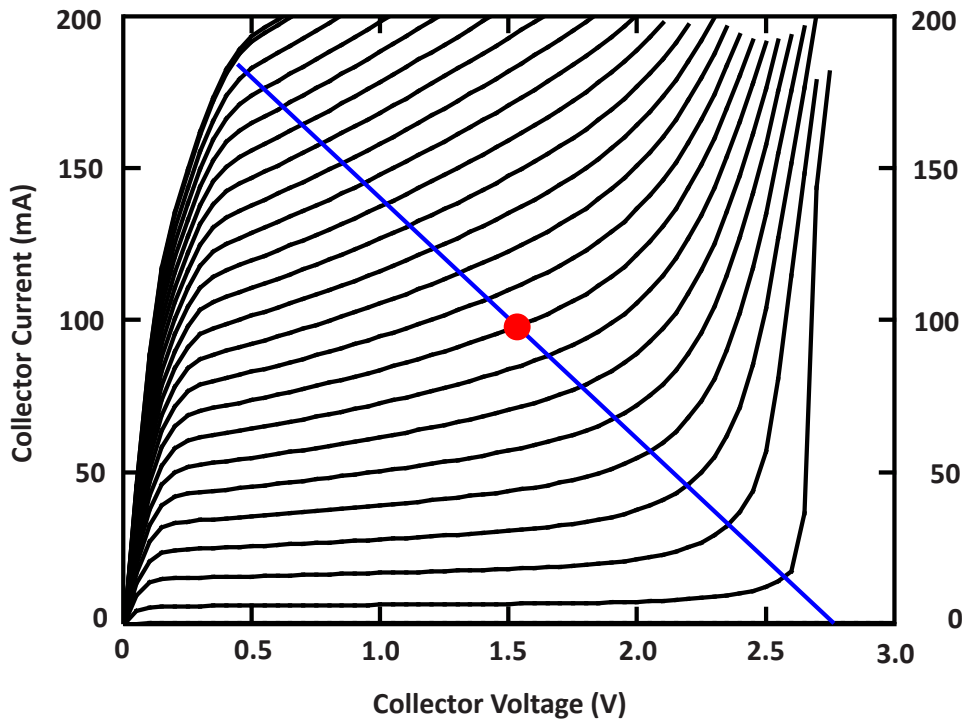


Figure 3.10: Simulated DC I-V plot of the output 60 µm device with the implemented bias scheme. A 15 Ω load line is illustrated.

The 60 µm output transistor's maximum available gain is plotted in Fig. 3.11. The transistor is biased at peak f_T and is unconditionally stable above 45 GHz. This high available gain is traded for bandwidth and power in subsequent load-pull optimizations.

The near-midband, simulated load-pull data in Fig. 3.12 is based on the design kit models. The optimal source and load impedances under large-signal conditions are around 10 Ω. The realized load has an impedance with real part approximately 13 Ω.

Extracted large-signal load-pull models are not available for this process, yet, and known limitations exist for the implemented VBIC models. Simulated versus measured performance in the same technology based on prior comparison studies [14, 173] indicate that the VBIC model is overly optimistic predicting saturated output powers and compression points of large-signal swings by 1–2 dB. The empirical measurement data is taken into consideration and provides a rough correction factor when viewing the optimistic load-pull simulations.

Optimal matching for wideband operation is achieved by performing source/load-pulls on Q1 and Q2 iteratively and by managing device input/output mismatch such that high output power is available throughout a wideband bandwidth.

Fig. 3.13 and Fig. 3.14 show simulated time-domain waveforms for the voltage and current swings at the 1dB compression point.

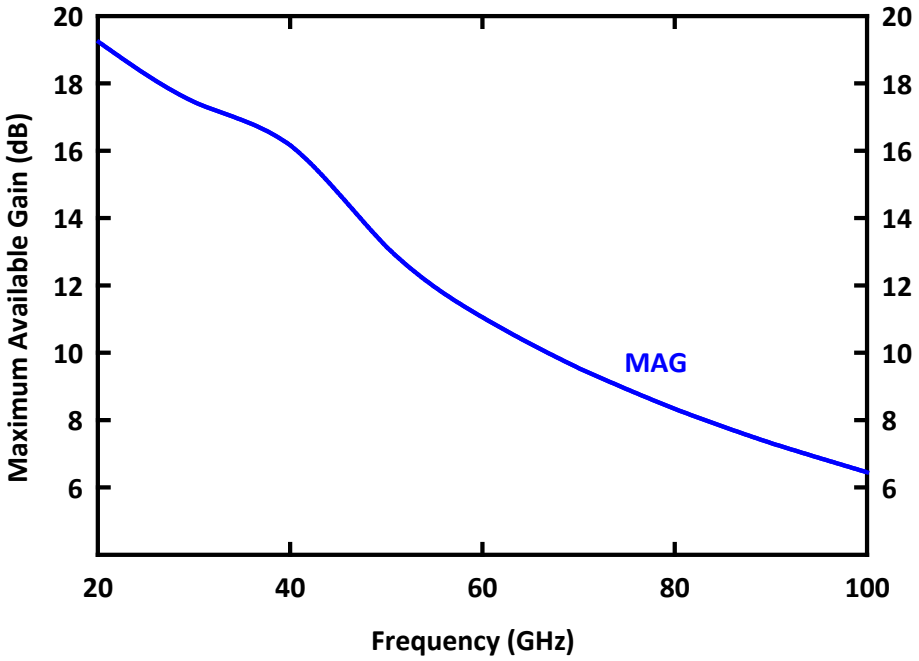


Figure 3.11: Maximum available gain for the 60 µm output transistor biased at peak f_T current. Below 45 GHz, the plot illustrates maximum stable gain (MSG).

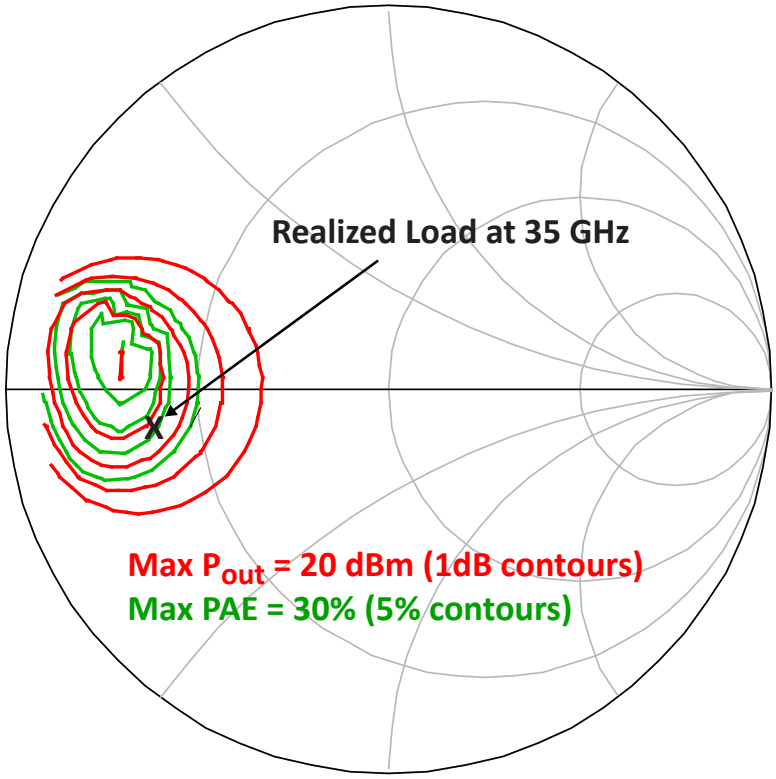


Figure 3.12: Load-pull contours at 35 GHz for the 60 µm output device.

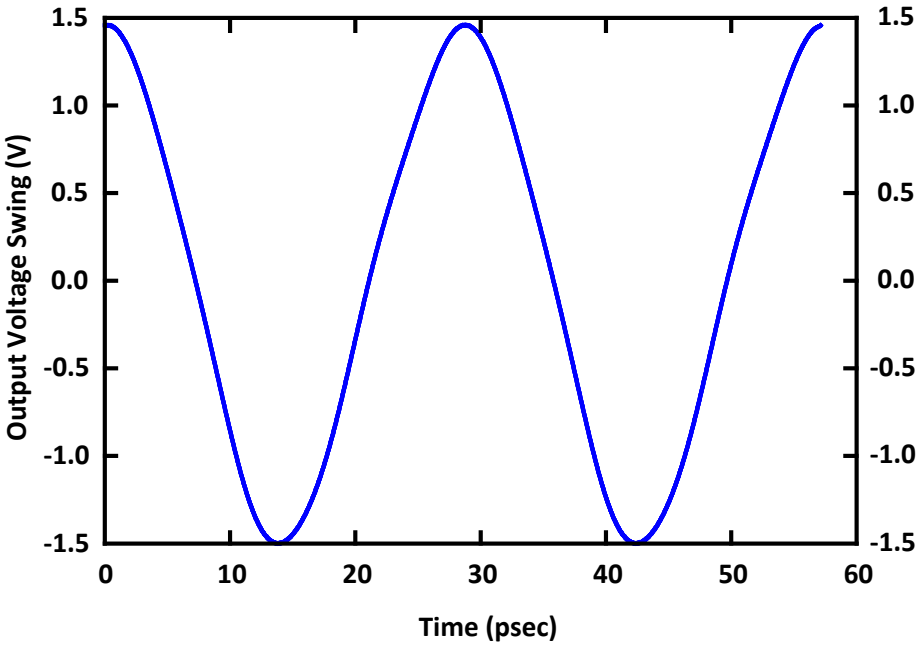


Figure 3.13: Simulated single-ended (half-PA) output voltage swing @ $P_{1\text{dB}}$ compression at 35 GHz.

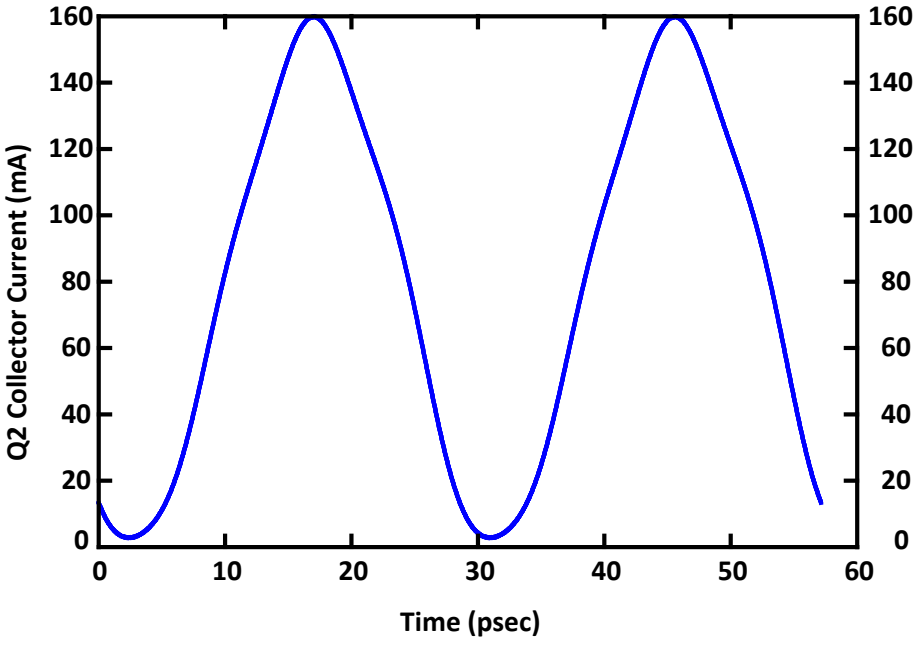


Figure 3.14: Simulated single-ended (half-PA) conduction angle Q2 collector current @ $P_{1\text{dB}}$ compression at 35 GHz.

3.2.8 Matching Networks

Integrated input and output matching networks present 100 Ω differential impedances at the input and output. Shielded 40 to 50 Ω microstrip transmission lines (Fig. 3.15, ϵ_{eff} = 3.9, loss = 0.4 dB/mm) ranging from 12 to 18μm widths are used extensively in the design to reduce coupling to nearby lines, to minimize coupling to the lossy 11 Ω-cm substrate, and to allow routing of DC feeds underneath the RF ground shields. The measured transmission line characteristics are presented in Fig. 3.16.

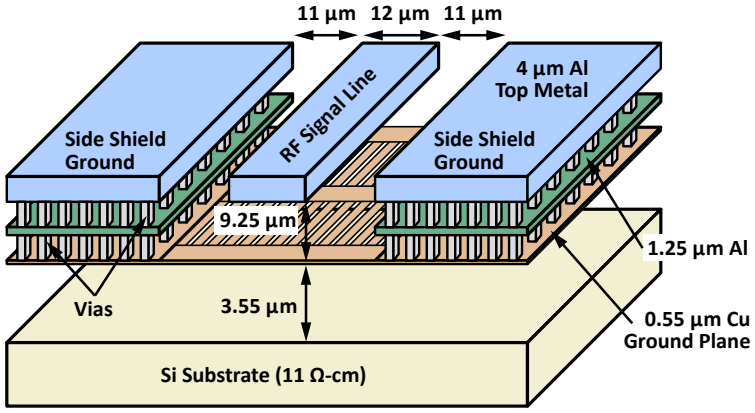


Figure 3.15: Shielded-microstrip transmission line cross-section. Four additional Cu metal layers between the ground plane and substrate serve as DC interconnects (not shown).

The matching networks present source and load impedances to the power transistors that result in positive gain slopes with increasing frequency, compensating for the transistors' f_T roll-off. This compensation results in a wideband response in the gain and output power. The input and output matching networks consist of two L-networks using shunt, shorted-circuited stubs and series MIM capacitors. Both input and output matching stubs present an equivalent inductance of 120 pH (Q = 50) at midband frequencies while both input and output matching MIM capacitors present 180 fF (Q = 20).

A single 550 fF series MIM capacitor serves both as interstage matching as well as DC decoupling of the two PA stages. The single MIM capacitor provides a compact and simple low-loss matching solution that allows the Q1 and Q2 power transistors to be placed closed together, compacting the layout and minimizing RF interconnect loss.

The grounded, short-circuited stubs of the input and output matching networks offer the additional benefit of ESD protection for the RF I/O pads– any ESD present on the RF probes is shunted to the global ground network. Another important benefit of the grounded,

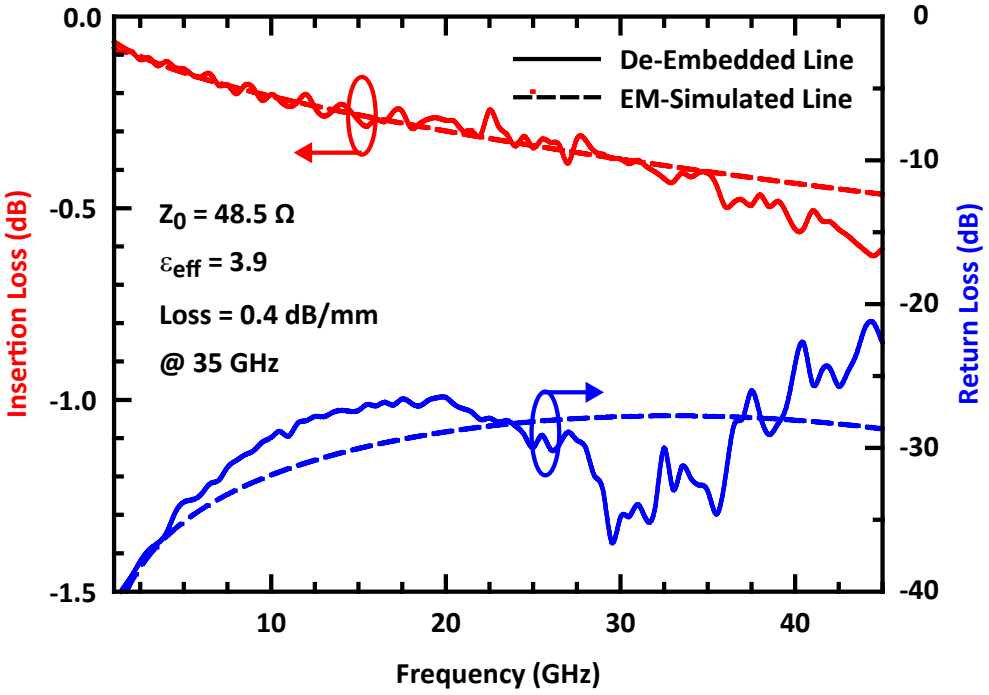


Figure 3.16: Measured (solid) and modeled (dashed) transmission line characteristics de-embedded from a 1mm test structure.

shunt stubs arises because all top-level metals in the RF signal path are tied to ground or to diffusion paths; thus, no dedicated diodes are necessary to comply with top-level metal antenna rules. The absence of extra diodes ensures that there is no degradation of intended RF operation, especially at high voltage swings. (The reverse-biased diodes could easily forward-bias at large signal swings.)

3.2.9 Collector Feeds

Quarter-wave transmission lines are preferred to lumped spiral inductors for collector feed implementation. Spiral inductors are impractical for the collector feeds from 26-40 GHz because of the low self-resonant frequency resulting from an electromigration-compliant linewidth (width $>$ 18μm). Furthermore, if the spiral inductor has more than one turn, the cross-over vias and metal become the limiting components in electromigration compliance.

The high currents involved in power design necessitate careful consideration of reliability constraints. Metal line widths and via arrays are sized to be electromigration compliant and conform to the process technology's reliability guidelines for 100°C operation for 100,000 power-on hours. Wide 18 μm microstrip-shielded quarter-wavelength bias chokes ($Z_0 = 40\ \Omega$, $Q = 16$) feed large DC bias currents (100 mA) to the power transistors while presenting RF

collector impedances greater than 150 Ω. Relative to the output transistor's low impedance (10 Ω), the distributed choke presents an effective open circuit within the operating bandwidth.

Line coupling between the collector feeds is negligible (less than -40 dB). Careful routing and use of the RF ground shields prevents any coupling between the DC base feeds and the collector feeds.

To stabilize the PA at very low frequencies (< 100 MHz), where transistor gain is extremely high, the power-transistor collector feed, base-biasing network, and supply bypass capacitors all present low RF impedances to all the power-transistor terminals (< 1 Ω to the collector, < 50 Ω to the base). Bias stability concerns arising from the measurement setup at very low frequencies (KHz to MHz) is further enhanced by the measurement setup's off-chip DC feed network (ferrite cores, additional bypass capacitors, and aggressive noise filtering).

3.3 Measurement Results

3.3.1 Measurement Methodology

For simplicity, the tests are performed single-ended with one half of the balanced power amplifier. Since the two, symmetric halves are well isolated and independent, a 3 dB power enhancement arises when the balanced amplifier is driven differentially (assuming no amplitude/phasing errors arising from the measurement setup) [14].

Three setups are used to measure the power amplifiers: a large-signal power measurement setup using power meters (Fig. 3.17), a setup using a network analyzer (PNA) for small-signal S-parameters and large-signal verification, and a highly linear setup for intermodulation (IP3) measurements (Fig. 3.18). In the large-signal power measurement setup, the calibrated and offset power meters measure the actual RMS power at the device-under-test's input and output ports. Input and output attenuators present broadband, constant 50-Ω loads to the chip input and output ports, minimizing measurement errors arising from VSWR. At each measurement point, an automated spectrum analyzer (PSA) is used verify the power measurement and to check the output spectrum for any oscillations or distortion products.

Low-frequency bias chokes (DC filtering, ferrite beads on needles, ferrite cores on DC feed lines, force/sense lines), shown in Fig. 3.19, filter out supply noise and further stabilize the measurement setup from very low frequency bias oscillations and maintain desired DC bias points.

Extended PA measurements (i.e. measurement over hours) introduce sources of error from device stressing [174]. Measurements presented here are taken after a PA break-in period (to stabilize measurements). Each PA is stressed over 3–4 hours at the saturated output power

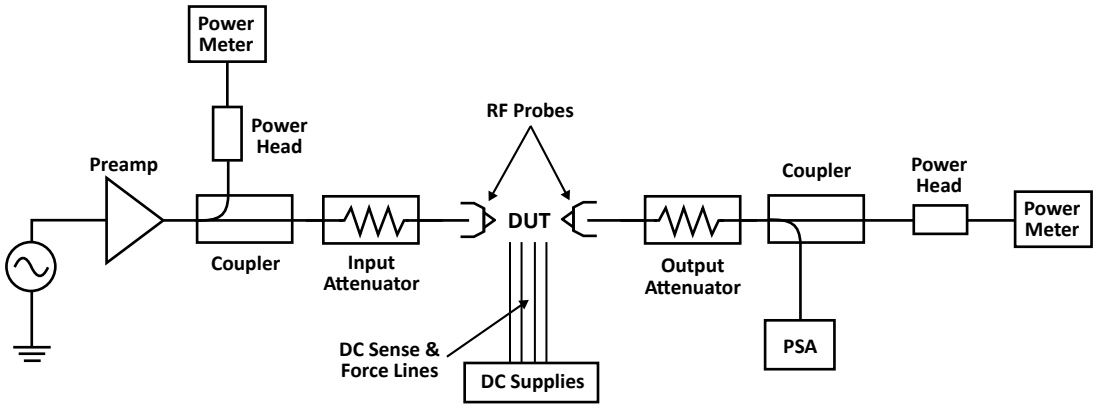


Figure 3.17: Large-signal power measurement setup using couplers and power meters. Input and output attenuators present constant, well-matched 50 Ω loads to the DUT.

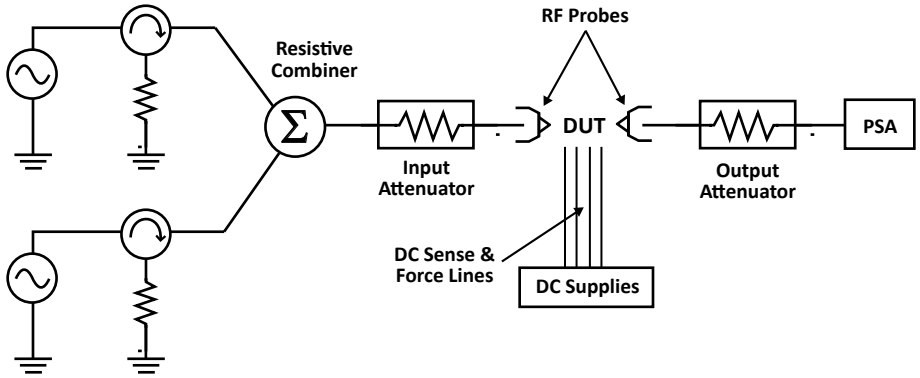


Figure 3.18: Highly-linear measurement setup to characterize intermodulation distortion.

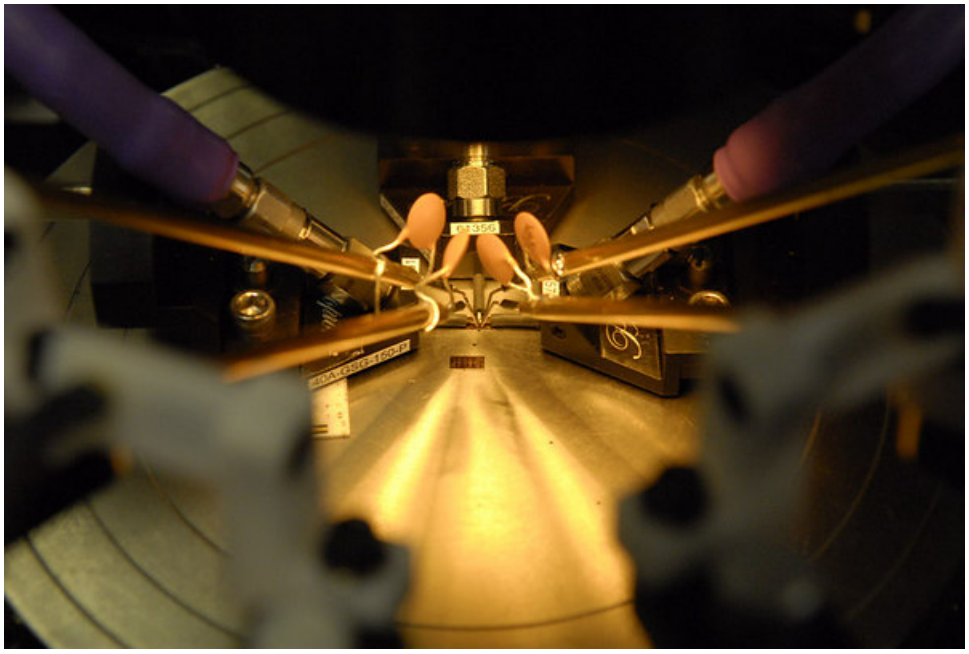


Figure 3.19: DC and RF probe setup. Additional capacitor bypass and inductive chokes on the DC feeds stabilize the measurement setup from very low frequency bias oscillations.

for sufficient break-in time. For the data acquisition, the wafer chuck is kept near ambient room temperature (20°C–25°C) by a temperature regulation system while the PA is given time to settle at each change of bias and input power.

A test setup from 18 to 50 GHz in 2.4 and 2.92 mm coaxial system has been automated via GPIB connections and Agilent VEE. Custom code has been written to carefully step through sinusoidal (CW) measurements over frequency, output powers, bias, and temperature. Preliminary measurements whow that P_{sat} variation from die to die is less than 0.3 dB.

Fig. 3.20 shows the complete lab setup.

3.3.2 Measurement Data

The PA measurements are taken directly on-wafer at room temperature and are referenced to the G–S–G–S–G probe transitions. The combined input and output transition losses, measured to be 0.35 to 0.50 dB from 26 to 40 GHz, have not been de-embedded from the measurements. The differential PA consumes a total quiescent power of 525 mW (375 mA) from a 1.4 V supply. The measured small signal gain (Fig. 3.21) is 13 dB at 32 to 33 GHz, and the 3 dB bandwidth is 25.5 to 41 GHz (a fractional bandwidth of 47%). The output-to-input isolation (not plotted) is better than 40 dB across the entire bandwidth.

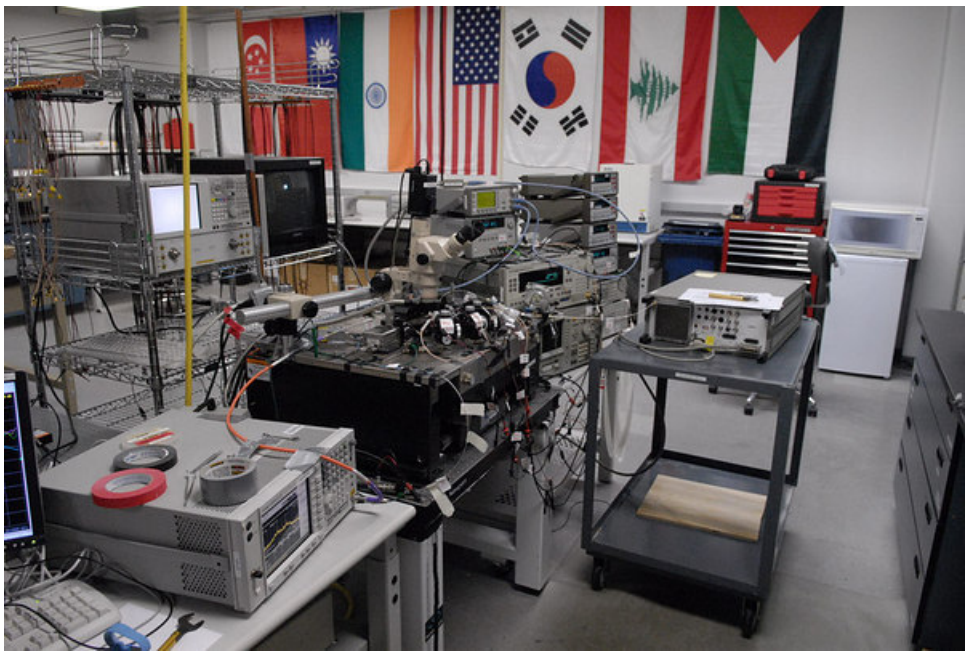


Figure 3.20: Complete lab setup at UC, San Diego.

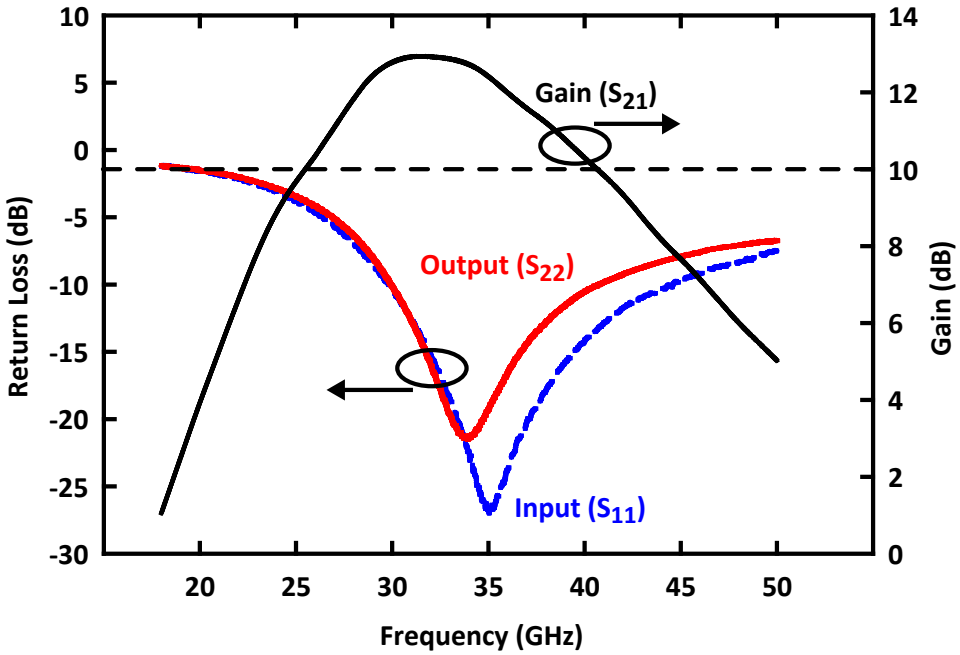


Figure 3.21: Measured small-signal return loss and gain. The output-to-input isolation is better than 40 dB across the entire operational band and is not plotted.

Fig. 3.22 presents the measured saturated and P_{1dB} output powers across frequency. The output saturated power (P_{sat}) is greater than 17 dBm from 25 to 40 GHz, and the output 1dB gain compression point (P_{1dB}) is greater than 15.5 dBm across 25 to 40 GHz. Fig. 3.23 presents the PAE performance across frequency. The peak PAE is 11.2% at 32 GHz, and PAE at P_{1dB} is greater than 5% from 24 to 40 GHz. The peak output power, gain, and PAE occur near the PA midband frequency at 32 to 33 GHz. A plot of the PA midband characteristics is presented in Fig. 3.24. The measured saturated output power is 19.4 dBm, and P_{1dB} is 16.4 dBm.

Summaries of the PA small-signal and large-signal results are presented in Table 3.1 and Table 3.2. Table 3.3 compares the PA design to other recent silicon-based, fully-integrated PAs in the Ka-band.

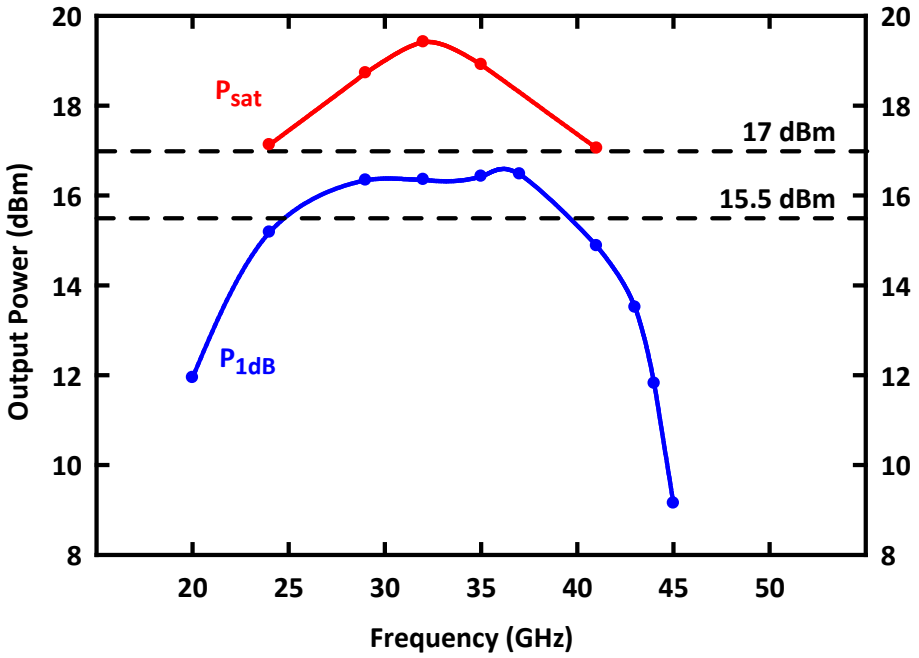


Figure 3.22: Measured saturated output power (P_{sat}) and output-referred power at the 1 dB gain compression point (P_{1dB}). P_{sat} measurements are limited by the available source power.

3.3.3 Simulated vs. Measured Data

The simulated DC bias points are very accurate (within 5–10% of measured values).

Fig. 3.25 shows the measured vs. simulated mismatch across frequency while Fig. 3.26 shows the data mismatch at the midband frequency. As expected, the VBIC models are overly optimistic predicting P_{1dB} and P_{sat}.

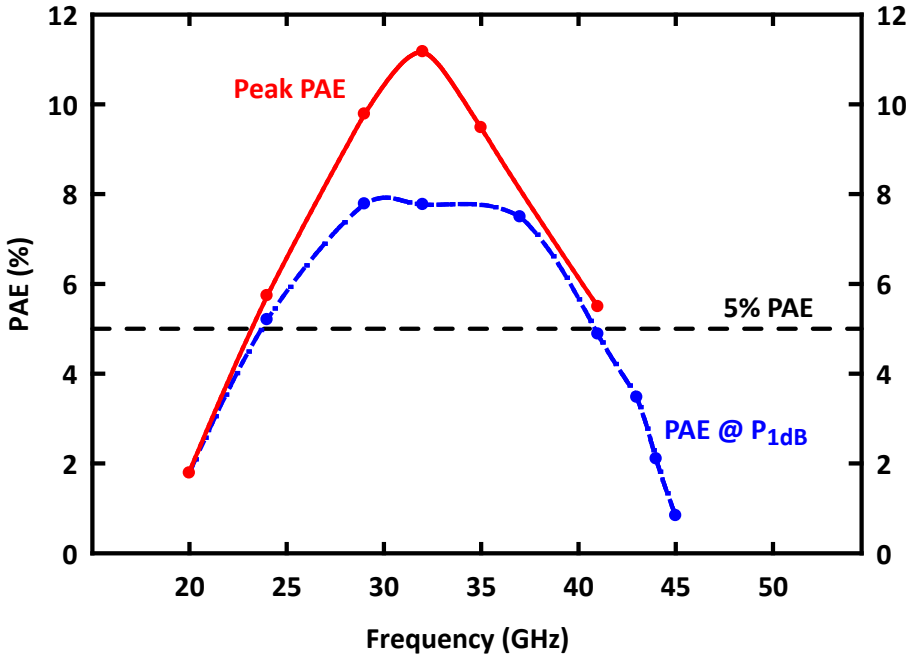


Figure 3.23: Measured peak PAE and PAE at $P_{1\text{dB}}$ output power. Peak PAE measurements are limited by available source power.

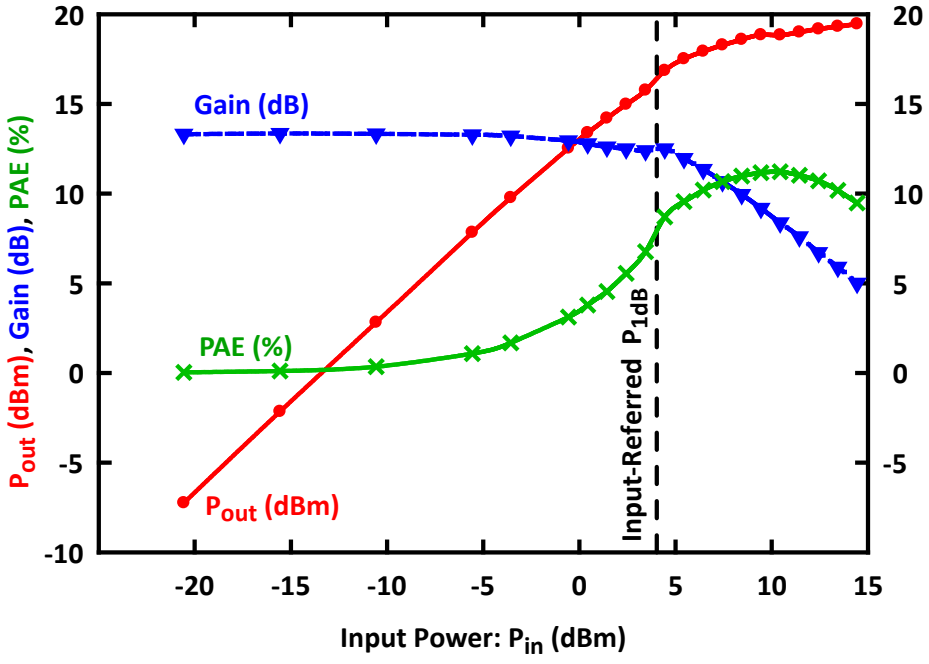


Figure 3.24: Measured midband large-signal characteristics at 32 GHz.

Table 3.1: Small-Signal Performance Summary

Input-Matching BW (> 8 dB RL)	28.5–49GHz
Output-Matching BW (> 8 dB RL)	29–45GHz
Maximum Gain	13dB
Small-Signal Gain > 10dB (3dB Bandwidth)	25.5–41GHz
Output-to-Input Isolation	> 40 dB
Quiescent Power Dissipation (375 mA from 1.4 V supply)	525 mW

Table 3.2: Large-Signal Performance Summary

Target Operation Bandwidth	26–40 GHz (42% BW)
Maximum Output P_{sat}	19.4 dBm
$P_{\text{sat}} > 17$ dBm	25–40 GHz
$P_{\text{1dB}} > 15.5$ dBm	25–40 GHz
Maximum PAE	11.2%

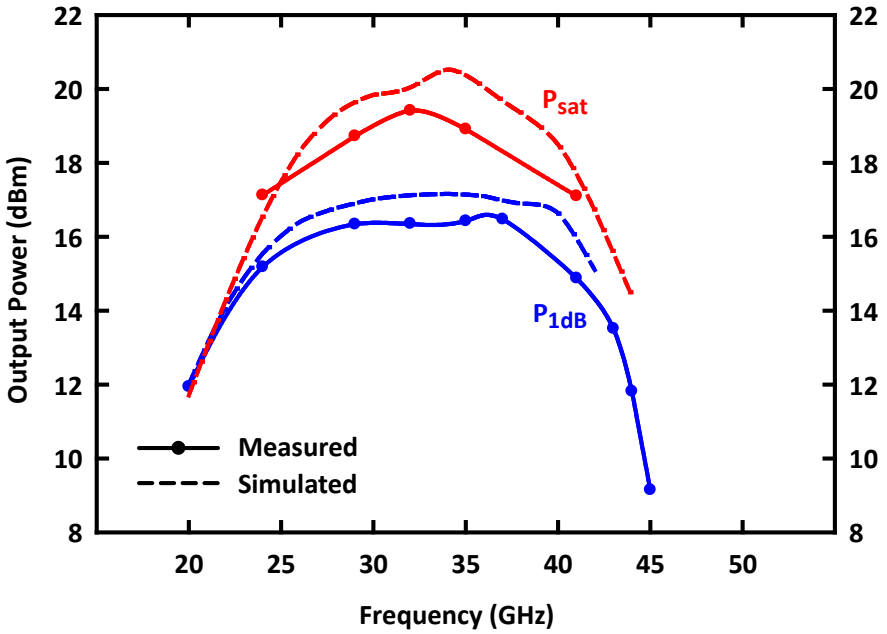


Figure 3.25: Simulated (dotted) vs. measured (solid) output saturated power and 1 dB compression points for the balanced PA.

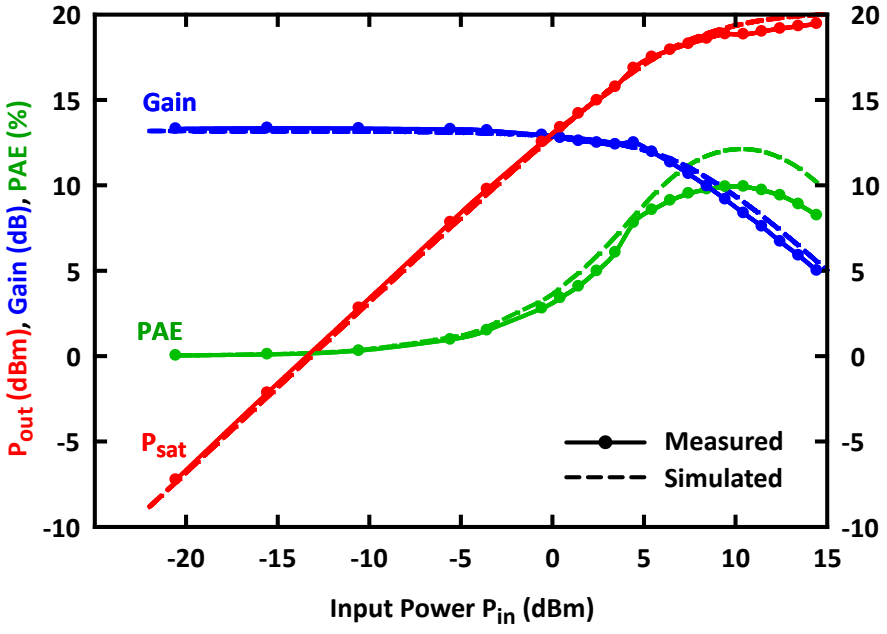


Figure 3.26: Simulated (dotted) vs. measured (solid) at midband frequency.

Small-signal measurements correlate nicely with the simulated results, as shown in Fig. 3.27- 3.29.

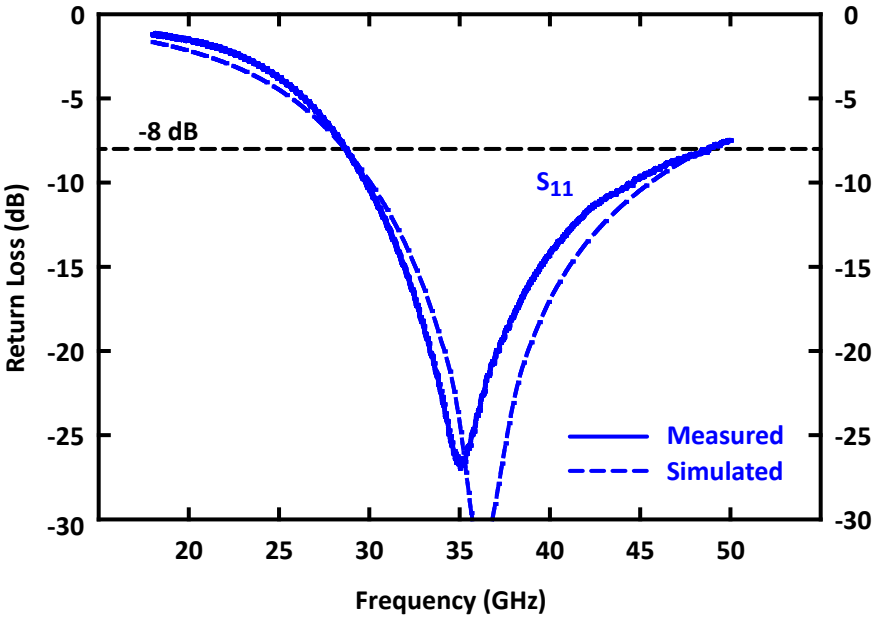


Figure 3.27: Simulated (dotted) vs. measured (solid) small-signal input return loss.

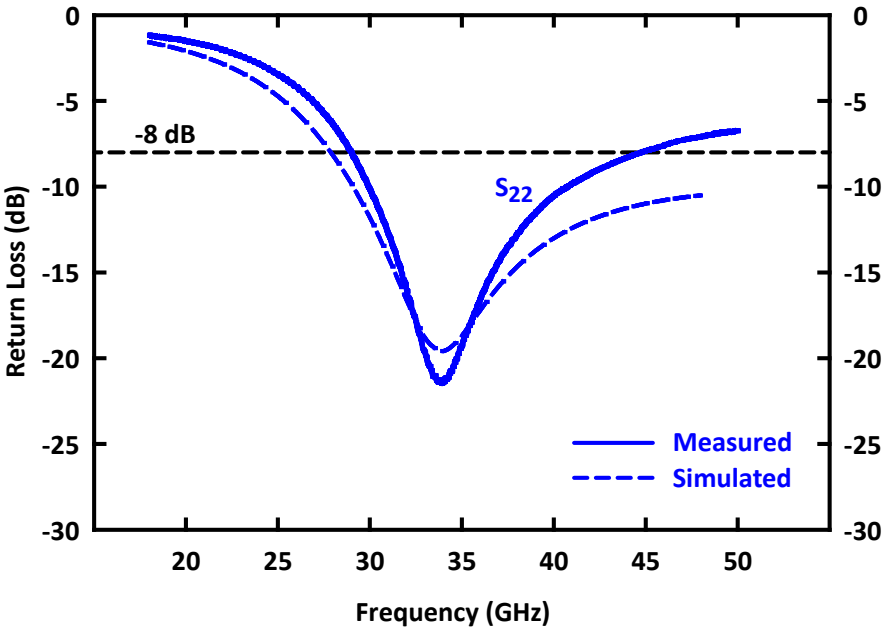


Figure 3.28: Simulated (dotted) vs. measured (solid) small-signal output return loss.

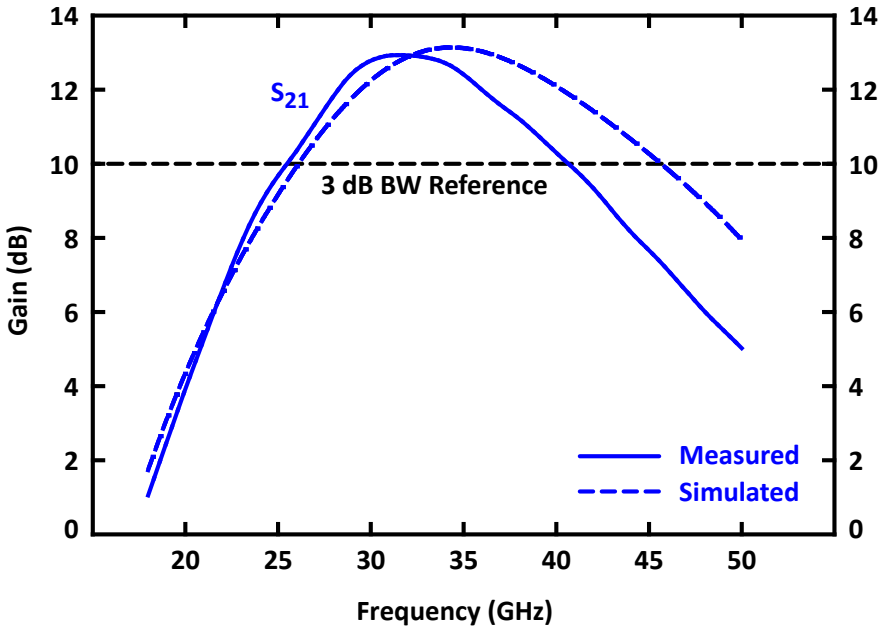


Figure 3.29: Simulated (dotted) vs. measured (solid) small-signal gain.

3.4 Summary of Completed Work

The SiGe PA presented here achieves a wide operational bandwidth while maintaining key performance parameters of gain, input/output matching, PAE, as well as linear and saturated output powers. Without employing distributed-amplifier (traveling-wave-amplifier) or negative-feedback design techniques, this integrated PA offers a 42% fractional-bandwidth performance within the Ka-frequency band. The measured results clearly demonstrate the capability and performance potential of SiGe PAs for wideband, linear power applications in highly-integrated SoC applications.

3.5 Addendum: Additional Data

Subsequent measurements have been taken at a collector voltage of 1.7 V with the quiescent currents re-optimized, improving saturated power performance across bandwidth at the cost of power-added efficiency (PAE). Fig. 3.30 shows the flattening of the P_{sat} response over frequency, providing a max P_{sat} of 19.5 dBm and P_{sat} greater than 15 dBm from 24 to 49 GHz. The higher collector voltage also results in a higher maximum small-signal gain of 15.8 dBm. The maximum PAE is 9.3%. Fig. 3.31 presents the small-signal input and output matching return losses. Fig. 3.32 shows that the output IP3 intermodulation metric is above 20 dBm from 25 to 40 GHz, with a peak OIP3 of 23 dBm.

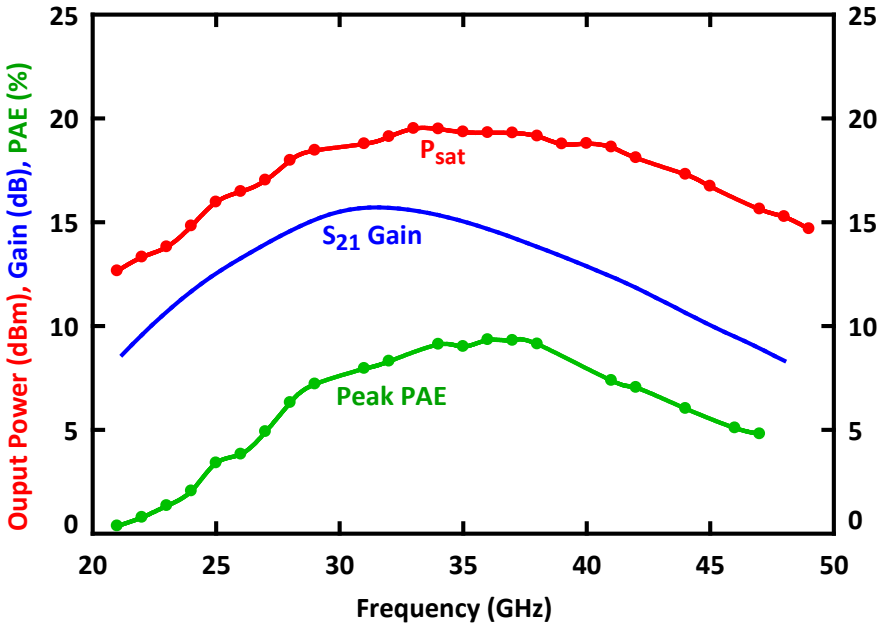


Figure 3.30: Ka-band power amplifier biased at 1.7 V. The max P_{sat} = 19.5 dBm, the max gain = 15.8 dB, and the max PAE = 9.3%.

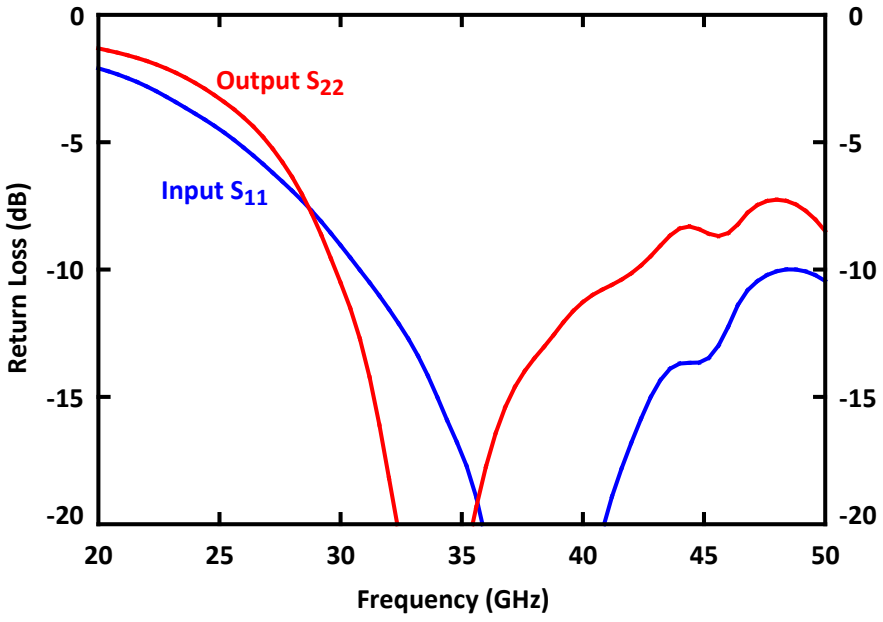


Figure 3.31: Small-signal input/output matching with 1.7 V collector bias.

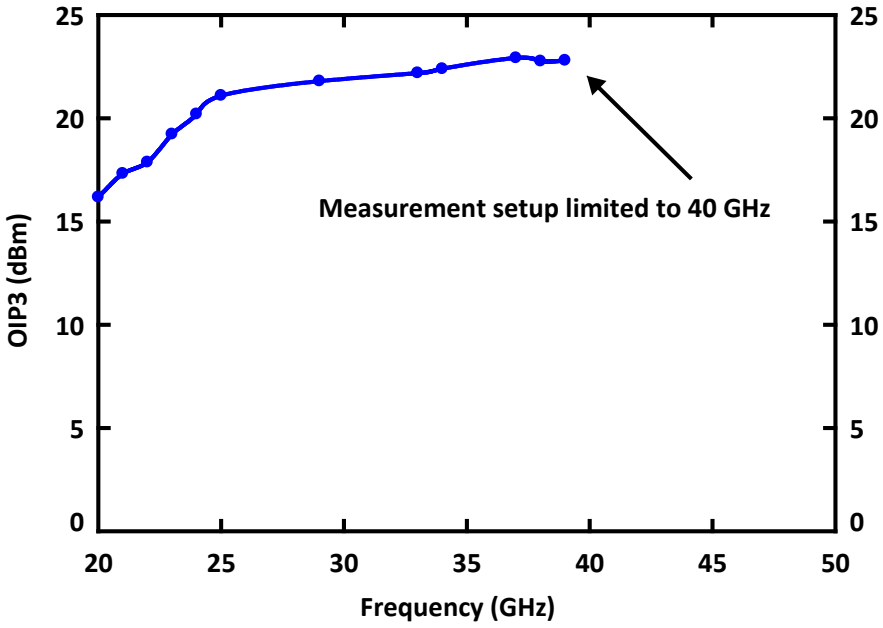


Figure 3.32: Linearity measurement based on output IP3 at 1 MHz offsets. The peak OIP3 = 23 dBm.

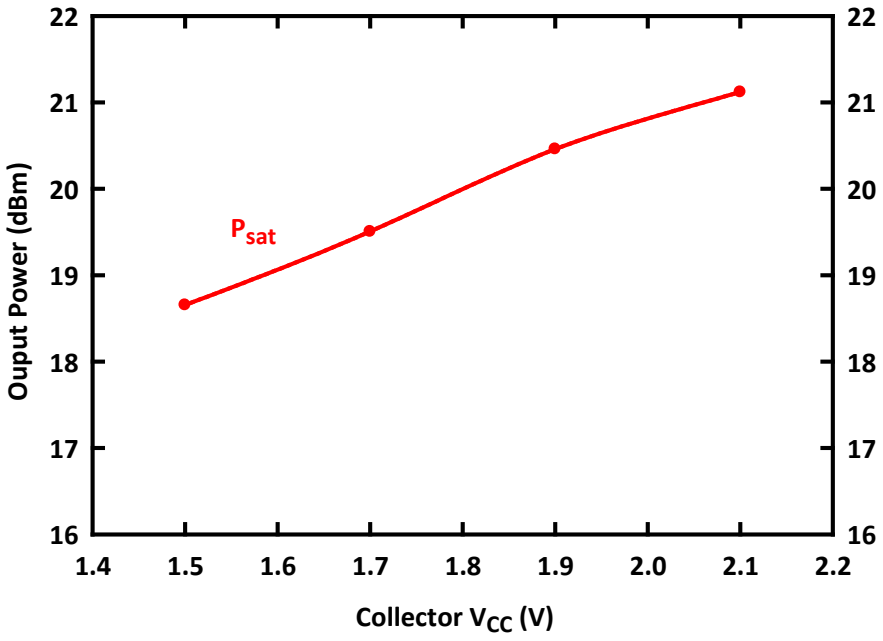


Figure 3.33: Saturated output power vs. collector voltage at midband frequency.

Finally, Fig. 3.33 shows that even higher output powers are achievable at higher collector voltage biases. If the collectors are biased at 2.1 V, the balanced power amplifier delivers 21.1 dBm (130 mW) of saturated output power. The increased power comes at the cost of reduced operational reliability and a degradation of small-signal return loss at the input and output (not shown).

3.5.1 Design Variations

After the successful demonstration that the Ka-band PAs can be pushed from $V_{CC} = 1.4$ V design values all the way to 2.1 V, a second iteration of the Ka-band design attempts to achieve even higher output powers with larger 80 µm and 90 µm transistors. The Class AB designs and their matching networks are more aggressively matched for high-power conditions near P_{1dB} output levels and higher gains, based on the same design techniques outlined earlier. The external base-biasing resistance is dropped from 40 Ω to 20 Ω, and many more MIM capacitors are integrated for even lower impedance references in the quarter-wavelength resonators (bias-insertion networks).

A sample of the measured performance of a balanced PA based on an 80 µm output device is shown in Fig. 3.34. The amplifier is biased at 1.9 V and offers a peak P_{sat} of 20.9 dBm and a maximum gain of 16.3 dBm.

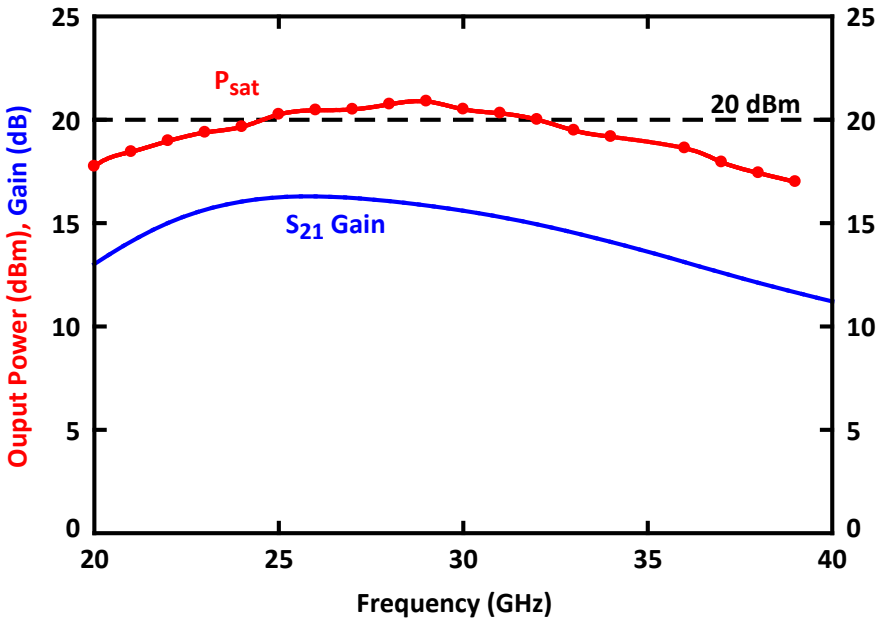


Figure 3.34: Output saturated power and small-signal gain from a balanced PA with 80µm output transistors. The peak $P_{sat} = 20.9$ dBm and the max gain = 16.3 dB at a collector bias of 1.9 V.

A variation with a slightly larger 90 µm output transistor improves on the output power as collector voltages are raised to V_{CC} = 2.2 V. The layout of the 90 µm HBT is shown in Fig. 3.35. The small-signal gain for the PA is presented in Fig. 3.36. The power optimization moves the peak small-signal gain center-frequency down to 27–28 GHz while the large-signal performance still peaks at 33 GHz.

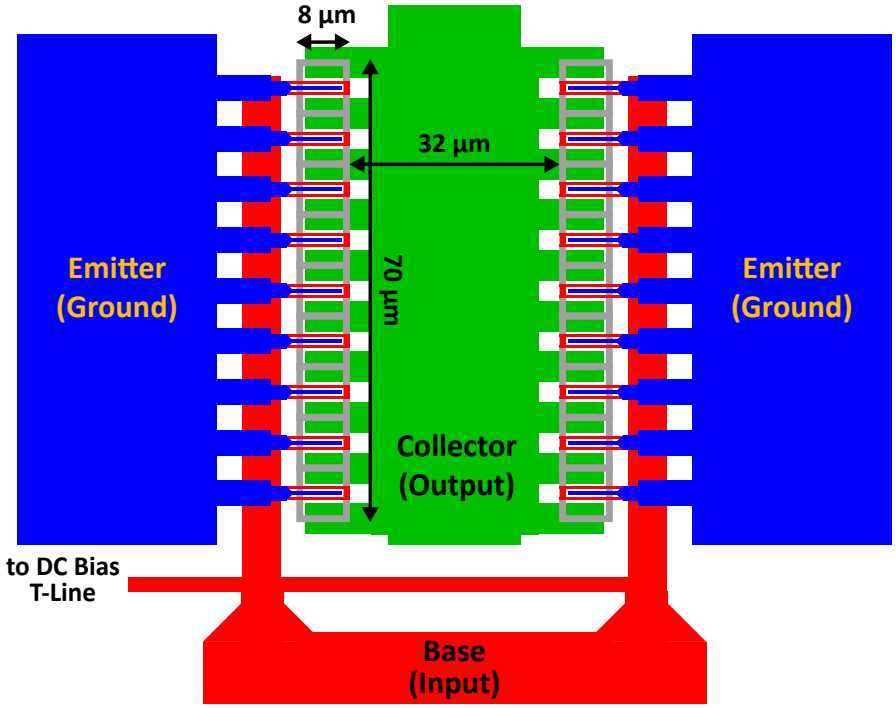


Figure 3.35: Ka-Band 90 µm output transistor: aggregated power transistor layout. Two 1 × 9 1D arrays of 5 µm cells are separated by 32 µm, which is the required deep-trench separation under the old 8HP local-density rules. For clarity, the figure only illustrates connectivity, and metal layering is not detailed. The emitters and collectors are wired up from the substrate to the top-level (AM) thick metal. The base is wired up to the top-level metal, as well, away from the power-cell core, to connect to RF transmission lines.

The measured large-signal performance achieves a P_{sat} = 24.2 dBm @ 33 GHz with V_{CC} = 2.2 V, P_{1dB} = 14.5 dBm, and peak-PAE = 22.3%. The small-signal gain is 14.6 dBm and has a 3dB-bandwidth from 20.5–36 GHz (55% BW). The design purposely trades off some linearity in Class AB operation for more output power, as evidenced by the lower P_{1dB} outputs despite having higher P_{sat} values.

Table 3.3 summarized two Ka-band designs: one from the original first iteration and a design from the second iteration with 90 µm output transistor and more aggressive power matching & bias-biasing resistance at 20 Ω. The table compares the Ka-band to all other

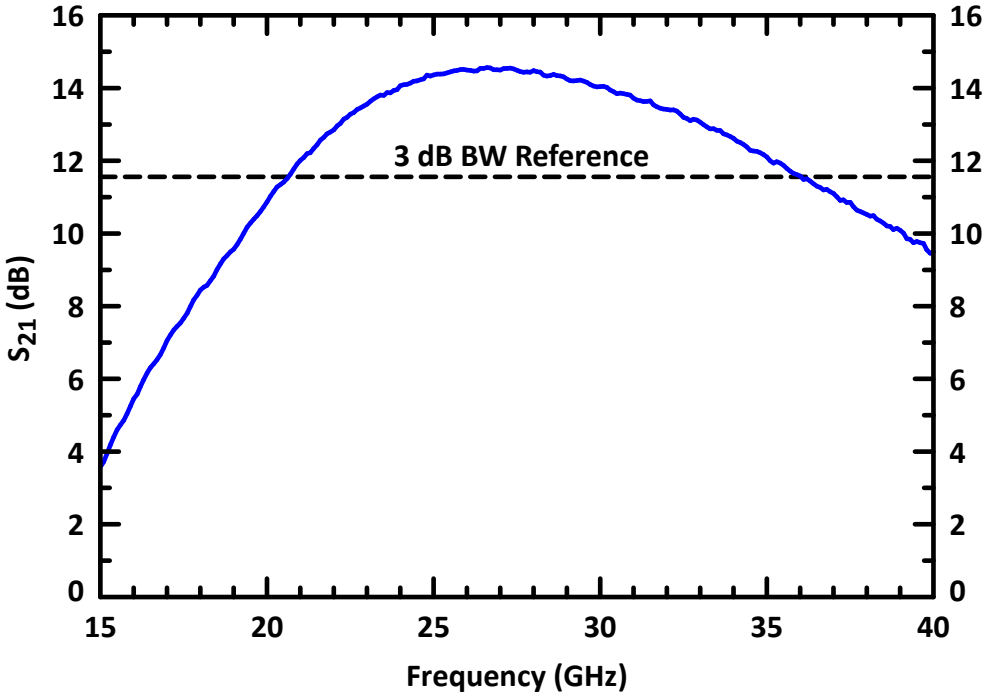


Figure 3.36: S_{21} (small-signal gain) for the Ka-band second-iteration PA with 90 µm output HBT.

published silicon PAs during the time of initial publication, including power-combined PAs and those in CMOS and other SiGe variants.

At the time of their publication, the presented Ka-band PAs achieved some of the highest output powers at Ka-band without power combining beyond differential operation, and 24.2 dBm is still the highest output-power result for SiGe non-power-combined linear PAs within Ka-band.

Ka-Band One-Way Linear PA Comparisons: SiGe

Table 3.4 compares the presented Ka-band PA to all one-way linear SiGe PAs near Ka-band, which includes recent publications. Power-combined SiGe PAs, aside from differential operation, are not included. Switch-mode SiGe PAs are not included.

Ka-Band One-Way Switch-Mode PA Comparisons: SiGe

Table. 3.5 compares recent research into non-linear, switch-mode PAs (Class E, F/F^{-1}, and others) near Ka-band. Only SiGe-based PAs are included in the table— no CMOS/SOI. Power-combined PAs, aside from differential operation, are not included in the table.

Table 3.3: Ka-band PA comparison at the time of original publication

Authors	Year	Technology	Lith (nm)	f_T (GHz)	Freq (GHz)	3-dB BW (GHz)	P_{sat} (dBm)	P_{1dB} (dBm)	Peak PAE (%)	S.S. Gain (dB)	Supply (V)	Topology	Stages	PA Mode	Input SE/Diff	Output SE/Diff	Baluns (if diff)	Power Combining	Class	Comments
This Work Design #2 @ Optimized Bias	2007	SiGe BiCMOS	130	200	**33**	20.5–36	**24.2**	14.5	**22.3**	14.6	2.2	CE	2	Diff	Diff	Diff	None	-	A/AB	3 dB added to SE measurements
This Work Design #1 @ Optimized Bias	2007	SiGe BiCMOS	130	200	**30**	26–40	**21.3**	> 15.5	**12.9**	15.5	2.1									
This Work Design #1 @ Nominal Bias	2007	SiGe BiCMOS	130	200	**32**	26–40	**19.4**	16.4	**11.2**	13.0	1.4									
Chartier [25]	2004	SiGe	500	80	**36**	33–39[a]	**NR**	10.0	**1.5**	15.3	7.0	Cascode	3	SE	SE	SE	-	-	A/AB	
Riemer [22]	2007	SiGe BiCMOS	130	200	**30**	29.5–34.3	**19.4**	15.4	**6.2**	46.8	1.4	CE	4	SE	SE	SE	-	4	A/AB	
Comeau [27]	2006	SiGe BiCMOS	180 CMOS	150	**24**	23.3–25.8	**22.0**	20.0	**14.0**	12.0	5.1	Cascode	1	SE	SE	SE	-	-	AB	
Kinayman [24]	2005	SiGe	180	80	**24**	20.5–27.5[a]	**12.0**	NR	**2.9**	18.0	5.0	CE	3	Truly Diff	Diff	Diff	None	-	A	
Komijani [28]	2005	CMOS	180	65	**24**	22.9–26	**14.5**	11.0	**6.5**	7.0	2.8	Cascode	2	SE	SE	SE	-	-	AB	
Chartier [25]	2004	SiGe	500	80	**24**	21.5–27.5[a]	**NR**	12.0	**2.7**	17.7	7.0	Cascode	3	SE	SE	SE	-	-	A/AB	
Cheung [20]	2005	SiGe BiCMOS	200	120	**22**	NR	**23.0**	NR	**19.7**	19.0	1.8	CB	3	Diff	SE	SE	Integrated	2	AB	

[a]Estimated from graph

Table 3.4: Ka-band one-way linear SiGe PA comparison

Authors	Year	Technology	Lith (nm)	f_T (GHz)	Freq (GHz)	3-dB BW (GHz)	P_{sat} (dBm)	P_{1dB} (dBm)	Peak PAE (%)	S.S. Gain (dB)	Topology	Stages	PA Mode	Input SE/Diff	Output SE/Diff	Baluns (if diff)	Class	Comments
This Work Design #2 @ Optimized Bias	2007	SiGe BiCMOS	130	200	**33**	20.5–36	**24.2**	14.5	**22.3**	14.6	CE	2	Diff	Diff	Diff	None	A/AB	3 dB added to SE measurements
This Work Design #1 @ Optimized Bias	2007	SiGe BiCMOS	130	200	**30**	26–40	**21.3**	> 15.5	**12.9**	15.5								
This Work Design #1 @ Nominal Bias	2007	SiGe BiCMOS	130	200	**32**	26–40	**19.4**	16.4	**11.2**	13.0								
Dabag [141]	2011	SiGe BiCMOS	120	200	**37.5**	32–47	**14.8**	NR	**26.2**	11.5	CE	1	SE	SE	SE	-	A	
Kalantari [142]	2012	SiGe BiCMOS	130	200	**38**	36–41	**23.0**	NR	**10.7**	18.4	CE with NeRF	3	SE	SE	SE	-	AB	Nested reactance feedback
					38	36–41	**21.3**	NR	**20.0**	18.7								
Chartier [25]	2004	SiGe	500	80	**36**	33–39[a]	**NR**	10.0	**1.5**	15.3	Cascode	3	SE	SE	SE	-	A/AB	
Liu [41]	2016	SiGe BiCMOS	130	250	**35**	30–40	**22.5**	20.5	**NR**	14[b]	4-Stacked	2	SE	SE	SE	-	AB	
Pei [143]	2014	SiGe:C BiCMOS	250	210	**34**	34–39[a]	**17**[a]	NR	**15.5**[a]	31[a]	Cascode	2	Diff	SE	SE	Integrated	A	
Ma [144]	2015	SiGe BiCMOS	180	NR	**32**	22–30[a]	**11.1**	9.6	**55.9**	16.0	2-Stacked	1	Diff	Diff	Diff	None	A/AB	
Farmer [145]	2011	SiGe BiCMOS	130	200	**30**	NR	**19.0**	16.6	**11.5**	13.0	4-Stacked	1	SE	SE	SE	-	A/AB	
Balducci [73]	2017	SiGe BiCMOS	250	120	**29.5**	27–31.5	**13.9**	12.3	**17.5**	16.5	Cascode	1	Diff	Diff	SE	Integrated output balun	A/AB	3 dB added to SE measurements
Chen [146]	2015	SiGe:C BiCMOS	250	210	**28**	NR	**13.0**	8.5[a]	**NR**	18[a]	Cascode	2	Diff	Digital	SE	Integrated	A	
Welp [147]	2016	SiGe	180	170	**24**	NR	**24.7**	23.7	**31.0**	18.5	Cascode	1	Truly Diff	Diff	Diff	None	A/AB	
Comeau [27]	2006	SiGe BiCMOS	180 CMOS	150	**24**	23.3–25.8	**22.0**	20.0	**14.0**	12.0	Cascode	1	SE	SE	SE	-	AB	
Kinayman [24]	2005	SiGe	180	80	**24**	20.5–27.5[a]	**12.0**	NR	**2.9**	18.0	CE	3	Truly Diff	Diff	Diff	None	A	
Chartier [25]	2004	SiGe	500	80	**24**	21.5–27.5[a]	**NR**	12.0	**2.7**	17.7	Cascode	3	SE	SE	SE	-	A/AB	

[a] Estimated from graph

[b] Includes gain/loss of preceding TX front-end circuits

Table 3.5: Ka-band switch mode PA comparison

Authors	**Year**	**Technology**	**Lith (nm)**	f_T **(GHz)**	**Freq (GHz)**	**3-dB BW (GHz)**	$P_{\mathbf{sat}}$ **(dBm)**	P_{1dB} **(dBm)**	**Peak PAE (%)**	**Topology**	**Stages**	**PA Mode**	**Class**
Datta [50]	2014	SiGe BiCMOS	130	NR	**45**	**20.2**	NR	**31.5**	10.5	CE	2	SE	E
					41	**18.1**		**35.5**	5.6	CE	1	SE	
					41	**23.4**		**34.9**	14.5	2-Stacked	2	SE	
					40	**22.2**		**20.8**	15.4	3-Stacked	2	SE	
Datta [149]	2013	SiGe BiCMOS	130	NR	**41**	**23.4**	NR	**34.9**	14.6	2-Stacked	2	SE	E
Mortazavi [150]	2016	SiGe BiCMOS	130	NR	**40.5**	**18.0**	16.0	**43.0**	22.0	CE	2	SE	F^{-1}
Mortazavi [151]	2016	SiGe BiCMOS	130	180	**38**	**16.5**	15.0	**38.5**	16.5	CE	2	SE	F^{-1}
					24	**18.0**	16.0	**50.0**	21.0				
Mortazavi [152]	2015	SiGe BiCMOS	130	180	**38**	**16.5**	15.0	**38.5**	15–17	CE	2	SE	F^{-1}
Sarkar [44]	2014	SiGe BiCMOS	130	200	**28**	**18.6**	15.5	**35.3**	15.3	Cascode	1	SE	J
Mortazavi [45]	2015	SiGe BiCMOS	130	180	**28**	**17.1**	15.0	**42.0**	21.2	CE	2	SE	F^{-1}
Mortazavi [46]	2014	SiGe BiCMOS	130	NR	**28**	**17.1**	15.0	**40.7**	10.3	CE	1	SE	F^{-1}/F

Ka-Band One-Way Linear PA Comparisons: CMOS/SOI

Table. 3.6 summarizes recent CMOS/SOI one-way linear PAs near Ka-band. Non-linear switch-mode CMOS/SOI PAs and power-combined PAs, aside from differential operation, are not included in the table.

Dot Plots: P_{sat} % Peak-PAE

Table. 3.37 presents a dot plot of reported P_{sat} for the presented Ka-band PAs as well as other one-way PAs from 24–45 GHz: SiGe linear one-way PAs, SiGe switch-mode one-way PAs, and CMOS/SOI linear one-way PAs.

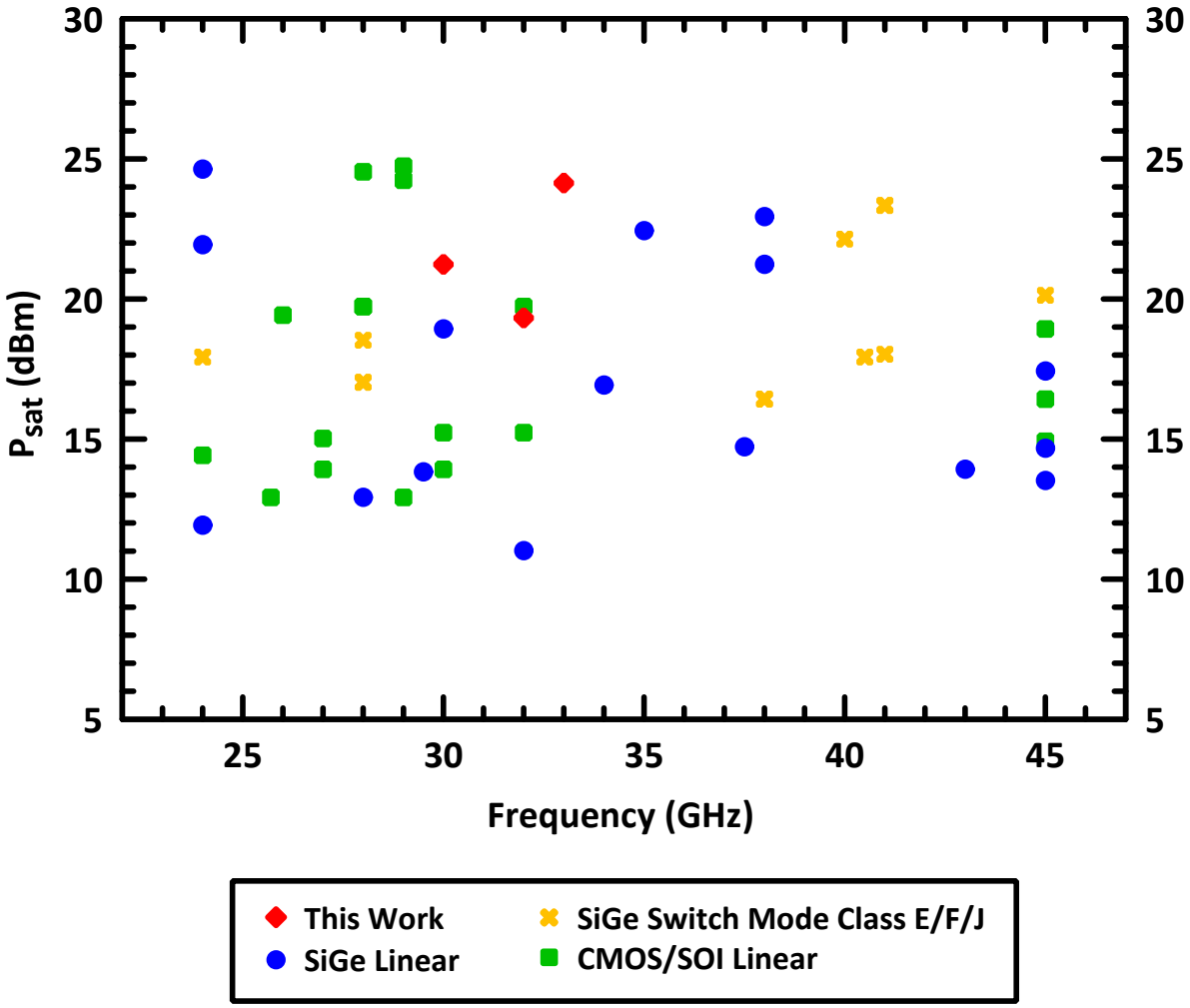


Figure 3.37: Saturated output power P_{sat} for Ka-band one-way PAs.

Table 3.38 presents a dot plot of the reported peak-PAE for the presented Ka-band PAs as well as other one-way PAs from 24–45 GHz: SiGe linear one-way PAs, SiGe switch-mode

Table 3.6: Ka-band one-way linear CMOS/SOI PA comparison

Authors	**Year**	**Technology**	**Lith (nm)**	**Freq (GHz)**	P_{sat} **(dBm)**	P_{1dB} **(dBm)**	**Peak PAE (%)**	**S.S. Gain (dB)**	**Topology**	**Stages**	**PA Mode**	**Input SE/Diff**	**Output SE/Diff**	**Baluns (if diff)**	**Class**
Indirayanti [158]	2017	Bulk CMOS	28	**32**	**19.8**	16.0	**21.0**	22.0	Doherty	2	Diff	SE	SE	Integrated	AB & C
Jia [159]	2016	Bulk CMOS	65	**32**	**15.3**	12.9	**32.9**	20.8	CS	2	Diff	SE	SE	Integrated	A/AB
Shakib [160]	2016	Bulk CMOS	28	**30**	**15.3**	14.3	**36.6**	16.3	CS	2	Diff	SE	SE	Integrated	NR
Shakib [161]	2016	Bulk CMOS	28	**30**	**14.0**	13.2	**35.5**	15.7	CS	2	Diff	SE	SE	Integrated	NR
Jayamon [35]	2016	CMOS SOI	45	**29**	**24.3**	NR	**29.0**	13.0	4-Stacked	1	SE	SE	SE	-	AB (close to B)
					24.8	21.0	**26.0**	NR							AB
Alsuraisry [162]	2017	CMOS	180	**29**	**13.0**	11.5	**25.3**	12.6	CS	2	SE	SE	SE	-	AB
Helmi [163]	2016	CMOS SOI	45	**28**	**24.6**	23.0	**15.0**	17.3	3-Stacked Triple-Cascode	1	SE	SE	SE	-	AB
Park [164]	2016	Bulk CMOS	28	**28**	**19.8**	18.6	**43.3**	13.6	2-Stacked	1	Truly Diff	SE	SE	Integrated	Deep AB
Shakib [165]	2017	CMOS	40	**27**	**15.1**	13.7	**33.7**	22.4	CS	3	Diff	SE	SE	Integrated	NR
Lee [166]	2008	CMOS	180	**27**	**14.0**	NR	**13.2**	14.5	CS	3	SE	SE	SE	-	NR
Kuo [167]	2013	CMOS	180	**26**	**19.5**	16.0	**10.2**	15.2	Darlington Cascode	2	Diff	Diff	Diff	External (de-embedded)	A
Vasylyev [168]	2005	CMOS	130	**25.7**	**13.0**	NR	**13.2**	8.4	Cascode	1	Diff	SE	SE	Integrated	AB
Komijani [28]	2005	CMOS	180	**24**	**14.5**	11.0	**6.5**	7.0	Cascode	2	SE	SE	SE	-	AB

one-way PAs, and CMOS/SOI linear one-way PAs.

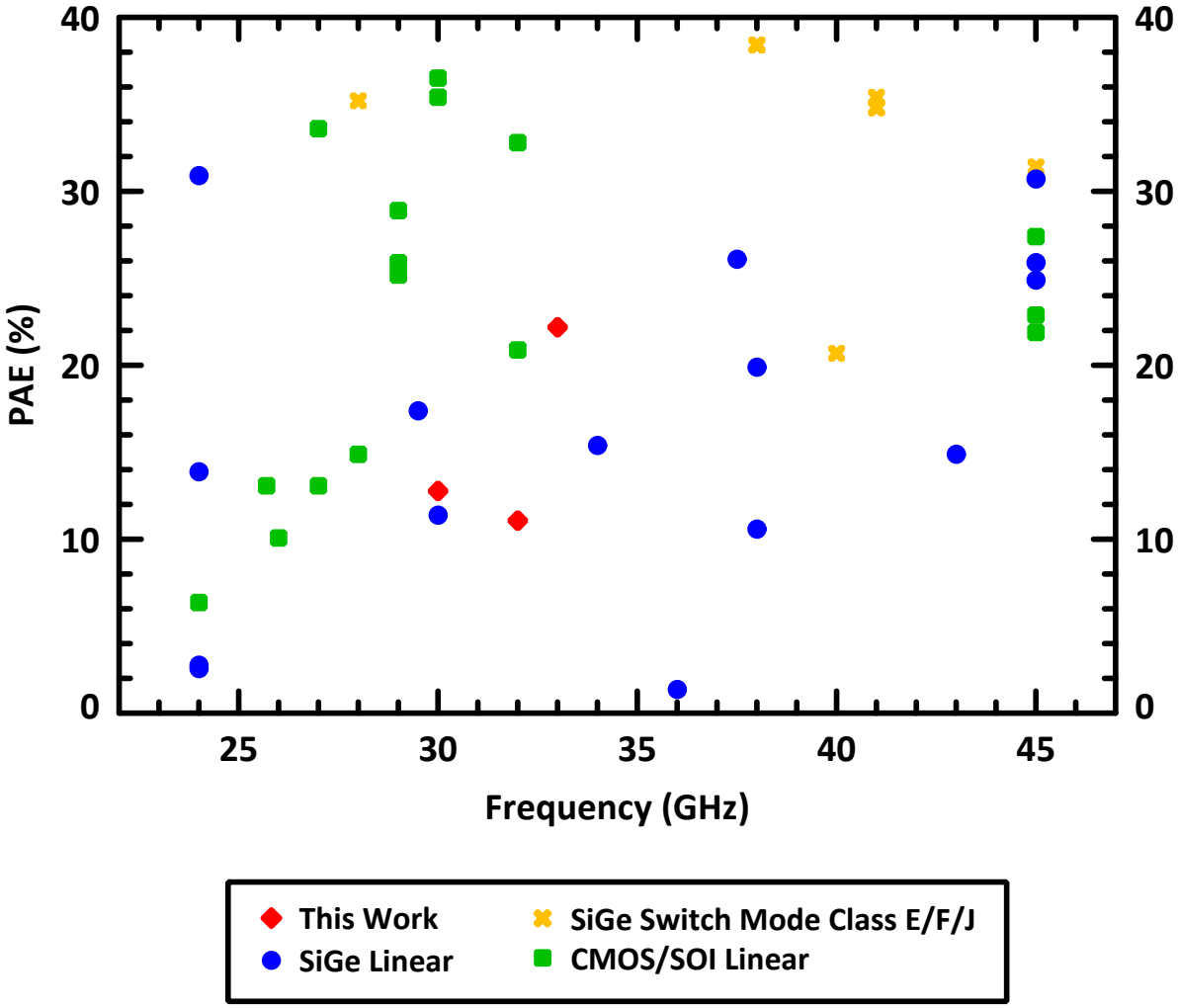


Figure 3.38: Survey of silicon power amplifiers and their power added efficiency (PAE) for Ka-band one-way (non-power-combined) PAs. The Ka-band PAs in this work (and breakout variants) are in red.

The values for the Ka-band SiGe linear, one-way power amplifiers in the dot plots are quoted from [24, 25, 25, 27, 41, 43, 73, 141–148].

The values for the Ka-band SiGe switch-mode power amplifiers in the dot plots are quoted from [44–46, 50, 149–152].

The values for the Ka-band CMOS/SOI linear, one-way power amplifiers in the dot plots are quoted from [28, 35, 158–169].

Chapter 4

Design Extensions and Future Work

This chapter presents a designed and measured full-integrated 3 × 3 power-amplifier array based on the W-band PA presented in Chapter 2. The latter part of this chapter outlines areas of future research opportunities for improving the mmW SiGe linear power amplifiers.

4.1 Design Extension: W-Band 3 × 3 PA Array

An extension of the W-band power amplifier design has been implemented into a 3 × 3 array of W-band power amplifiers. The W-band PA presented previously in Chapter 2 has been successfully integrated into a fully-integrated 3 × 3 wafer-scale quasi-optical spacial power-combining system [9]. The work has been done in collaboration with Yusuf Atesal, Berke Cetinoneri, and Ramadan Alhalabi at Univ. of California, San Diego, with Prof. Gabriel Rebeiz. Y. Atesal & B. Cetinoneri added and re-tuned an additional 4th input-driver gain stage to the W-band power amplifier, designed RF input/output and DC distribution/combining networks, and completed lab measurements of the 3 × 3 array. R. Alhalabi designed the microstrip-on-quartz antennas that are aligned & bonded to the SiGe chip with PA and antenna feeds.

This section focuses on matters pertaining to the standalone W-band PA integration into the wafer-scale array and only shows selected measurements results of the collaborative work. More technical details regarding the RF distribution, antenna design, and pattern measurements can be found published in [9].

The prior W-band PA work presented earlier is used as the basis for the PA elements within the 3 × 3 PA array operating at 94 GHz. The previously presented W-band design W2, centered at 94 GHz, is the PA element used within the array. A fourth common-emitter gain stage is added to the W-band PA element in order to increase the overall system gain. A separate 15 dB 4-stage medium-power (6 dBm) input-driver amplifier is adapted from the PA element and optimized for lower output powers.

Fig. 4.1 presents the schematic of the W-band PA with additional common-emitter input stage. The extra gain stage uses an additional 2×5 μm $= 10$ μm aggregated unit-cell transistor with matching networks re-optimized for small-signal gains. The additional gain stage follows the same unit-cell approach and re-uses unit-cell elements.

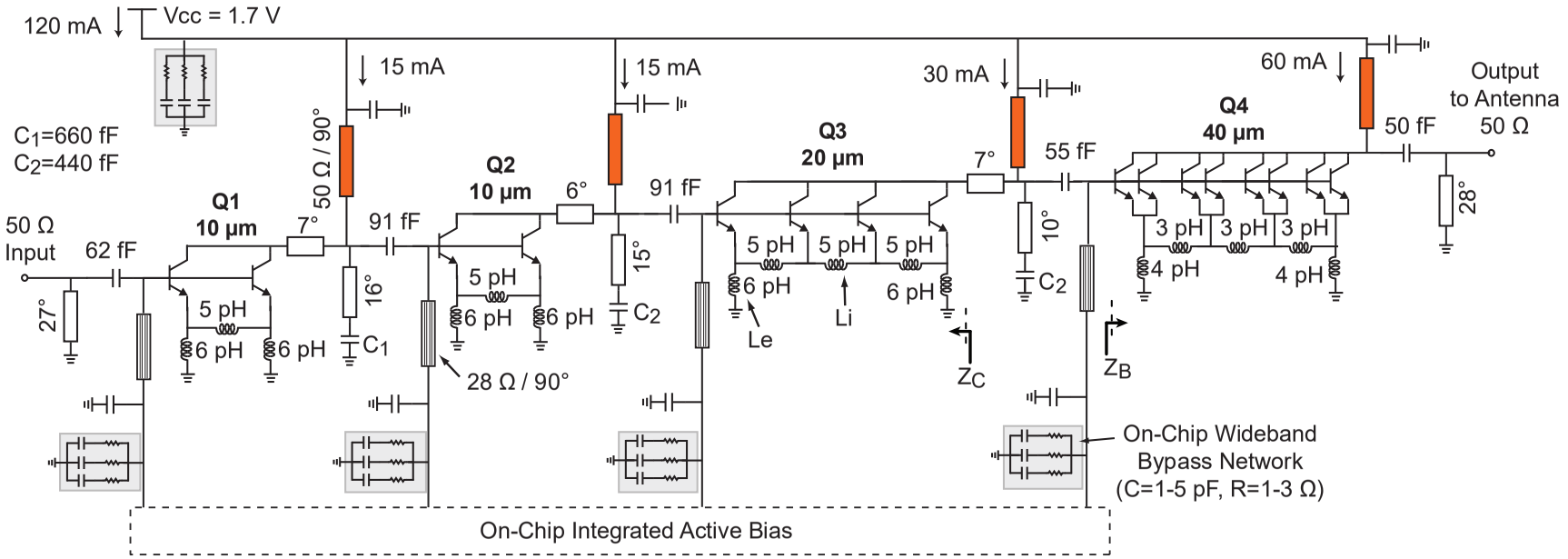


Figure 4.1: Circuit schematic of the W-band, 4-stage common-emitter power amplifier at 94 GHz. Unlabeled capacitor are 440 fF self-resonant unit-cell MIM capacitors used as RF-shorts.

New early versions of the improved HiCUM, high-current HBT models became available and allowed re-centering and re-optimization of design values based on improved transistor modeling. Measured W-band PA data from Chapter 2 has been included in the overall design optimization, as well.

The 48 mm^2 SiGe chip integrates multiple 4-stage PAs, integrated active-bias circuits, RF power distribution networks, DC distribution networks, and antenna feedlines. The SiGe chip is then bonded to a separate quartz substrate with an array of antennas. Fig. 4.2 presents the chip photomicrograph of the fabricated 3 × 3 PA array in GF SiGe 8HP.

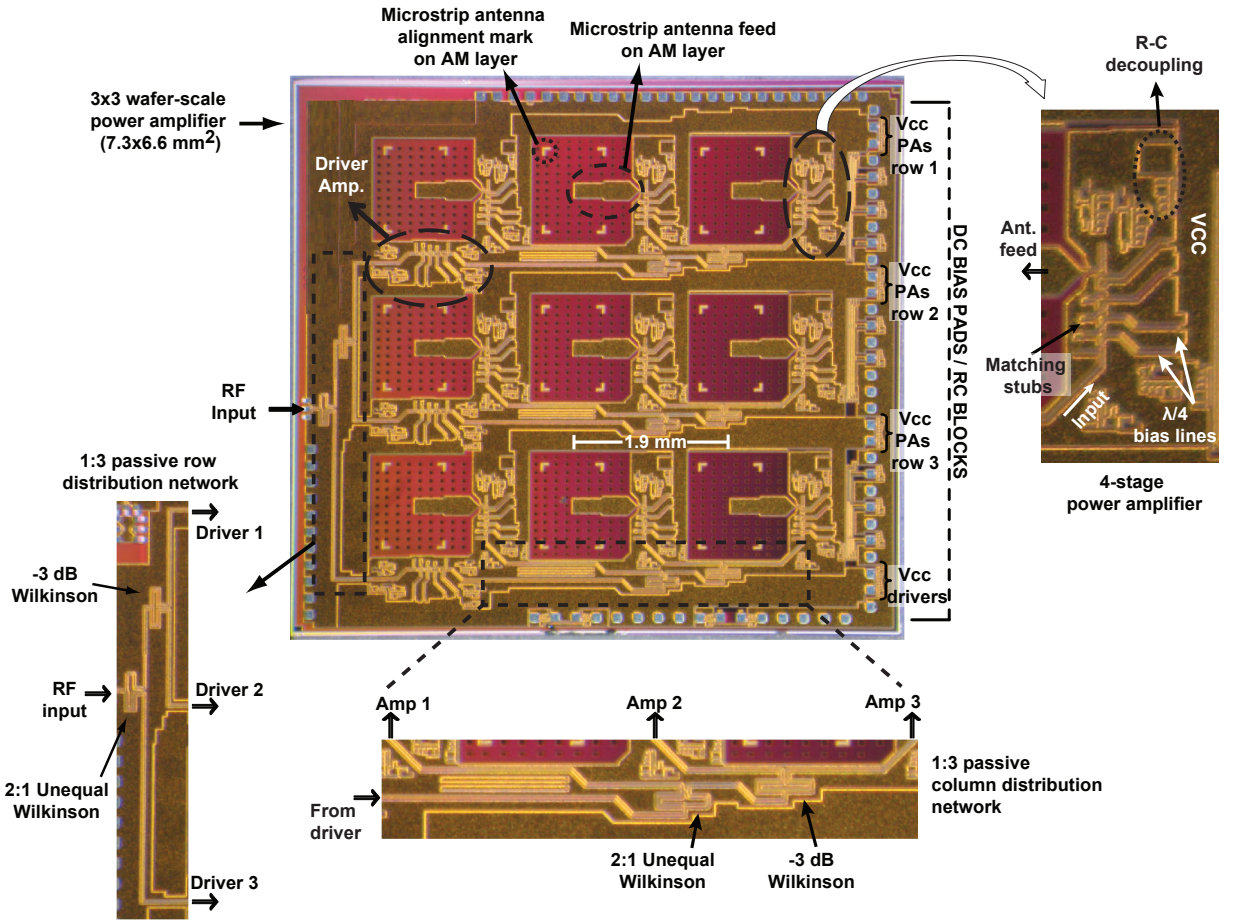


Figure 4.2: Photomicrograph of implemented 7.3 mm × 6.6 mm SiGe chip: 3 × 3 PA array without antenna substrate

The integrated array avoids on-chip output-combining losses in silicon by combining the PA output powers in free space via a quasi-optical power-combining array: spatial power combining using integrated antennas. Power-combining efficiency is important in keeping the total chip power dissipation within electromigration and thermal limits. Planar on-chip

power combining suffers from large combining losses for 8–16 elements or more because of the limited on-chip Q of RF passives at mmW frequencies.

The RF input signal is distributed across the PA array using Wilkinson couplers and shielded transmission lines. The RF input power is split equally into three branches using a combination of unequal 2:1 Wilkinson dividers and equal Wilkinson dividers, as illustrated in Fig. 4.3. Total interconnect loss from transmission lines and Wilkinson dividers are equalized among the 9 array elements. PA elements with one fewer Wilkinson divider have longer transmission feed lines, which equalizes the total RF signal loss. An active 15 dB driver amplifier is embedded within the divider network to compensate for the losses in order to provide -3.5 dBm input power into each of the 9 PA elements.

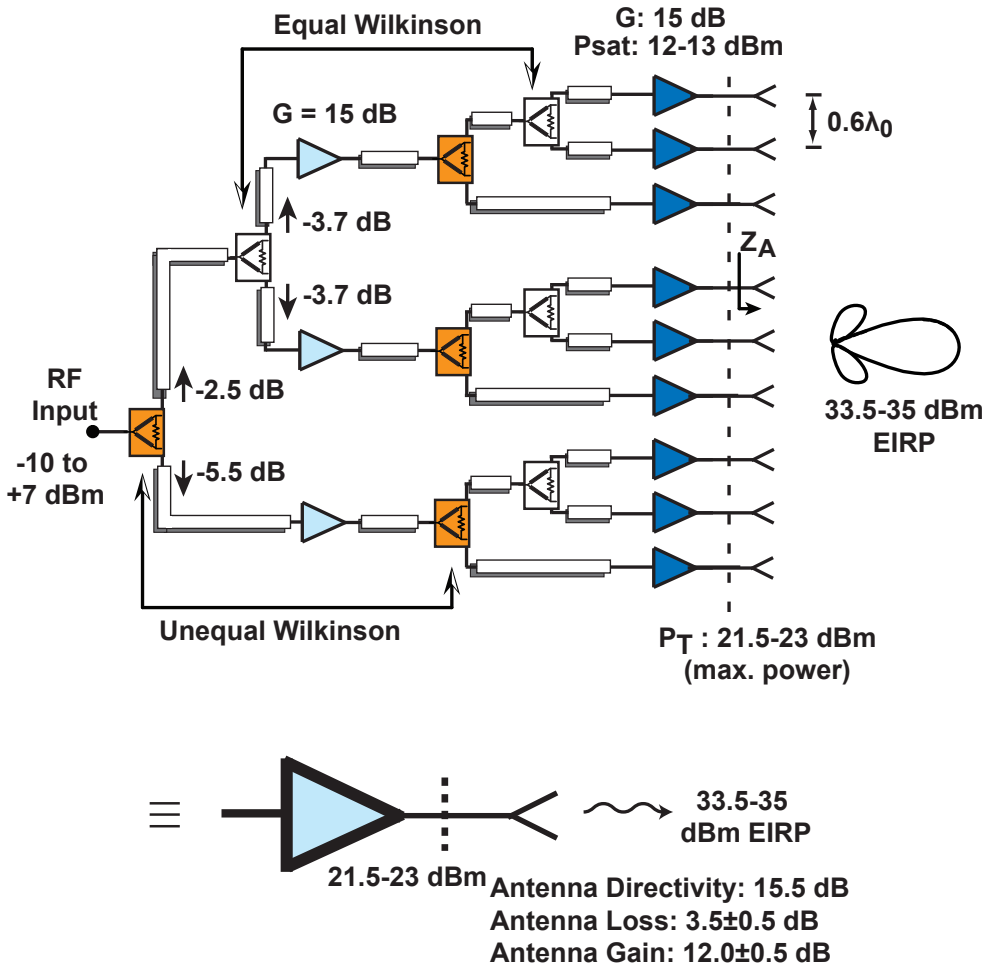


Figure 4.3: Block diagram of 3 × 3 W-band PA array: all elements except the microstrip antenna are implemented within the same SiGe chip & substrate. The feed-probes for the microstrip antennas are integrated within the SiGe chip.

Fig. 4.4 illustrates the SiGe on-chip powers throughout the RF distribution network. Active gain amplifiers, derived from re-optimizations of the output-PA element for lower output powers, are used to overcome power-splitting losses. The final RF input power to each 3 × 3 PA element is equalized throughout the array.

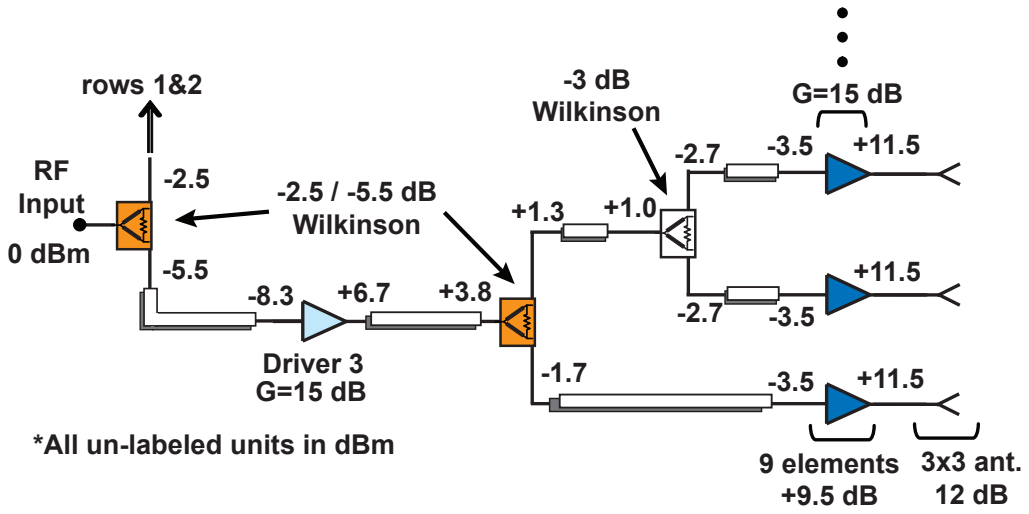


Figure 4.4: RF distribution networks within the PA array chip with active amplifiers to overcome power-splitting losses.

The outputs of the power amplifiers are combined quasi-optically in free-space using an array of antennas rather than on-chip power-combining techniques. Input combining losses can often be overcome with higher gain driver amplifiers to ensure power amplifier output stages are fully in saturation, but losses at the *output* of a power amplifier directly affect the attainable maximum output powers. Planar power-combining techniques within silicon are large & lossy at W-band, so on-chip output-power combining techniques such as Wilkinson combiners and transformers face challenging RF losses. These on-chip techniques also consume large areas, and power handling for large arrays of PAs becomes especially challenging as metal lines must scale considerably to be electromigration compliant while avoiding RF interconnect losses.

The output power combining is done in the quasi-optical domain with high-efficiency microstrip antennas integrated in a separate substrate bonded onto the SiGe chip. The microstrip antenna feeds are fully integrated on-chip as done in [175, 176]. Quasi-optical power combining in free space results in nearly perfect spatial power combining, not including the microstrip patch antenna-feed and antenna losses.

Fig. 4.5 shows the photomicrograph of the bonded quartz substrate onto the SiGe die.

The antenna elements are spaced $0.6 \times \lambda_0 = 1.9$ mm apart at 94 GHz. The microstrip

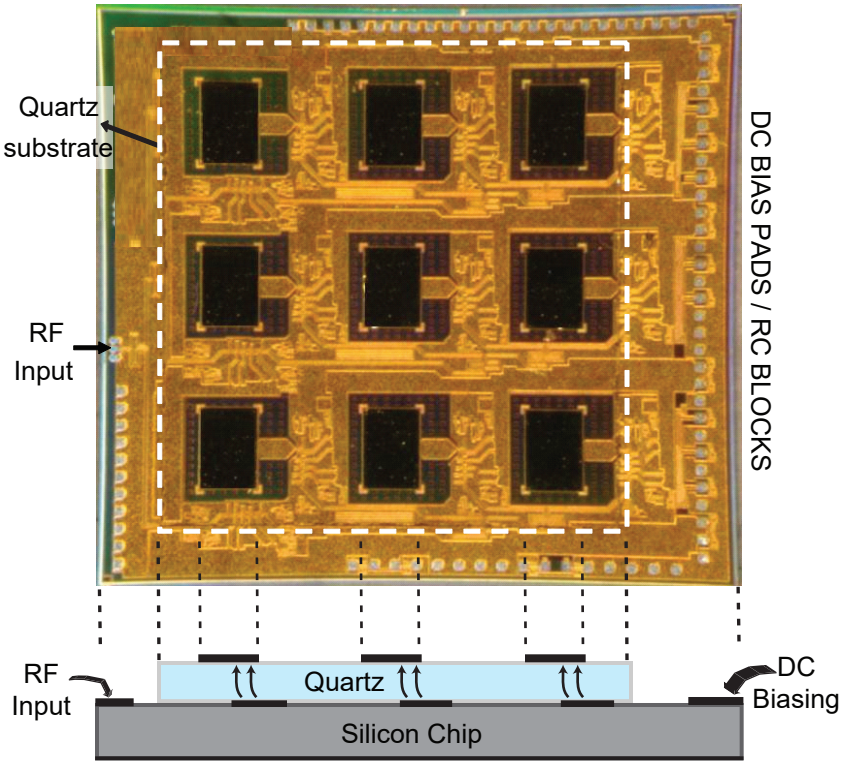


Figure 4.5: Photomicrograph with the top-view of the combined 3 × 3 antenna array bonded to the SiGe chip W-band PA array below.

patch antennas reside on a 125 μm-thick quartz substrate that is electromagnetically-coupled to the PA outputs. The antenna feed on top-layer AM metal of the SiGe die couples via fringing-field to the microstrip patch antenna, and no openings or vias are required on the SiGe chip or quartz substrate for the feed. The RF global ground plane (on layer MQ) on the SiGe chip also acts as the ground plan for the SiGe-integrated feed as well as the microstrip antenna on the quartz substrate.

The spatial power combining of 9 PA elements leads to an equivalent power gain of $10\log_{10}(9) = 9.5$ dB over a single PA element. A total of 21–23 dBm is available on-chip to feed the subsequent microstrip antenna array. The output-combining efficiency is 45%, including the antenna losses. The measured radiated power effective isotropic radiated power (EIRP) of the entire PA array is 33–35 dBm at 90–98 GHz.

Fig. 4.6 presents the measured E-plane and H-plane array patterns at 94 GHz. The measured 3dB-beamwidth is 28° at 94 GHz. Cross-polarization levels are below 25 dB between the E-plane and H-plane at 90–90 GHz. The measured patterns demonstrate that all 9 array elements radiate in-phase in a single mode, and coherent summing is successfully achieved in the spacial domain.

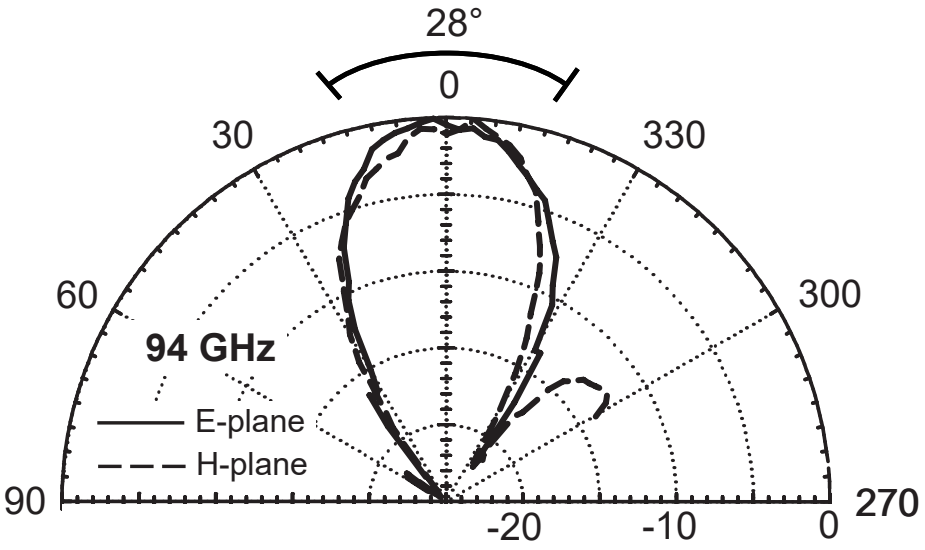
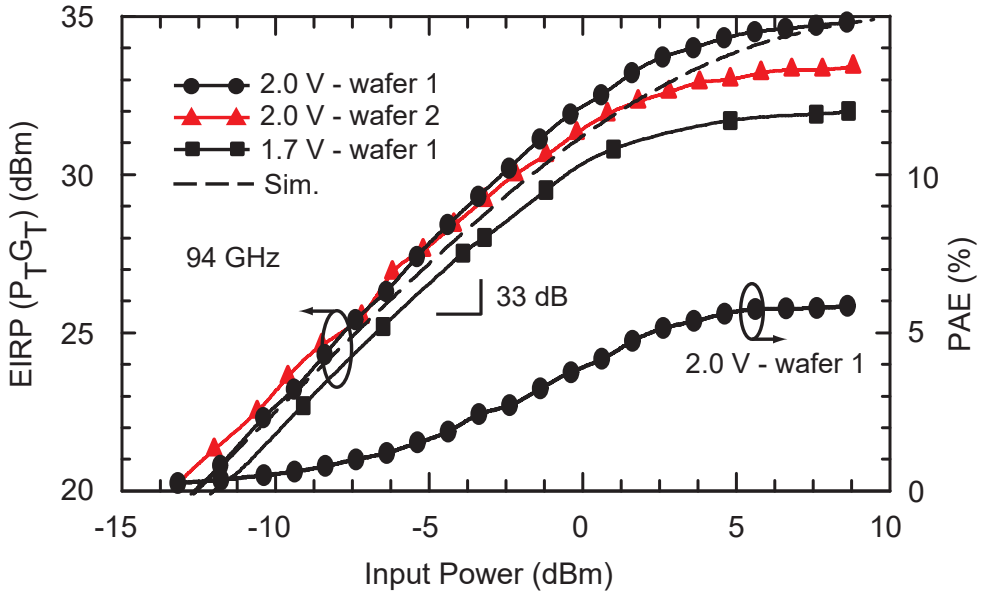


Figure 4.6: Measured E-plane and H-plane array patterns at 94 GHz with 28° 3dB-beamwidth.

Fig. 4.7a shows the measured EIRP of the antenna array versus array input power at 94 GHz over different PA collector voltages for two separate array assemblies. The input power P_{in} is the power into the SiGe chip. The slope of EIRP versus input power is the free-space gain of the system, plotted in Fig. 4.7b. The free-space gain is composed of the +15 dB gain input-driver amplifier, +15 dB gain PA element, +12dB gain antenna gain,

-9 dB Wilkinson divider and transmission-lines (combined losses). (The on-chip small-signal gain from array input to total-die-power output is 21 dB, which includes on-chip losses.)



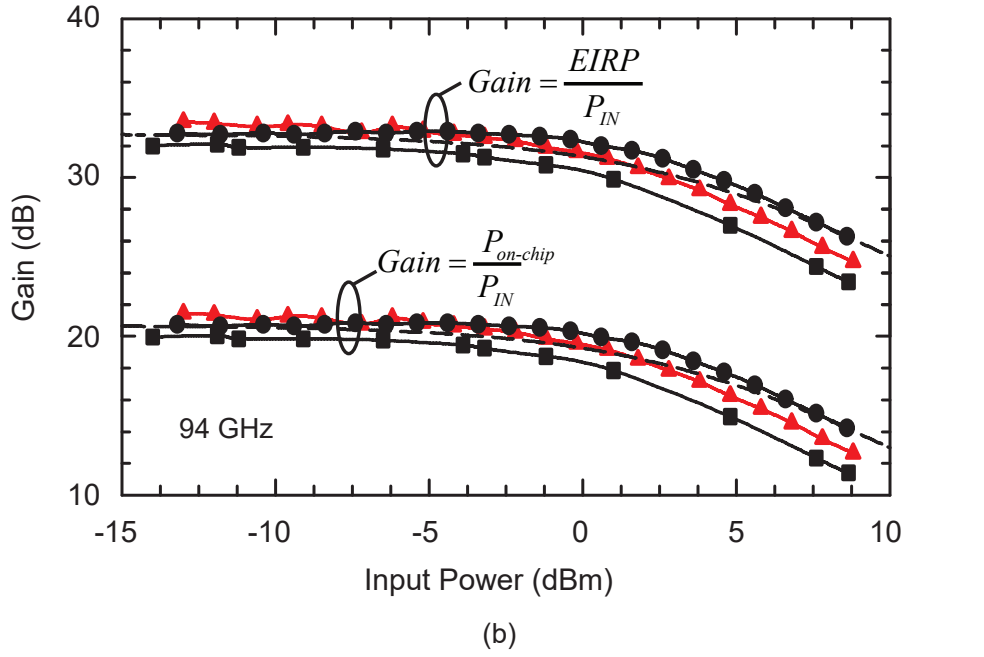


Figure 4.7: (a) Measured array EIRP versus input power at 94 GHz. (b) Measured quasi-optical array gain at 94 GHz for two different quartz-bonded (antenna) SiGe dies.

Fig. 4.8 plots the measured EIRP response for the PA array over frequency over different

PA collector voltages for two separate array assemblies. Measurements are limited by source power below 89 GHz and above 99 GHz.

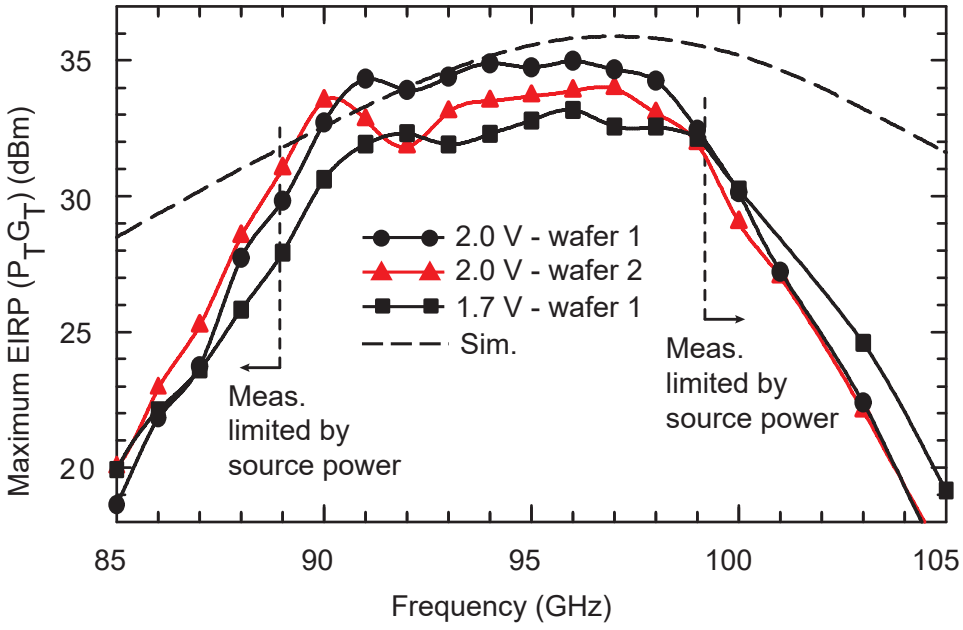


Figure 4.8: Measured array maximum EIRP versus frequency for two different quartz-bonded (antenna) SiGe dies. Measurements above 99 GHz and below 89 GHz are limited by the available source power in the measurement setup.

Table 4.1 summarizes the key measurements for the W-band 3 × 3 PA array (with antennas). This work demonstrates the first wafer-scale power amplifier array in 0.12 μm SiGe BiCMOS technology. The PA array operates at 2.0 V with 1.6 A currents, and the output power from the 9 PA elements on the SiGe die is 21–23 dBm at 90–98 GHz. The 21–23 dBm power is fed into the bonded microstrip antenna array. The radiated power from the antenna array is 17.5–19.5 dBm from 90–98 GHz. The effective isotropic radiated power (EIRP), which includes the antenna array-pattern gain, is 33–35 dBm (3.16 W) from 90–98 GHz. (EIRP measurements above 98 GHz are limited by the available source power of the measurement system.)

Table 4.2 compares the W-band PA array work with other quasi-optically combined arrays in different technologies and power-combining structures.

The integrated SiGe PA array is a highly scalable design, and integrated antennas eliminate significant losses from mmW transitions that would otherwise be necessary. The total output

[1] $P_{radiated}$
[2] EIRP is not applicable in waveguides. These entries are marked as NA in the table.
[3] 2-FETs per cell

Table 4.1: 3 × 3 wafer-scaled PA array measured results.

Parameter	**Measured Value**
EIRP	33–35 dBm @ 90–98 GHz
$P_{\text{on-chip}}$	21–23 dBm @ 90–98 GHz
P_{radiated}	17.5–19.5 dBm @ 90–98 GHz
S_{11}	-10 dB @ 84–105 GHz
Bias Conditions	1.66 A @ 2 V
$PAE_{\text{on-chip}}$	5.0–5.8%
PAE_{radiated}	2.25–2.6%
Power Combined Efficiency	100% (45% with antenna loss)

Table 4.2: Comparison of quasi-optically combined arrays

Authors	**Freq (GHz)**	**Tech**	**Power Comb. Method**	**#Elements**	**P_T (mW)**	**EIRP (W)**
This Work [9]	90–98	SiGe	Antenna/RFIC	9 (3 × 3)	90[1]	3.16
Hacker [177]	93	InGaAs	Grid/Waveguide	196 (14 × 14)	684	NA[2]
Chung [178]	79	InP	Grid/Waveguide	64 (8 × 8)	264	NA
Gouker [179]	44	GaAs	Antenna/Hybrid	256 (16 × 16)	5,900	11,570
Liu [180]	40	GaAs	Grid/Space	36 (6 × 6)	670	NR
DeLisio [181]	30	GaAs	Grid/Waveguide	242 (11 × 11)[3]	23,000	NA
Hubert [182]	29	GaAs	Antenna/Hybrid	36 (6 × 6)	8,000	NR

power scales directly with the number of elements N while EIRP scales with N^2. The design is scalable in silicon and allows for very large wafer-scaled designs greater than 200 mm in diameter. Increasing the number of elements beyond 64 results in watt-level on-chip mmW powers at 94 GHz that is competitive with traditional III-V solutions at much lower costs. Integrating additional amplitude and phase control on each PA element results in a transmit phased array capable of beam steering with adjustable sidelobe levels.

4.2 Future Work

This section discusses research opportunities for future work in mmW power amplifiers within the same SiGe 8HP 0.12 μm foundry technology.

Some immediate items can be addressed with access to necessary measurement equipment or additional wafer samples of the PA designs.

- Measure the PAs fully differentially with differential-input drive and differential output load.
- Measure small signal IP3 at W-band as well as large-signal linearity with a digitally-modulated wideband waveform with significant AM envelopes.
- Fully characterize the performance of the PAs over temperature.
- With more available source power at W-band, fully compress PA designs to obtain P_{1dB} and P_{sat} at higher frequencies.
- Characterize the RF performance shifts over multiple wafer samples to detail the tunability of the PA over process via the bias-tuning provisions.
- Obtain more refined custom models of the HBT thermal coupling with more parts from different wafers to expand the sample size.
- PAE can be further improved by reducing the quiescent bias currents in the bias circuits, particularly those of the first two stages. Transistor widths of the Q1 and Q2 input stages can also be reduced, which leads to reduced quiescent bias currents while still biased near peak-ft current densities. Gain and bandwidth can be recovered, as well, from reduced impedance transformation in the matching networks because of higher impedances presented by smaller HBTs.
- The experimental W5 and W6 designs with MOM capacitors above 100 GHz need to be re-tuned to recover gain and efficiency. All quarter-wavelength resonators should be

shortened to raise the center frequency. The self-resonant 440 fF MIM capacitor should be reduced in value to re-center the design more.

Several opportunities for design optimization still exist for the PA designs to increase output powers and efficiencies within the same technology, GF SiGe 8HP, and without complete re-design of the circuits. At W-band frequencies, at which the PA designs are significantly limited by available gain, general improvements in RF performance can be made by increasing the gain of PA stage, which can be achieved by decreasing passive losses and by optimizing the HBT matching.

- As demonstrated, the W-band PA design approach in common-emitter configuration is useful for generating significant linear mmW output powers up to a frequency of approximately 0.45 $\times$ f_T, after which HBT gain limitations prevail. Beyond this frequency, the effective gain of a common-emitter stage nears just 3 dB/stage. To push the linear common-emitter design to even higher frequencies, losses must be reduced— gain recovered— or designs need to move to more advanced technologies with higher transistor gains at 90 GHz.

- The Ka-band can be significantly improved by applying new dense 2D, overlapping-DT unit-cell packing in the aggregated power-cells at Ka-band. Significant gain can be recovered from wiring parasitics, leading to improved efficiencies and PAE. Unlike W-band densely-packed cells, the very large Ka-band power-cell transistors (e.g., 90 μm) may require re-visiting thermal-stability from mutual thermal coupling and emitter ballasting resistances.

- The latest HiCUM models are much improved over the early VBIC HBT models, allowing further design optimizations at higher collector voltages and currents. Rather than being limited to 1.7 V design optimizations, the PA can be tuned in simulations at 2.3 V, which can provide improvements in PAE and output powers.

- Additional gain can be recovered (around 0.6 dB) by creating even lower loss, higher Q inverted M_{234} base-feed transmission lines. By decreasing the losses of the characteristic base-feed transmission-line, the input impedance of the quarter-wavelength resonator is higher, which decreases the loss associated with using the resonator, especially at the base of Q1, which has the highest node impedance at 20 Ω. Overall bandwidth of the PA will improve, as well, since a lower loss base-insertion feed will result in larger impedances for wider bandwidths.

  There are two primary means of reducing the M_{234} losses: (1) much wider lines and (2) M1 RF ground shield from the lossy substrate. Using wider lines requires allowing

the structure to be heavily cheesed (up to 50% local density), and the characteristic impedance of the line drops, as well, which goes against achieve higher resonator impedances, but there is an optimization possible between the parameters. Adding an M1 (lowest metal) ground shield will also drop the characteristic impedance of the line, but reduce the losses caused by inducing eddy currents in the lossy substrate. The complication is that M2 sits only 0.35 μm from M1, so M2 would likely need to be dropped from the base-feed stack in favor of a wider M_{34} transmission-line structure.

- The MIM capacitor Q = 7–10 is a severe limitation for W-band PA designs, and it is the limiting component for resonator losses and matching-network losses. Using MOM capacitors exclusively can boost Q to 20–30 at 90 GHz, but layouts must be re-worked to account for devices that may be 10× larger in area. If more elements are added to the matching-network in order to lower Q_{MN} for even wider bandwidth response, MIM capacitor values tend to increase, which becomes problematic for MOM capacitors and their very low 0.1 fF/μm^2 capacitance density.

  Vertical-natural capacitors (VNCAP), are available in more advanced processes with more complex metal stack-ups in the BEOL, offer an excellent solution. They are densely, closely-separated, interdigitated fingers made from stacked metals that can provide up to 1.2 fF/μm^2 of capacitance, which is as much as the 8HP MIM capacitor. The VNCAP is high Q like the MOM, but much higher in capacitance density. Although still not as compact as a MIM, the VNCAP is much more practical than the single-layer MOM, and enables W-band designs to reduce losses significantly. The VNCAP is not currently available in 8HP, but future BEOL changes in 8HP could enable the much-needed device.

- With higher MIM Q components, available either by future process improvements of the MIM or availability of the vertical-natural capacitor above, matching networks can be made more wideband for even PA higher bandwidths. The current W-band PAs have small-signal & large-signal 3dB-bandwidths around 20% while the Ka-band designs have 3dB-bandwidths from 40–50%. The current designs use only 2-element matching networks to avoid losses from the MIM capacitor, but improved MIM Q enables more complex networks without penalties in matching-network transducer losses.

- New dual-MIM capacitors in the 8HP back-end-of-line process improvements allows for 3× the capacitance density (3.1 fF/μm^2 of the single-MIM capacitor used in the PA designs. The higher capacitance density allows for even larger capacitance bypass on chip, making the wideband bypass array even more effective with very large capacitors

within the structure.

- 2nd harmonic terminations, currently $< 2\ \Omega$ at W-band, can be made even lower for greater harmonic rejection. Lower impedance terminations at $2f_0$ can be achieved by using even wider collector lines, beyond what is required for electromigration compliance, and additional MIM capacitors in the wideband bypass-capacitor array. The optimizations would allow for improved PA efficiencies with gain compression in Class AB operation.

- The differential PA can be made truly differential with common-mode rejection characteristics by decoupling the AC and DC grounding at the plane of symmetry, as discussed in Section 2.4.2. Creating common-mode nodes with only virtual AC grounds or common-mode nodes with emitter-degeneration inductances achieves true common-mode noise rejection, and even-ordered distortions are cancelled in the differential operation, improving overall linearity.

- The area consumption can be reduced by folding and bending the top-level AM transmission lines to be more area efficient. The new $3\times$ density dual-MIM capacitors in 8HP shrink the area of the 16 wideband capacitor bypass arrays (or 8 in the half-PA). The shared large ground area between the two PA-halves is currently limited by the RF input/output pads and external probe configuration. This large ground area can be shrunk considerably in transmitter-integrated designs.

- Side shielding can be implemented on the base-feed quarter-wavelength transmission lines (M_{234}) so that the layout can be even more compacted. A shielded inverted tub structure can be formed by adding grounded M2, M3, and M4 stack that connects to the MQ RF ground plane and runs along the M_{234} transmission line on each side, much like the top-level AM transmission lines, but underneath the RF ground plane.

- New advances in the back-end-of-line processing has made through-silicon vias (TSVs) available to the 8HP process recently. TSVs can help reduce local grounding parasitics and improve thermal wicking of heat near the power-cell transistors. TSVs may be particularly helpful in converting the pseudo-differential to a truly differential one with common-mode rejection, as discussed in an earlier point.

- ESD diode clamp can be added to the DC bias pads for full ESD compliance. The shunt shorted transmission-line stubs at the immediate inputs and output of the PA make the RF I/O ESD compliant already, and the rest of the DC pads can be brought into compliance easily without affecting RF performance.

- Linearity is an important metric to characterize and measure since large-signal linearity simulations of high AM-envelope signals in PA compression (high output powers) are not very accurate due to modeling limitations of the HBT at voltage and current extremes, especially at 90 GHz. Optimizing a PA and any supporting system-level linearization structures for linearity performance is especially difficult.

  Measuring the large-signal linearity (not just small-signal two-tone linearity) aids in understanding all tradeoffs with gain, PAE, output power, and compression-point characteristics. Large-signal linearity based on real digitally-modulated waveforms, in particular, is not captured well by the HBT models, so measurements are necessary, which is difficult from an instrumentation and measurement-technique standpoint at W-band.

- New electro-thermal co-simulators and models are being developed for silicon processes. As these become available and mature, optimal custom layouts for aggregated transistors can be done in the simulator along with EM simulations. Electro-thermal-EM co-simulations would provide many HBT power-cell optimization opportunities without having to iterate multiple fab turns for characterization, which is non-trivial in terms of lab measurements.

- Foundry-supplied EM substrates, along with advanced simulators and ever-increasing compute capabilities, are now able to include complex process tolerances into EM simulations. Previously, only pre-computed compact SPICE models of scalable standard cells had this capability in passive devices. Having physical tolerances modeled in an EM substrate file (metal thicknesses, dielectric thicknesses, sheet resistances, bulk resistivities, dielectric losses, etc.) allows EM designs at mmW frequencies to simulate exactly how process tolerances will affect custom EM structures. Although the advanced capabilities are currently not available in 8HP, the general trend in new analog/RF/mmW processes is to include such EM-substrate modeling in the design kits.

- CAD compute power has steadily increased, and EM simulations that used to take a week now take minutes. The increased compute and RAM capabilities enable finite-element method (FEM) simulator to be used more often to model structures that significant 3D aspects to them (e.g., stacked-metal transistor-grounding structures and MOM capacitor leads). Planar method-of-moments simulators have difficult accurately modeling structures with vertical distributed-current components.

- With all possible improvements in the PA design at W-band, additional gain can be

recovered (at least 1 dB). This allows sizing the output transistor larger for even more maximum output powers from a single PA element.

- Finally, once optimizations in the current design approach are fully realized, new topologies such as cascode output stages and stacked transistors offer opportunities to push output powers another > 0.5 dB higher. Many of the design approaches and techniques presented in this work can also be applied to other topologies and different silicon technologies across a wide frequency range.

Advanced processes bring the benefits of improved passive components: for example, more complex metal stackup with more layers (up to 15 currently) and higher Q capacitors. Additionally, fabrication technology improvements result in higher tolerances for all structures devices, which is critical for practical mmW designs. All models, passive and active, tend to improve with process-technology advancements. Finally, the cost of developing larger circuits in better technologies always goes down with time.

As technology continues to improve and push higher f_T and f_{MAX} devices, transistor breakdown voltages BV_{CEO} will continue to decrease, as well, so the approaches presented in this work can be applied as SiGe technology advances and even higher frequencies & output powers are sought.

Appendices

Appendix A

Additional Equations

Harmonic Analysis for Reduced Conduction Angle Operation

When the quiescent current of a device is biased below Class A operation, the conduction angle is reduced and harmonic components are generated. Equations describing collector currents and their harmonic content over different conduction angles and amplifier classes of operation can be written. The harmonic content of the current waveform can be derived as follows [63].

Let α be the conduction angle and I_{max} be the peak current. Assuming the peak current is maintained at I_{max} as quiescent current is reduced, the DC component of the current is given by

$$I_{DC} = \frac{1}{2\pi} \int_{-\alpha/2}^{\alpha/2} \frac{I_{max}}{1 - \cos(\alpha/2)} \left(\cos(\theta) - \cos(\alpha/2)\right) d\theta \tag{A.1}$$

and the nth harmonic of the current is given by

$$I_n = \frac{1}{\pi} \int_{-\alpha/2}^{\alpha/2} \frac{I_{max}}{1 - \cos(\alpha/2)} \left(\cos(\theta) - \cos(\alpha/2)\right) \cos(n\theta) d\theta. \tag{A.2}$$

Simplifying the expressions above for the DC and fundamental components of the current,

$$I_{DC} = \frac{I_{max}}{2\pi} \cdot \frac{2\sin(\alpha/2) - \alpha\cos(\alpha/2)}{1 - \cos(\alpha/2)} \tag{A.3}$$

$$I_1 = \frac{I_{max}}{2\pi} \cdot \frac{\alpha - \sin(\alpha)}{1 - \cos(\alpha/2)}. \tag{A.4}$$

MSG & Rollett Stability Factor

Rollett stability factor based on transistor two-port S-parameter data [104]:

$$K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}S_{21}|} \tag{A.5}$$

$$\Delta = S_{11}S_{22} - S_{12}S_{21} \tag{A.6}$$

Maximum stable gain (MSG), defined when $K < 1$:

$$\text{MSG (dB)} = 20\log_{10}\left(10\frac{|S_{21}|}{|S_{12}|}\right) \tag{A.7}$$

Noise Figure

Noise figure is given by

$$NF = 10\log_{10} F. \tag{A.8}$$

Using the gain method, the system noise figure can be calculated as follows:

$$F_{sys} = \frac{\text{Total Output Noise Power}}{\text{Output Noise Power Due to Input Source Only}} \tag{A.9}$$

$$= \frac{P_{cold}}{2kT \cdot BW \cdot Gain_{sys}} \tag{A.10}$$

The denominator is multiplied by 2 because both the upper and lower sideband of input source noise is downconverted into the measured IF band. $kT(\text{in dB}) = -174dBm/Hz$.

Using the Y factor method, the system noise figure can be calculated as follows:

$$F_{sys} = \frac{ENR}{(\frac{P_{hot}}{P_{cold}} - 1)} \tag{A.11}$$

DUT noise figure can be de-embedded for a cascaded system, A - DUT - B, as follows:

$$F_{sys} = F_A + \frac{F_{DUT} - 1}{Gain_A} + \frac{F_B - 1}{Gain_A Gain_{DUT}} \tag{A.12}$$

$$F_{DUT} = Gain_A(F_{sys} - F_A) - \frac{F_B - 1}{Gain_{DUT}} + 1 \tag{A.13}$$

Appendix B

2-Element Matching: Complex Load to Complex Source

Fig. illustrates the configuration for presenting a matched load to a complex source impedance Z_1 with a 2-element L-matching network consisting only of a reactive series impedance Z_{series} and a reactive parallel impedance Z_{parallel}. The 2-element matching network transforms the complex load impedance Z_2 to $Z_{IN} = Z_1^*$.

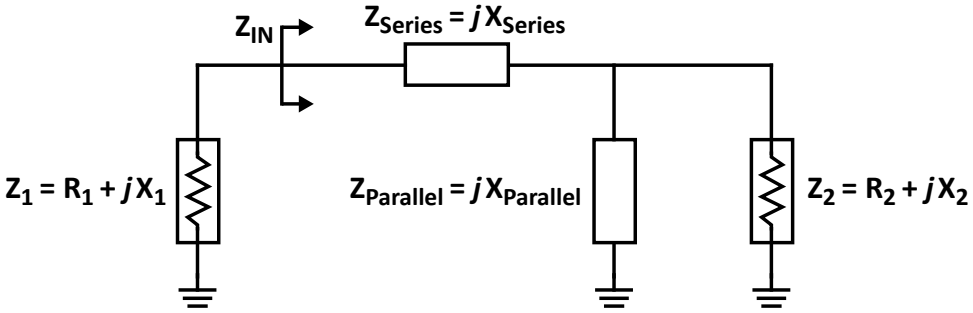


Figure B.1: Two-element L-matching network with complex-load and complex-source impedances.

Expressing the complex impedances with their resistive (R) and reactive (X) components [111],

$$Z_1 = R_1 + jX_1 \tag{B.1}$$

$$Z_2 = R_2 + jX_2 \tag{B.2}$$

$$Z_{\text{series}} = jX_{\text{series}} \tag{B.3}$$

$$Z_{\text{parallel}} = jX_{\text{parallel}}. \tag{B.4}$$

The input impedance presented to the source is given by $Z_{\text{series}} + (Z_{\text{parallel}} \parallel Z_2)$ or

$$Z_{IN} = Z_{\text{series}} + \frac{Z_{\text{parallel}} Z_2}{Z_{\text{parallel}} + Z_2}. \tag{B.5}$$

By setting $Z_{IN} = Z_1^*$, the series and parallel impedances can be calculated as follows:

$$X_{\text{series}} = -\left(X_1 \pm R_1 Q\right) \tag{B.6}$$

$$X_{\text{parallel}} = \frac{X_2 \pm R_2 Q}{\frac{R_2}{R_1} - 1} \tag{B.7}$$

Only one unique solution of X_{series} and X_{parallel} will provide the desired configuration. The loaded quality factor Q_{loaded} for the resulting matching network is given by

$$Q_{\text{loaded}} = \sqrt{\frac{R_2}{R_1} - 1 + \frac{X_2^2}{R_1 R_2}}. \tag{B.8}$$

Taking the specific configuration used in the Ka-band and W-band power amplifiers, all the matching networks use a series capacitor and a parallel inductor (shorted-circuit transmission-line stub). Using a capacitor for the series component and an inductor for the parallel component, the required capacitance and inductance values are now given by

$$C = \frac{1}{2\pi f_0 X_{\text{series}}} \tag{B.9}$$

$$L = \frac{X_{\text{parallel}}}{2\pi f_0}. \tag{B.10}$$

For the input and output matching networks, the load Z_2 is purely resistive (50 Ω). If $Z_2 = 50\ \Omega$, then $R_2 = 50\ \Omega$ and $X_2 = 0$. In this case, the expressions for Q_{loaded} and X_{parallel} simplify, with the assumption $R_2 > R_1$, to the familiar

$$Q_{\text{loaded}} = \sqrt{\frac{R_2}{R_1} - 1} \tag{B.11}$$

$$X_{\text{parallel}} = \frac{R_2}{Q_{\text{loaded}}}. \tag{B.12}$$

X_{series} follows from the relationships

$$Q_{\text{loaded}} = Q_{\text{l,series}} = Q_{\text{l,parallel}} = \sqrt{\frac{R_{\text{parallel}}}{R_{\text{series}}} - 1} \tag{B.13}$$

where the loaded series $Q_{\text{l,series}}$ is defined look into X_{series} & R_1 in series; and the loaded parallel $Q_{\text{l,parallel}}$ is defined as looking at only R_2 & X_{parallel} in parallel.

$$Q_{\text{l,series}} = \frac{X_{\text{series}}}{R_{\text{series}}} = \frac{X_{\text{series}}}{R_1} \tag{B.14}$$

$$Q_{\mathrm{l,parallel}} = \frac{R_{\mathrm{parallel}}}{X_{\mathrm{parallel}}} = \frac{R_2}{X_{\mathrm{parallel}}} \tag{B.15}$$

Finally, X_{series} is

$$X_{\mathrm{series}} = Q_{\mathrm{loaded}} R_{\mathrm{series}} = R_{\mathrm{series}} \sqrt{\frac{R_{\mathrm{parallel}}}{R_{\mathrm{series}}} - 1} \tag{B.16}$$

$$X_{\mathrm{series}} = R_1 \sqrt{\frac{R_2}{R_1} - 1} \tag{B.17}$$

Appendix C

VBIC vs. HiCUM HBT Models

Fig. C.1 compares the DC-IV curves between the older VBIC HBT model and the newer HiCUM HBT model for the 5 µm unit-cell device. Identical biasing conditions ranges are applied for both models with the same forced base-current range and I_B steps. The VBIC model predicts higher currents than the HiCUM model.

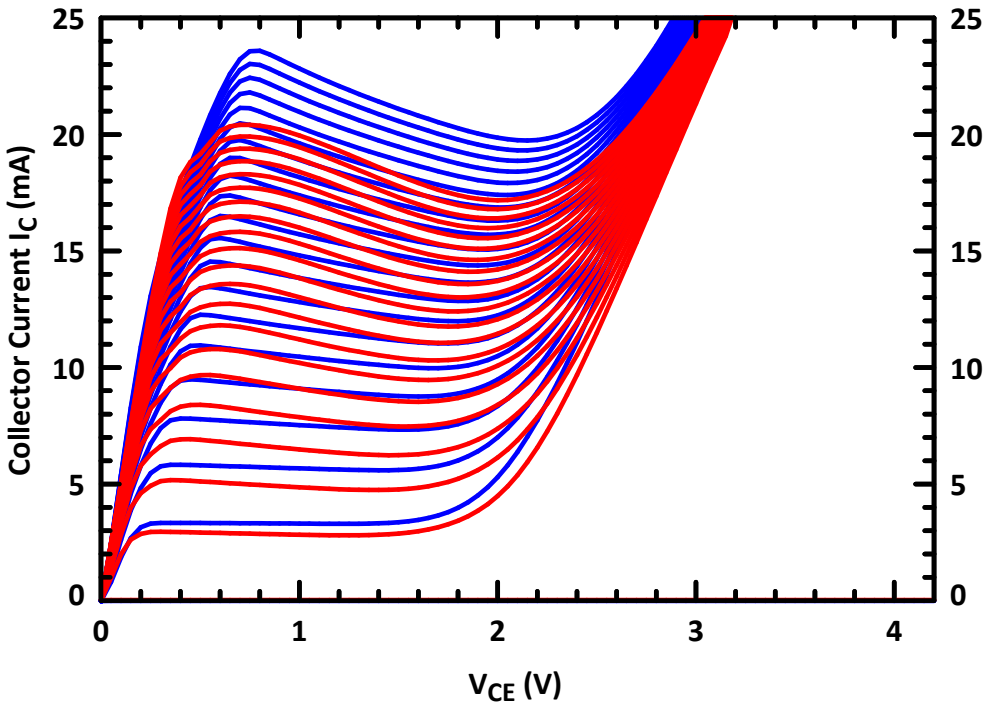


Figure C.1: DC-IV comparison for 5 µm npn HBT with newer HiCUM model (red) and older VBIC model (blue). The HBT is biased with an ideal current source with the same range and current-steps for each model.

Fig. C.2 illustrates the maximum available gain (MAG) for the implemented aggregated 40 µm power-cell HBT at 90 GHz. The older VBIC transistor model simulates 0.6 dB more gain near peak-f_T collector currents of 60 µm. The HiCUM model should be more reflect

hardware more accurately at higher current densities. VBIC predicts a higher maximum available gain (MAG) than the HiCUM model. The roll-off of MAG from the Kirk effect is much steeper with increasing current in the HiCUM model.

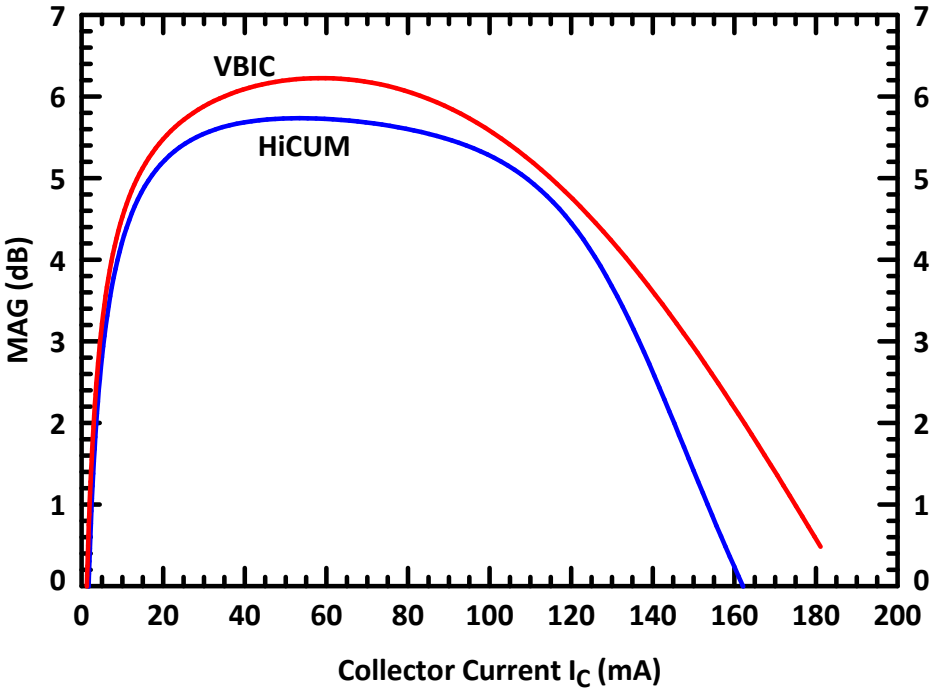


Figure C.2: Q3 40 µm maximum available gain (MAG) @ 90 GHz vs. collector current I_C for V_{CC} = 1.7 V: HiCUM (blue) and VBIC (red) transistor models.

In general, the VBIC and HiCUM modeling is the most accurate for smaller devices around 5.0 µm. The model parameters are likely extracted, fit, and optimized to smaller transistors and scaled up for larger ones. Foundry measured-versus-modeled data shows that the original VBIC model for SiGe npn HBTs is most accurate for 3–6 µm emitter lengths in the peak-f_T npn HBT devices [77]. Beyond 6 µm, the modeling accuracy of both DC characteristics and f_T deteriorates, impacting IV curves and RF parameters such as gain and output matching. The measured foundry data also shows that both the VBIC and HiCUM small-signal S-parameter models suffer reduced accuracy above 60 GHz.

Appendix D

Common Emitter vs. Cascode Topology

This section compares the chosen common-emitter topology with the cascode-amplifier topology along with challenges, advantages, and disadvantages of each configuration for mmW PAs. Although the cascode configuration is not used in any of the presented PA designs, understanding the benefits and limitations of the circuit highlights the some key benefits of the chosen common-emitter topology.

Fig. D.1 depicts a cascode configuration with ideal inductive biasing. The cascode consists of a CE stage followed by a common-base (CB) stage. When comparing cascode to CE, the cascode topology should be considered a two-stage amplifier with coupled currents.

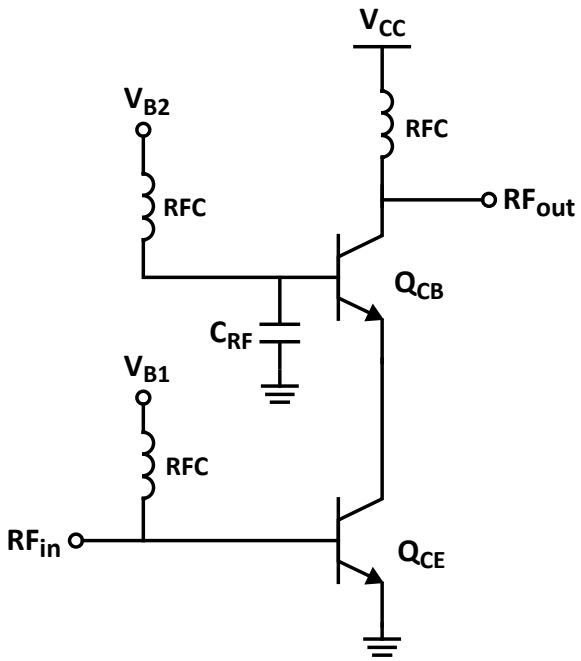


Figure D.1: Cascode topology (CE + CB) with ideal inductive, RF-choke (RFC) biasing.

A cascode amplifier acts as a two-stage amplifier: common-emitter (CE) to common-base (CB) amplifier. The input common-emitter stage Q_{CE} provides the current gain and low

voltage gain while the following common-base stage Q_{CB} provides the voltage gain and low current gain. Voltage gain of the bottom HBT is low, which reduces the effective Miller capacitance, which is the voltage gain multiplied by the collector-base capacitance ($A_v \times C_{CB}$) seen at the input of the CE transistor. Current gain is set by the CE stage while the voltage gain depends on the resistance presented to the collector of the CB output. The base of the CB stage is AC-grounded, ideally, so the collector voltage and emitter voltage of the CB transistor are held nearly constant. Thus, the CB stage presents a low-impedance load to the collector of the CE stage.

The cascode offers higher collector voltage operation and higher output-collector breakdown voltages as well as the possibility for higher optimal output impedances. The cascode can also offer additional bandwidth improvements since the Miller effect on the collector-base capacitance of the input common-emitter transistor is greatly reduced because the voltage gain is only ~ 1 V/V. The upper, common-base device— the cascode recovers the gain by being primarily a voltage-gain stage.

The cascode configuration can, ideally, offer higher gains at 90 GHz compared to two common-emitter (CE) stages, but RF/AC grounding challenges, which are extremely layout dependent, will reduce the achievable gains in W-band. Base inductive & resistive degeneration in the cascode transistor and additional emitter degeneration in the lower cascode transistor will impact the achievable gains since grounding this terminal is non-trivial within W-band. The cascode stage can be more difficult to stabilize on-chip at low frequencies since gain can be extremely high and associated biasing networks are more complicated than a common-emitter stage.

The CB base node suffers from base degeneration. Ideally this node is a perfect AC ground at 90 GHz. Any wiring inductance or resistance adds to base degeneration. At 90 GHz, the MIM Q is only 7–10, resulting in unwanted base degeneration. Moreover, capacitors cannot be so large (> 500 fF) that they operate past resonance and add additional degeneration inductances.

In the cascode, both the CB output collector metal interconnect and the CB emitter metal interconnect, which connects to the CE collector, must be electromigration compliant. This necessitates wiring to top-level AM metal for both CB emitter and collector. The common-emitter configuration faces the same constraints with the emitter and collector, but the emitter can be simply grounded to a ground bus that pulls over the transistor array and connects to large global DC/RF grounds surrounding the HBTs. The common-emitter base node is pull out using lower level M2, M3, and M4 copper lines. In the common-emitter topology, these base feedlines do not need to be low RF impedance as in the case of the cascode CB base wiring. Whatever means of AC-grounding the CB base must deal with

the additional resistance and inductance of the base lines that subsequently connect to an AC-shorting device such as a MIM capacitor.

The common-emitter configuration does offer some key benefits in its simplicity

- In contrast to the cascode, the common emitter (CE) is the simplest transistor amplifier configuration to bias and ground. The emitter is both AC and DC-grounded to the global chip ground without additional bypass capacitor elements. This allows simple, low-inductance grounding schemes to minimize inductive & resistive degeneration. Compared to common base (CB) and cascode topologies, common-emitter has no additional AC grounding required at any of the other terminals. Additional AC-grounding in the CB and cascode topology is necessary at the base with RF bypass capacitors while a CE ground is a true DC & AC ground. At 90 GHz, low impedance RF shorts are difficult to achieve, and there are fewer parasitics associated with the simple CE grounding scheme.

- Having independent current & voltage bias fine-tuning provisions in the common emitter stages is more flexible than tuning the coupled & stacked common-emitter and common-base stages in the cascode, which share both current and voltage biasing for both transistors. Comparing a two-stage common emitter to a cascode, the common-emitter configuration allows more biasing freedom to tune and adjust the DC biasing points in order to tune the RF responses of the two stages indepdently.

- BV_{CBO} is the ultimate collector-to-base breakdown voltage limitation in a single high-performance npn HBT in any circuit topology, common emitter or common base. By having extremely low external base resistances, the common-emitter operational breakdown voltage BV_{CER} can still approach the common-base (and thus cascode) voltage and power limits of operating at $BV_{\mathrm{CBO}} = 6.0$ V [183].

- The common-base (CB) stage is not as linear, inherently, as common emitter (CE) from an intermodulation (IM) standpoint. Under power matching conditions, the CE is more linear than cascode and CB [184, 185]. In the common-base transistor, nonlinear collector-base capacitance dominates the nonlinearity at high current densities while avalanche multiplication dominates at low currents [186]. Since the common-emitter topology does not contain a common base stage, the CE results in much higher linearity, e.g., as compared in [187] for an envelope-tracking transmitter architecture for WiMAX polar transmitters at 2.4 GHz.

In W-band simulations, the cascode topology does not offer appreciable advantages in maximizing output power in simulations after non-idealities are considered— base degeneration

from AC-grounding limitations on the cascode CB-transistor base and inter-cell wiring parasitics. The maximum output power is approximately 0.5 dB higher, but the design is considerably more difficult to tune on the measurement bench since the voltages and currents of the CE and CB transistors are tightly coupled. The design approaches in the presented common-emitter PA are still applicable to cascode configurations, however, and both common-emitter and cascode are valid, proven configurations in mmW power amplifiers with different merits and challenges.

Appendix E

Metal Density Limitations

GF 8HP (and many other advanced silicon processes) features copper metal interconnects, and the copper IC fabrication process requires local-density requirements for process uniformity within the copper layers. The 8HP process requires mandatory auto-cheesing and auto-filling of copper layers M1, M2, M3, and M4 on mmW design. Cheesing and filling are required to ensure planarity of the copper metal layers in the copper damascene process. 8HP designers have no control over the automated process, which presents challenges in modeling mmW structures. At best, structures can request reduced filling or cheesing under or around critical structures, but cannot eliminate the fill or cheese altogether. The cheese & fill impact the accuracy of simulated EM models, in particular. Informal estimates by the foundry approximated that fill & cheese would affect EM modeling with shifts around a 3–4% percent above 30 GHz.

All copper layers within 8HP receive auto-generated (by the foundry) copper fill and holes (cheesing) in order to create uniform copper pattern densities in each layer for proper planarization during the copper formation and etching processes. The 5 lowest-level metals (closest to the silicon substrate) are made from copper: layers M1, M2, M3, M4, and MQ. These layers have automatic fill generated to meet minimum density requirements for planarity in the copper damascene process.

The mandatory cheesing– local-density reduction of copper metals with holes– of wide copper metals reduces the effective linewidth by 30% (for wide lines under 50 μm and 52% for lines over 50 μm. With special requests, the fill can be reduced down to 8% around critical structures like inductors, but the fill cannot be eliminated completely. The aluminum metals (AM top-layer and LY below it) are not subject to mandatory cheesing and are more easily modeled in EM simulations.

The foundry also adds fill to ensure the local density of a copper metal layer is $> 10\%$ or creates holes (cheesing) to make local density $< 85\%$. This process takes place within all 50 μm $\times$ 50 μm area windows across the chip. Designers must account for a possible additional margin for the fill & cheese process. Empty areas are filled up to 15% minimum

local density. Large copper planes are cheesed down to 80% local copper densities. The RF ground plane (on MQ, the uppermost copper layer) and wide wiring over 50 μm are cheesed to make those areas 70% local copper densities.

Bits of floating copper metal fill within the oxide dielectric shorten the effective path of the electric field. The net effect is increased capacitance. A 20% fill factor (by area) reduces the effective dielectric thickness by ∼20%. The fill effect can modeled either as a decrease in effective dielectric thickness or an increase in effective dielectric constant between the layers.

During the design phase, incorporating the effects of auto-cheesing & filling on transmission line structures is not possible in EM simulations, and small shifts in the EM-modeled response are expected. The pattern fill and pattern cheese (holes) can shift transmission-line parameters by a few percent at mmW frequencies. The small shifts in modeling are not critical for these W-band PA designs because of the low-loaded-Q, wideband matching networks, which are more tolerant to process shifts and shift together because of applied unit-cell approaches in the passive mmW structures.

Bibliography

[1] H. Hashemi and S. Raman, *mm-Wave Silicon Power Amplifiers and Transmitters*. Cambridge, United Kingdom: Cambridge University Press, 2016.

[2] A. Joseph, D. Coolbaugh, M. Zierak, R. Wuthrich, P. Geiss, Z. He *et al.*, "A 0.18μm BiCMOS technology featuring 120/100 GHz ($f_t/f_{\max}$) HBT and ASIC-compatible CMOS using copper interconnect," in *Proc. IEEE Bipolar/BiCMOS Circuits Technol. Meeting (BCTM)*, Minneapolis, MN, Oct. 2001, pp. 143–146.

[3] A. Joseph, D. Coolbaugh, D. Harame, G. Freeman, S. Subbanna, M. Doherty *et al.*, "0.13 μm 210 GHz f_t SiGe HBTs—expanding the horizons of SiGe BiCMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, vol. 1, San Francisco, CA, Feb. 2002, pp. 138–140.

[4] D. C. Ahlgren, G. Freeman, S. Subbanna, R. Groves, D. Greenberg, J. Malinowski *et al.*, "A SiGe HBT BiCMOS technology for mixed signal RF applications," in *Proc. IEEE Bipolar/BiCMOS Circuits Technol. Meeting (BCTM)*, Minneapolis, MN, Sep. 1997, pp. 195–197.

[5] S. A. St. Onge, D. L. Harame, J. S. Dunn, S. Subbanna, D. C. Ahlgren, G. Freeman *et al.*, "A 0.24 μm SiGe BiCMOS mixed-signal RF production technology featuring a 47 GHz f_t HBT and 0.18 μm L_{eff} CMOS," in *Proc. IEEE Bipolar/BiCMOS Circuits Technol. Meeting (BCTM)*, Minneapolis, MN, Sep. 1999, pp. 117–120.

[6] M. Steer, *Microwave and RF Design: A Systems Approach*, 2nd ed. Edison, NJ: SciTech Publishing, 2013.

[7] M. Chang and G. M. Rebeiz, "A 26 to 40GHz wideband SiGe balanced power amplifier IC," in *IEEE Radio Freq. Integr. Circuits (RFIC) Symp. Dig.*, Honolulu, HI, Jun. 2007, pp. 729–732.

[8] ——, "A wideband high-efficiency 79–97 GHz SiGe linear power amplifier with >90 mW output," in *Proc. IEEE Bipolar/BiCMOS Circuits Technol. Meeting (BCTM)*, Monterey, CA, Oct. 2008, pp. 69–72.

[9] 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 Trans. Microw. Theory Techn.*, vol. 59, no. 4, pp. 954–965, Apr. 2011.

[10] S. Raman, T. H. Chang, I. Abdomerovic, C. L. Dohrman, C. Maxey, and M. J. Rosker, "The DARPA COSMOS and ELASTx programs: Towards next generation linearized microwave/mm-wave transmitters," in *Proc. IEEE Compound Semicond. Integr. Circuits Symp. (CSICS)*, Monterey, CA, Oct. 2010, pp. 1–4.

[11] I. Abdomerovic, W. D. Palmer, P. M. Watson, R. Worley, and S. Raman, "Leveraging integration: Toward efficient linearized all-silicon IC transmitters," *IEEE Microw. Mag.*, vol. 15, no. 3, pp. 86–96, May 2014.

[12] H. Chireix, "High power outphasing modulation," *Proc. Inst. of Radio Engineers*, vol. 23, no. 11, pp. 1370–1392, Nov. 1935.

[13] S. Reynolds, B. Floyd, U. Pfeiffer, and T. Zwick, "60GHz transceiver circuits in SiGe bipolar technology," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, vol. 1, San Francisco, CA, Feb. 2004, pp. 442–538.

[14] U. R. Pfeiffer, S. K. Reynolds, and B. A. Floyd, "A 77 GHz SiGe power amplifier for potential applications in automotive radar systems," in *IEEE Radio Freq. Integr. Circuits (RFIC) Symp. Dig.*, Fort Worth, TX, Jun. 2004, pp. 91–94.

[15] B. A. Floyd, S. K. Reynolds, U. R. Pfeiffer, T. Zwick, T. Beukema, and B. Gaucher, "SiGe bipolar transceiver circuits operating at 60 GHz," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 156–167, Jan. 2005.

[16] U. R. Pfeiffer, D. Goren, B. A. Floyd, and S. K. Reynolds, "SiGe transformer matched power amplifier for operation at millimeter-wave frequencies," in *Proc. Eur. Solid-State Circuits Conf. (ESSCIRC)*, Grenoble, France, Sep. 2005, pp. 141–144.

[17] U. R. Pfeiffer and D. Goren, "A 20 dBm fully-integrated 60 GHz SiGe power amplifier with automatic level control," *IEEE J. Solid-State Circuits*, vol. 42, no. 7, pp. 1455–1463, Jul. 2007.

[18] S. Glisic and C. Scheytt, "A 13.5-to-17 dBm P1dB, selective, high-gain power amplifier for 60 GHz applications in SiGe," in *Proc. IEEE Bipolar/BiCMOS Circuits Technol. Meeting (BCTM)*, Monterey, CA, Oct. 2008, pp. 65–68.

[19] U. R. Pfeiffer and D. Goren, "A 23-dBm 60-GHz distributed active transformer in a silicon process technology," *IEEE Trans. Microw. Theory Techn.*, vol. 55, no. 5, pp. 857–865, May 2007.

[20] T. S. D. Cheung and J. R. Long, "A 21–26-GHz SiGe bipolar power amplifier MMIC," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2583–2597, Dec. 2005.

[21] A. Komijani and A. Hajimiri, "A wideband 77-GHz, 17.5-dBm fully integrated power amplifier in silicon," *IEEE J. Solid-State Circuits*, vol. 41, no. 8, pp. 1749–1756, Aug. 2006.

[22] P. J. Riemer, J. S. Humble, J. F. Prairie, J. D. Coker, B. A. Randall, B. K. Gilbert, and E. S. Daniel, "Ka-band SiGe HBT power amplifier for single-chip T/R module applications," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Honolulu, HI, Jun. 2007, pp. 1071–1074.

[23] S. T. Nicolson, K. H. K. Yau, S. Pruvost, V. Danelon, P. Chevalier, P. Garcia, A. Chantre, B. Sautreuil, and S. P. Voinigescu, "A low-voltage SiGe BiCMOS 77-GHz automotive radar chipset," *IEEE Trans. Microw. Theory Techn.*, vol. 56, no. 5, pp. 1092–1104, May 2008.

[24] N. Kinayman, A. Jenkins, D. Helms, and I. Gresham, "Design of 24 GHz SiGe HBT balanced power amplifier for system-on-a-chip ultra-wideband applications," in *IEEE Radio Freq. Integr. Circuits (RFIC) Symp. Dig.*, Long Beach, CA, Jun. 2005, pp. 91–94.

[25] S. Chartier, E. Sonmez, and H. Schumacher, "24 and 36 GHz SiGe HBT power amplifiers," in *Proc. Topical Meeting Silicon Monolithic Integr. Circuits in RF Syst.*, Atlanta, GA, Sep. 2004, pp. 251–254.

[26] E. Afshari, H. Bhat, X. Li, and A. Hajimini, "Electrical funnel: A broadband signal combining method," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, Feb. 2006, pp. 751–760.

[27] J. P. Comeau, J. M. Andrews, and J. D. Cressler, "A monolithic 24 GHz, 20 dBm, 14% PAE SiGe HBT power amplifier," in *Proc. Eur. Microw. Conf.*, Manchester, United Kingdom, Sep. 2006, pp. 419–422.

[28] A. Komijani, A. Natarajan, and A. Hajimiri, "A 24-GHz, +14.5-dBm fully integrated power amplifier in 0.18-μm CMOS," *IEEE J. Solid-State Circuits*, vol. 40, no. 9, pp. 1901–1908, Sep. 2005.

[29] *International Technology Roadmap for Semiconductors 2.0*, 2015.

[30] 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 *Proc. Int. Electron Devices Meeting (IEDM)*, Washington, DC, Dec. 2007, pp. 255–258.

[31] S. Narasimha, P. Chang, C. Ortolland, D. Fried, E. Engbrecht, K. Nummy *et al.*, "22nm high-performance SOI technology featuring dual-embedded stressors, epi-plate high-K deep-trench embedded DRAM and self-aligned via 15LM BEOL," in *Proc. Int. Electron Devices Meeting (IEDM)*, San Francisco, CA, Dec. 2012, pp. 3.3.1–3.3.4.

[32] C. H. Lin, B. Greene, S. Narasimha, J. Cai, A. Bryant, C. Radens *et al.*, "High performance 14nm SOI FinFET CMOS technology with $0.0174\mu m^2$ embedded DRAM and 15 levels of Cu metallization," in *Proc. Int. Electron Devices Meeting (IEDM)*, San Francisco, CA, Dec. 2014, pp. 3.8.1–3.8.3.

[33] J. Kim, H. Dabag, P. Asbeck, and J. F. Buckwalter, "Q-band and W-band power amplifiers in 45-nm CMOS SOI," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 6, pp. 1870–1877, Jun. 2012.

[34] 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 J. Solid-State Circuits*, vol. 49, no. 5, pp. 1148–1157, May 2014.

[35] J. A. Jayamon, J. F. Buckwalter, and P. M. Asbeck, "Multigate-cell stacked FET design for millimeter-wave CMOS power amplifiers," *IEEE J. Solid-State Circuits*, vol. 51, no. 9, pp. 2027–2039, Sep. 2016.

[36] J. P. Raskin, "SOI technology pushes the limits of CMOS for RF applications," in *Proc. IEEE Topical Meeting Silicon Monolithic Integr. Circuits in RF Syst. (SiRF)*, Austin, TX, Jan. 2016, pp. 17–20.

[37] A. J. Joseph, D. L. Harame, B. Jagannathan, D. Coolbaugh, D. Ahlgren, J. Magerlein *et al.*, "Status and direction of communication technologies—SiGe BiCMOS and RFCMOS," *Proc. IEEE*, vol. 93, no. 9, pp. 1539–1558, Sep. 2005.

[38] R. L. Schmid, A. C. Ulusoy, S. Zeinolabedinzadeh, and J. D. Cressler, "A comparison of the degradation in RF performance due to device interconnects in advanced SiGe HBT and CMOS technologies," *IEEE Trans. Electron Devices*, vol. 62, no. 6, pp. 1803–1810, Jun. 2015.

[39] A. Mazzanti, L. Larcher, R. Brama, and F. Svelto, "Analysis of reliability and power efficiency in cascode class-E PAs," *IEEE J. Solid-State Circuits*, vol. 41, no. 5, pp. 1222–1229, May 2006.

[40] A. Chakrabarti and H. Krishnaswamy, "High-power high-efficiency class-E-like stacked mmWave PAs in SOI and bulk CMOS: Theory and implementation," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 8, pp. 1686–1704, Aug. 2014.

[41] C. Liu, Q. Li, Y. Li, X. D. Deng, H. Tang, R. Wang, H. Liu, and Y. Z. Xiong, "A Ka-band single-chip SiGe BiCMOS phased-array transmit/receive front-end," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 11, pp. 3667–3677, Nov. 2016.

[42] M. Fuqan, M. Jahn, and A. Stelzer, "A SiGe-based high-gain power amplifier for E-band communication systems," in *Proc. Eur. Microw. Integr. Circuit Conf. (EuMIC)*, Nuremberg, Germany, Oct. 2013, pp. 149–152.

[43] P. Y. Wu and J. F. Buckwalter, "A Q-band/W-band dual-band power amplifier in 0.12μm SiGe BiCMOS process," in *Proc. IEEE Compound Semicond. Integr. Circuits Symp. (CSICS)*, Oct. 2013, pp. 1–4.

[44] A. Sarkar and B. Floyd, "A 28-GHz class-J power amplifier with 18-dBm output power and 35% peak PAE in 120-nm SiGe BiCMOS," in *Proc. IEEE Topical Meeting Silicon Monolithic Integr. Circuits in RF Syst. (SiRF)*, Newport Beach, CA, Jan. 2014, pp. 71–73.

[45] S. Y. Mortazavi and K. J. Koh, "A 28-GHz inverse class-F power amplifier with coupled-inductor based harmonic impedance modulator," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, San Jose, CA, Sep. 2015, pp. 1–4.

[46] ——, "A class F^{-1}/F 24-to-31GHz power amplifier with 40.7% peak PAE, 15dBm OP1dB, and 50mW Psat in 0.13μm SiGe BiCMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, Feb. 2014, pp. 254–255.

[47] R. Levinger, R. B. Yishay, O. Katz, B. Sheinman, N. Mazor, R. Carmon, and D. Elad, "High-performance E-band transceiver chipset for point-to-point communication in SiGe BiCMOS technology," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 4, pp. 1078–1087, Apr. 2016.

[48] A. Y. K. Chen, Y. Baeyens, Y. K. Chen, and J. Lin, "An 83-GHz high-gain SiGe BiCMOS power amplifier using transmission-line current-combining technique," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 4, pp. 1557–1569, Apr. 2013.

[49] K. Datta and H. Hashemi, "A 29dBm 18.5% peak PAE mm-Wave digital power amplifier with dynamic load modulation," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, Feb. 2015, pp. 1–3.

[50] ——, "Performance limits, design and implementation of mm-Wave SiGe HBT class-E and stacked class-E power amplifiers," *IEEE J. Solid-State Circuits*, vol. 49, no. 10, pp. 2150–2171, Oct. 2014.

[51] H. C. Lin and G. M. Rebeiz, "A 70–80 GHz SiGe amplifier with peak output power of 27.3 dBm," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 7, pp. 2039–2049, Jul. 2016.

[52] H.-C. Lin and G. M. Rebeiz, "A 110–134-GHz SiGe amplifier with peak output power of 100–120 mW," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 12, pp. 2990–3000, Dec. 2014.

[53] S. Balasubramanian, S. Boumaiza, H. Sarbishaei, T. Quach, P. Orlando, J. Volakis, G. Creech, J. Wilson, and W. Khalil, "Ultimate transmission," *IEEE Microwave Magazine*, vol. 13, no. 1, pp. 64–82, Jan. 2012.

[54] *Attenuation by atmospheric gases*, International Telecommunication Union Radiocommunication Sector (ITU-R) Recommendation P.676-11, Sep. 2016.

[55] S. Zihir, O. D. Gurbuz, A. Karroy, S. Raman, and G. M. Rebeiz, "A 60 GHz single-chip 256-element wafer-scale phased array with EIRP of 45 dBm using sub-reticle stitching," in *IEEE Radio Freq. Integr. Circuits (RFIC) Symp. Dig.*, Phoenix, AZ, May 2015, pp. 23–26.

[56] A. M. Niknejad and H. Hashemi, *mm-Wave Silicon Technology: 60GHz and Beyond*. New York, NY: Springer, 2008.

[57] B. C. Kane, L. A. Geis, M. A. Wyatt, D. G. Copeland, and J. A. Mogensen, "Smart phased array SoCs: A novel application for advanced SiGe HBT BiCMOS technology," *Proc. IEEE*, vol. 93, no. 9, pp. 1656–1668, Sep. 2005.

[58] P. Song, A. C. Ulusoy, R. L. Schmid, S. N. Zeinolabedinzadeh, and J. D. Cressler, "W-band SiGe power amplifiers," in *Proc. IEEE Bipolar/BiCMOS Circuits Technol. Meeting (BCTM)*, Coronado, CA, Sep. 2014, pp. 151–154.

[59] P. Song, M. A. Oakley, A. C. Ulusoy, M. Kaynak, B. Tillack, G. A. Sadowy, and J. D. Cressler, "A class-E tuned W-band SiGe power amplifier with 40.4% power-added efficiency at 93 GHz," *IEEE Microw. Wireless Compon. Lett.*, vol. 25, no. 10, pp. 663–665, Oct. 2015.

[60] S. T. Nicolson, K. A. Tang, K. H. K. Yau, P. Chevalier, B. Sautreuil, and S. P. Voinigescu, "A low-voltage 77-GHz automotive radar chipset," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Honolulu, HI, Jun. 2007, pp. 487–490.

[61] B. Razavi, *RF Microelectronics*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2012.

[62] J. L. B. Walker, *Handbook of RF and Microwave Power Amplifiers*. Cambridge, United Kingdom: Cambridge University Press, 2012.

[63] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*, 2nd ed. Norwood, MA: Artech House, 2006.

[64] D. M. Snider, "A theoretical analysis and experimental confirmation of the optimally loaded and overdriven RF power amplifier," *IEEE Trans. Electron Devices*, vol. 14, no. 12, pp. 851–857, Dec. 1967.

[65] F. H. Raab, P. Asbeck, S. Cripps, P. B. Kenington, Z. B. Popovic, N. Pothecary, J. F. Sevic, and N. O. Sokal, "Power amplifiers and transmitters for RF and microwave," *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 3, pp. 814–826, Mar. 2002.

[66] J.-S. Rieh, B. Jagannathan, D. R. Greenberg, M. Meghelli, A. Rylyakov, F. Guarin *et al.*, "SiGe heterojunction bipolar transistors and circuits toward terahertz communication applications," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 10, pp. 2390–2408, Oct. 2004.

[67] J.-S. Rieh, D. Greenberg, A. Stricker, and G. Freeman, "Scaling of SiGe heterojunction bipolar transistors," *Proc. IEEE*, vol. 93, no. 9, pp. 1522–1538, Sep. 2005.

[68] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, 3rd ed. Hoboken, NJ: Wiley, 2007.

[69] S. A. Maas, *Nonlinear Microwave and RF Circuits*, 2nd ed. Norwood, MA: Artech House, 2003.

[70] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 5th ed. Hoboken, NJ: Wiley, 2009.

[71] N. Demirel, E. Kerherve, D. Pache, and R. Plana, "Design techniques and considerations for mmWave SiGe power amplifiers," in *Proc. SBMO/IEEE MTT-S Int. Microw. Optoelectronics Conf. (IMOC)*, Belem, Brazil, Nov. 2009, pp. 37–41.

[72] N. Demirel, E. Kerherve, R. Plana, D. Pache, and D. Belot, "59–71GHz wideband MMIC balanced power amplifier in a 0.13um SiGe technology," in *Proc. Eur. Microw. Conf.*, Rome, Italy, Sep. 2009, pp. 1852–1855.

[73] M. Balducci, H. Schumacher, and S. Chartier, "A Ka-band low power and high-efficiency differential power amplifier in 0.25-µm BiCMOS," in *Integr. Nonlinear Microw. Millimetre-wave Circuits Workshop (INMMiC)*, Graz, Austria, Apr. 2017, pp. 1–3.

[74] M. Rickelt, H. M. Rein, and E. Rose, "Influence of impact-ionization-induced instabilities on the maximum usable output voltage of Si-bipolar transistors," *IEEE Trans. Electron Devices*, vol. 48, no. 4, pp. 774–783, Apr. 2001.

[75] C. T. Kirk, "A theory of transistor cutoff frequency (fT) falloff at high current densities," *IEEE Trans. Electron Devices*, vol. 9, no. 2, pp. 164–174, Mar. 1962.

[76] E. Johnson, "Physical limitations on frequency and power parameters of transistors," in *IRE International Convention Record*, vol. 13, New York, NY, Mar. 1965, pp. 27–34.

[77] *BiCMOS8HP Model Reference Guide*, V1700 ed., GlobalFoundries, Essex Junction, VT, 2017.

[78] M. Rickelt and H. M. Rein, "A novel transistor model for simulating avalanche-breakdown effects in Si bipolar circuits," *IEEE J. Solid-State Circuits*, vol. 37, no. 9, pp. 1184–1197, Sep. 2002.

[79] A. Hastings, *The Art of Analog Layout*, 2nd ed. Upper Saddle River, NJ: Pearson, 2006.

[80] C. M. Grens, P. Cheng, and J. D. Cressler, "Reliability of SiGe HBTs for power amplifiers— part I: Large-signal RF performance and operating limits," *IEEE Trans. Device Mater. Rel.*, vol. 9, no. 3, pp. 431–439, Sep. 2009.

[81] P. Cheng, C. M. Grens, and J. D. Cressler, "Reliability of SiGe HBTs for power amplifiers— part II: Underlying physics and damage modeling," *IEEE Trans. Device Mater. Rel.*, vol. 9, no. 3, pp. 440–448, Sep. 2009.

[82] M. A. Oakley, U. S. Raghunathan, B. R. Wier, P. S. Chakraborty, and J. D. Cressler, "Large-signal reliability analysis of SiGe HBT cascode driver amplifiers," *IEEE Trans. Electron Devices*, vol. 62, no. 5, pp. 1383–1389, May 2015.

[83] C. M. Grens, J. D. Cressler, J. M. Andrews, Q. Liang, and A. J. Joseph, "Reliability issues associated with operating voltage constraints in advanced SiGe HBTs," in *Proc. IEEE Int. Rel. Physics Symp.*, San Jose, CA, Apr. 2005, pp. 409–414.

[84] L. L. Spina, V. d'Alessandro, S. Russo, N. Rinaldi, and L. K. Nanver, "Influence of concurrent electrothermal and avalanche effects on the safe operating area of multifinger bipolar transistors," *IEEE Trans. Electron Devices*, vol. 56, no. 3, pp. 483–491, Mar. 2009.

[85] P. Cheng, C. M. Grens, A. Appaswamy, P. S. Chakraborty, and J. D. Cressler, "Modeling mixed-mode DC and RF stress in SiGe HBT power amplifiers," in *Proc. IEEE Bipolar/BiCMOS Circuits Technol. Meeting (BCTM)*, Monterey, CA, Oct. 2008, pp. 133–136.

[86] G. G. Fischer and S. Glisic, "Temperature stability and reliability aspects of 77 GHz voltage controlled oscillators in a SiGe:C BiCMOS technology," Orlando, FL, Jan. 2008, pp. 171–174.

[87] R. Anholt, *Electrical and Thermal Characterization of MESFETs, HEMTs, and HBTs*. Norwood, MA: Artech House, 1995.

[88] R. C. Joy and E. S. Schlig, "Thermal properties of very fast transistors," *IEEE Trans. Electron Devices*, vol. 17, no. 8, pp. 586–594, Aug. 1970.

[89] J. S. Rieh, J. Johnson, S. Furkay, D. Greenberg, G. Freeman, and S. Subbanna, "Structural dependence of the thermal resistance of trench-isolated bipolar transistors," in *Proc. IEEE Bipolar/BiCMOS Circuits Technol. Meeting (BCTM)*, Minneapolis, MN, Oct. 2002, pp. 100–103.

[90] J.-S. Rieh, D. Greenberg, Q. Liu, A. J. Joseph, G. Freeman, and D. C. Ahlgren, "Structure optimization of trench-isolated SiGe HBTs for simultaneous improvements in thermal and electrical performances," *IEEE Trans. Electron Devices*, vol. 52, no. 12, pp. 2744–2752, Dec. 2005.

[91] A. Pacelli, P. Palestri, and M. Mastrapasqua, "Compact modeling of thermal resistance in bipolar transistors on bulk and SOI substrates," *IEEE Trans. Electron Devices*, vol. 49, no. 6, pp. 1027–1033, Jun. 2002.

[92] J. S. Rieh, D. Greenberg, B. Jagannathan, G. Freeman, and S. Subbanna, "Measurement and modeling of thermal resistance of high speed SiGe heterojunction bipolar transistors," Ann Arbor, MI, Sep. 2001, pp. 110–113.

[93] S. Balanethiram, R. D'Esposito, A. Chakravorty, S. Fregonese, D. Celi, and T. Zimmer, "Efficient modeling of distributed dynamic self-heating and thermal coupling in multi-finger SiGe HBTs," *IEEE Trans. Electron Devices*, vol. 63, no. 9, pp. 3393–3398, Sep. 2016.

[94] S. Lehmann, Y. Zimmermann, A. Pawlak, and M. Schrter, "Characterization of the static thermal coupling between emitter fingers of bipolar transistors," *IEEE Trans. Electron Devices*, vol. 61, no. 11, pp. 3676–3683, Nov. 2014.

[95] S. Balanethiram, A. Chakravorty, R. D'Esposito, S. Fregonese, and T. Zimmer, "Efficient modeling of static self-heating and thermal-coupling in multi-finger SiGe HBTs," in *Proc. IEEE Bipolar/BiCMOS Circuits Technol. Meeting (BCTM)*, Boston, MA, Oct. 2015, pp. 68–71.

[96] P. L. Hower, "Quantitative models for second breakdown in transistors," in *IEEE Power Electronics Specialists Conf.*, Jun. 1973, pp. 67–73.

[97] R. P. Arnold and D. S. Zoroglu, "A quantitative study of emitter ballasting," *IEEE Trans. Electron Devices*, vol. 21, no. 7, pp. 385–391, Jul. 1974.

[98] G. B. Gao, M. S. Unlu, H. Morkoc, and D. L. Blackburn, "Emitter ballasting resistor design for, and current handling capability of AlGaAs/GaAs power heterojunction bipolar transistors," *IEEE Trans. Electron Devices*, vol. 38, no. 2, pp. 185–196, Feb. 1991.

[99] W.-R. Zhang, D.-Y. Jin, J.-W. Yang, L.-J. He, Y.-P. Sha, Y. Wang, and W. Zhang, "Improvement of thermal stability of multi-finger power SiGe HBTs using emitter-ballasting-resistor-free designs," in *Proc. Int. Conf. Solid-State Integr. Circuit Technol. (ICSICT)*, Shanghai, China, Oct. 2006, pp. 221–223.

[100] M. G. Adlerstein, "Thermal stability of emitter ballasted HBT's," *IEEE Trans. Electron Devices*, vol. 45, no. 8, pp. 1653–1655, Aug. 1998.

[101] H. Veenstra, G. A. M. Hurkx, D. van Goor, H. Brekelmans, and J. R. Long, "Analyses and design of bias circuits tolerating output voltages above BVCEO," *IEEE J. Solid-State Circuits*, vol. 40, no. 10, pp. 2008–2018, Oct. 2005.

[102] R. Singh, D. L. Harame, and M. M. Oprysko, *Silicon Germanium: Technology, Modeling, and Design*. Piscataway, NJ: IEEE Press, 2004.

[103] W. R. Eisenstadt and Y. Eo, "S-parameter-based IC interconnect transmission line characterization," *IEEE Trans. Compon., Hybrids, Manuf. Technol.*, vol. 15, no. 4, pp. 483–490, Aug. 1992.

[104] G. Gonzalez, *Microwave Transistor Amplifiers: Analysis and Design*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1997.

[105] D. M. Pozar, *Microwave Engineering*, 4th ed. Hoboken, NJ: Wiley, 2012.

[106] F. T. Ulaby and U. Ravaioli, *Fundamentals of Applied Electromagnetics*, 7th ed. Upper Saddle River, NJ: Pearson, 2015.

[107] P. A. Rizzi, *Microwave Engineering Passive Circuits*. Upper Saddle River, NJ: Prentice Hall, 1988.

[108] K. Wang, K. Ma, and K. S. Yeo, "Design consideration for 60 GHz SiGe power amplifier with ESD protection," in *Proc. Int. SoC Des. Conf.*, Jeju, Korea, Nov. 2011, pp. 5–8.

[109] ——, "V-band high gain SiGe power amplifier with wideband ESD protection," in *Proc. IEEE Int. Wireless Symp. (IWS)*, Shenzhen, China, Mar. 2015, pp. 1–4.

[110] *BiCMOS8HP Design Manual*, V1620 ed., Global Foundries, Essex Junction, VT, 2016.

[111] M. Bozanic and S. Sinha, *Power Amplifiers for the S-, C-, X- and Ku-bands: An EDA Perspective*, 2nd ed. Switzerland: Springer, 2016.

[112] N. Nallam and S. Chatterjee, "Design of concurrent multi-band matching networks," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Rio de Janeiro, Brazil, May 2011, pp. 201–204.

[113] J. E. Carlstrom, R. L. Plambeck, and D. D. Thornton, "A continuously tunable 65–115-GHz Gunn oscillator," *IEEE Trans. Microw. Theory Techn.*, vol. 33, no. 7, pp. 610–619, Jul. 1985.

[114] J. W. May and G. M. Rebeiz, "A W-band SiGe 1.5V LNA for imaging applications," in *IEEE Radio Freq. Integr. Circuits (RFIC) Symp. Dig.*, Atlanta, GA, Jun. 2008, pp. 241–244.

[115] N. Kalantari and J. F. Buckwalter, "A 91 to 110-GHz tapered constructive wave power amplifier in a 0.12μm SiGe BiCMOS process," in *Proc. IEEE Bipolar/BiCMOS Circuits Technol. Meeting (BCTM)*, Austin, TX, Oct. 2010, pp. 125–128.

[116] N. Demirel, E. Kerherv, R. Plana, and D. Pache, "79GHz BiCMOS single-ended and differential power amplifiers," in *Proc. Eur. Microw. Conf.*, Paris, France, Sep. 2010, pp. 1690–1693.

[117] K. Datta and H. Hashemi, "High-breakdown, high-$f_{\max}$ multiport stacked-transistor topologies for the W-band power amplifiers," *IEEE J. Solid-State Circuits*, vol. 52, no. 5, pp. 1–15, Jan. 2017.

[118] ——, "75–105 GHz switching power amplifiers using high-breakdown, high-fmax multiport stacked transistor topologies," in *IEEE Radio Freq. Integr. Circuits (RFIC) Symp. Dig.*, San Francisco, CA, May 2016, pp. 306–309.

[119] Z. Xu, Q. J. Gu, and M. C. F. Chang, "A 100–117 GHz W-band CMOS power amplifier with on-chip adaptive biasing," *IEEE Microw. Wireless Compon. Lett.*, vol. 21, no. 10, pp. 547–549, Oct. 2011.

[120] K. J. Tsai, J. L. Kuo, and H. Wang, "A W-band power amplifier in 65-nm CMOS with 27GHz bandwidth and 14.8dBm saturated output power," in *IEEE Radio Freq. Integr. Circuits (RFIC) Symp. Dig.*, Montreal, QC, Jun. 2012, pp. 69–72.

[121] 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 MOS," in *IEEE Radio and Wireless Symp.*, Austin, TX, Jan. 2013, pp. 256–258.

[122] A. Agah, J. Jayamon, P. Asbeck, J. Buckwalter, and L. Larson, "A 11% PAE, 15.8-dBm two-stage 90-GHz stacked-FET power amplifier in 45-nm SOI CMOS," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Seattle, WA, Jun. 2013, pp. 1–3.

[123] D. A. Chan and M. Feng, "A compact W-band CMOS power amplifier with gain boosting and short-circuited stub matching for high power and high efficiency operation," *IEEE Microw. Wireless Compon. Lett.*, vol. 21, no. 2, pp. 98–100, Feb. 2011.

[124] 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% efficiency," in *IEEE Radio Freq. Integr. Circuits (RFIC) Symp. Dig.*, San Francisco, CA, May 2016, pp. 262–265.

[125] N. Rohani, J. Zhang, J. Lee, and J. Bai, "A 28-nm CMOS 76–81-GHz power amplifier for automotive radar applications," in *Proc. IEEE Topical Meeting Silicon Monolithic Integr. Circuits in RF Syst. (SiRF)*, Phoenix, AZ, Jan. 2017, pp. 85–87.

[126] R. B. Yishay, R. Carmon, O. Katz, and D. Elad, "A high gain wideband 77GHz SiGe power amplifier," in *IEEE Radio Freq. Integr. Circuits (RFIC) Symp. Dig.*, Anaheim, CA, May 2010, pp. 529–532.

[127] V. Giammello, E. Ragonese, and G. Palmisano, "A 15-dBm SiGe BiCMOS PA for 77-GHz automotive radar," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 11, pp. 2910–2918, Nov. 2011.

[128] ——, "A transformer-coupling current-reuse SiGe HBT power amplifier for 77-GHz automotive radar," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 6, pp. 1676–1683, Jun. 2012.

[129] R. B. Yishay and D. Elad, "A 17.5-dBm D-band power amplifier and doubler chain in SiGe BiCMOS technology," in *Proc. Eur. Microw. Integr. Circuit Conf. (EuMIC)*, Rome, Italy, Oct. 2014, pp. 53–56.

[130] Y. Zhao, J. R. Long, M. Spirito, and A. B. Akhnoukh, "A +18dBm, 79–87.5GHz bandwidth power amplifier in 0.13μm SiGe-BiCMOS," in *Proc. IEEE Bipolar/BiCMOS Circuits Technol. Meeting (BCTM)*, Atlanta, GA, Oct. 2011, pp. 17–20.

[131] Y. Zhao and J. R. Long, "A wideband, dual-path, millimeter-wave power amplifier with 20 dBm output power and PAE above 15% in 130 nm SiGe-BiCMOS," *IEEE J. Solid-State Circuits*, vol. 47, no. 9, pp. 1981–1997, Sep. 2012.

[132] R. B. Yishay, R. Carmon, O. Katz, B. Sheinman, and D. Elad, "A 20dBm E-band power amplifier in SiGe BiCMOS technology," in *Proc. Eur. Microw. Integr. Circuit Conf. (EuMIC)*, Amsterdam, Netherlands, Oct. 2012, pp. 699–702.

[133] F. Ahmed, M. Furqan, K. Aufinger, and A. Stelzer, "A SiGe-based broadband 100–180-GHz differential power amplifier with 11 dBm peak output power and >1.3THz GBW," in *Proc. Eur. Microw. Integr. Circuit Conf. (EuMIC)*, London, United Kingdom, Oct. 2016, pp. 257–260.

[134] N. Sarmah, P. Chevalier, and U. R. Pfeiffer, "160-GHz power amplifier design in advanced SiGe HBT technologies with P_{sat} in excess of 10 dBm," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 2, pp. 939–947, Feb. 2013.

[135] M. Furqan, F. Ahmed, H. Rucker, and A. Stelzer, "A 140-180-GHz broadband amplifier with 7 dBm OP1dB and 400 GHz GBW in SiGe BiCMOS," in *Proc. IEEE Compound Semicond. Integr. Circuits Symp. (CSICS)*, New Orleans, LA, Oct. 2015, pp. 1–4.

[136] N. Sarmah, B. Heinemann, and U. R. Pfeiffer, "A 135–170 GHz power amplifier in an advanced sige HBT technology," in *IEEE Radio Freq. Integr. Circuits (RFIC) Symp. Dig.*, Seattle, WA, Jun. 2013, pp. 287–290.

[137] L. Zhang, J. Wen, L. Sun, and T. Wu, "A three stage, fully differential D-band power amplifier," in *Proc. Int. Conf. Solid-State Integr. Circuit Technol. (ICSICT)*, Oct. 2014, pp. 1–3.

[138] D. Hou, Y. Z. Xiong, W. L. Goh, W. Hong, and M. Madihian, "A D-band cascode amplifier with 24.3 dB gain and 7.7 dBm output power in 0.13 um SiGe BiCMOS technology," *IEEE Microw. Wireless Compon. Lett.*, vol. 22, no. 4, pp. 191–193, Apr. 2012.

[139] H. C. Lin and G. M. Rebeiz, "A 112–134 GHz SiGe amplifier with peak output power of 120 mW," in *IEEE Radio Freq. Integr. Circuits (RFIC) Symp. Dig.*, Tampa, FL, Jun. 2014, pp. 163–166.

[140] R. B. Yishay and D. Elad, "A 17.8 dBm 110–130 GHz power amplifier and doubler chain in SiGe BiCMOS technology," in *IEEE Radio Freq. Integr. Circuits (RFIC) Symp. Dig.*, Phoenix, AZ, May 2015, pp. 391–394.

[141] H. T. Dabag, J. Kim, L. E. Larson, J. F. Buckwalter, and P. M. Asbeck, "A 45-GHz SiGe HBT amplifier at greater than 25% efficiency and 30 mW output power," in *Proc. IEEE Bipolar/BiCMOS Circuits Technol. Meeting (BCTM)*, Atlanta, GA, Oct. 2011, pp. 25–28.

[142] N. Kalantari and J. F. Buckwalter, "A nested-reactance feedback power amplifier for Q-band applications," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 6, pp. 1667–1675, Jun. 2012.

[143] Y. Pei, Y. Chen, and D. M. W. Leenaerts, "A broad-band BiCMOS transmitter front-end for 27–36GHz phased array systems," in *Proc. IEEE Bipolar/BiCMOS Circuits Technol. Meeting (BCTM)*, Coronado, CA, Sep. 2014, pp. 64–67.

[144] K. Ma, T. B. Kumar, and K. S. Yeo, "A reconfigurable K-/Ka-band power amplifier with high PAE in 0.18-um SiGe BiCMOS for multi-band applications," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 12, pp. 4395–4405, Dec. 2015.

[145] T. J. Farmer, A. Darwish, B. Huebschman, E. Viveiros, H. A. Hung, and M. E. Zaghloul, "Millimeter-wave SiGe HBT high voltage/high power architecture implementation," *IEEE Microw. Wireless Compon. Lett.*, vol. 21, no. 10, pp. 544–546, Oct. 2011.

[146] Y. Chen, Y. Pei, and D. M. W. Leenaerts, "A fully integrated 30GHz 16-QAM single-channel phased array transmitter with 5.9% EVM at 6dB back-off," in *Proc. Eur. Solid-State Circuits Conf. (ESSCIRC)*, Graz, Austria, Sep. 2015, pp. 92–95.

[147] B. Welp, K. Noujeim, and N. Pohl, "A wideband 20 to 28 GHz signal generator MMIC with 30.8 dBm output power based on a power amplifier cell with 31% PAE in SiGe," *IEEE J. Solid-State Circuits*, vol. 51, no. 9, pp. 1975–1984, Sep. 2016.

[148] W. Tai, L. R. Carley, and D. S. Ricketts, "A Q-band SiGe power amplifier with 17.5 dBm saturated output power and 26% peak PAE," in *Proc. IEEE Bipolar/BiCMOS Circuits Technol. Meeting (BCTM)*, Atlanta, GA, Oct. 2011, pp. 146–149.

[149] K. Datta, J. Roderick, and H. Hashemi, "Analysis, design and implementation of mm-Wave SiGe stacked class-E power amplifiers," in *IEEE Radio Freq. Integr. Circuits (RFIC) Symp. Dig.*, Seattle, WA, Jun. 2013, pp. 275–278.

[150] S. Y. Mortazavi and K. J. Koh, "A 43% PAE inverse class-F power amplifier at 39–42 GHz with a $\lambda/4$-transformer based harmonic filter in 0.13-µm SiGe BiCMOS," in *IEEE MTT-S Int. Microw. Symp. Dig.*, San Francisco, CA, May 2016, pp. 1–4.

[151] ——, "Integrated inverse class-F silicon power amplifiers for high power efficiency at microwave and mm-Wave," *IEEE J. Solid-State Circuits*, vol. 51, no. 10, pp. 1–15, Aug. 2016.

[152] ——, "A 38 GHz inverse class-F power amplifier with 38.5% peak PAE, 16.5 dB gain, and 50 mW Psat in 0.13-µm SiGe BiCMOS," in *IEEE Radio Freq. Integr. Circuits (RFIC) Symp. Dig.*, Phoenix, AZ, May 2015, pp. 211–214.

[153] W. Tai, L. R. Carley, and D. S. Ricketts, "A 0.7W fully integrated 42GHz power amplifier with 10% PAE in 0.13µm SiGe BiCMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, Feb. 2013, pp. 142–143.

[154] N. Kalantari and J. F. Buckwalter, "A 19.4 dBm, Q-band class-E power amplifier in a 0.12 µ m SiGe BiCMOS process," *IEEE Microw. Wireless Compon. Lett.*, vol. 20, no. 5, pp. 283–285, May 2010.

[155] J. Essing, D. Leenaerts, and R. Mahmoudi, "A 27GHz, 31dBm power amplifier in a 0.25µm SiGe:C BiCMOS technology," in *Proc. IEEE Bipolar/BiCMOS Circuits Technol. Meeting (BCTM)*, Coronado, CA, Sep. 2014, pp. 143–146.

[156] S. Daneshgar and J. F. Buckwalter, "A 22 dBm, 0.6 mm^2 D-band SiGe HBT power amplifier using series power combining sub-quarter-wavelength baluns," in *Proc. IEEE Compound Semicond. Integr. Circuits Symp. (CSICS)*, New Orleans, LA, Oct. 2015, pp. 1–4.

[157] Y. Zhao, J. R. Long, and M. Spirito, "A 60GHz-band 20dBm power amplifier with 20% peak PAE," in *IEEE Radio Freq. Integr. Circuits (RFIC) Symp. Dig.*, Baltimore, MD, Jun. 2011, pp. 1–4.

[158] P. Indirayanti and P. Reynaert, "A 32 GHz 20 dBm-PSAT transformer-based Doherty power amplifier for multi-Gb/s 5G applications in 28 nm bulk CMOS," in *IEEE Radio Freq. Integr. Circuits (RFIC) Symp. Dig.*, Honolulu, HI, Jun. 2017, pp. 45–48.

[159] H. Jia, C. C. Prawoto, B. Chi, Z. Wang, and C. P. Yue, "A 32.9% PAE, 15.3 dBm, 21.6–41.6 GHz power amplifier in 65nm CMOS using coupled resonators," in *Proc. IEEE Asian Solid-State Circuits Conf. (ASSCC)*, Toyama, Japan, Nov. 2016, pp. 345–348.

[160] S. Shakib, H. C. Park, J. Dunworth, V. Aparin, and K. Entesari, "A 28GHz efficient linear power amplifier for 5G phased arrays in 28nm bulk CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, Jan. 2016, pp. 352–353.

[161] ——, "A highly efficient and linear power amplifier for 28-GHz 5G phased array radios in 28-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 51, no. 12, pp. 3020–3036, Dec. 2016.

[162] H. Alsuraisry, S. T. Yen, J. H. Tsai, and T. W. Huang, "Ka-band up-link CMOS/GaAs power amplifier design for satellite-based wireless sensor," in *Topical Workshop Internet of Space (TWIOS)*, Phoenix, AZ, Jan. 2017, pp. 1–3.

[163] S. R. Helmi, J. H. Chen, and S. Mohammadi, "High-efficiency microwave and mm-Wave stacked cell CMOS SOI power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 7, pp. 2025–2038, Jul. 2016.

[164] B. Park, S. Jin, D. Jeong, J. Kim, Y. Cho, K. Moon, and B. Kim, "Highly linear mm-Wave CMOS power amplifier," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 12, pp. 4535–4544, Dec. 2016.

[165] S. Shakib, M. Elkholy, J. Dunworth, V. Aparin, and K. Entesari, "A wideband 28GHz power amplifier supporting 8 × 100MHz carrier aggregation for 5G in 40nm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, Feb. 2017, pp. 44–45.

[166] J.-W. Lee and S.-M. Heo, "A 27 GHz, 14 dBm CMOS power amplifier using 0.18 μm common-source MOSFETs," *IEEE Microw. Wireless Compon. Lett.*, vol. 18, no. 11, pp. 755–757, Nov. 2008.

[167] C. W. Kuo, H. K. Chiou, and H. Y. Chung, "An 18 to 33 GHz fully-integrated Darlington power amplifier with Guanella-type transmission-line transformers in 0.18 μm CMOS technology," *IEEE Microw. Wireless Compon. Lett.*, vol. 23, no. 12, pp. 668–670, Dec. 2013.

[168] A. Vasylyev, P. Weger, and W. Simburger, "Ultra-broadband 20.5-31 GHz monolithically-integrated CMOS power amplifier," *IET Electron. Letters*, vol. 41, no. 23, pp. 1281–1282, Nov. 2005.

[169] S. Pornpromlikit, H. T. Dabag, B. Hanafi, J. Kim, L. E. Larson, J. F. Buckwalter, and P. M. Asbeck, "A Q-band amplifier implemented with stacked 45-nm CMOS FETs," in *Proc. IEEE Compound Semicond. Integr. Circuits Symp. (CSICS)*, Oct. 2011, pp. 1–4.

[170] J.-S. Rieh, B. Jagannathan, H. Chen, K. Schonenberg, S.-J. Jeng, M. Khater, D. Ahlgren, G. Freeman, and S. Subbanna, "Performance and design considerations for high speed SiGe HBTs of fT/fmax=375 GHz/210 GHz," in *IEEE Conf. Indium Phosphide and Related Materials*, May 2003, pp. 374–377.

[171] J. P. Comeau, "Integration issues associated with monolithic silicon-germanium microwave radar systems," Ph.D. dissertation, Georgia Institute of Technol., Atlanta, GA, Dec. 2006.

[172] E. Jarvinen, S. Kalajo, and M. Matilainen, "Bias circuits for GaAs HBT power amplifiers," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 1, Phoenix, AZ, May 2001, pp. 507–510.

[173] U. R. Pfeiffer and A. Valdes-Garcia, "Millimeter-wave design considerations for power amplifiers in an SiGe process technology," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 1, pp. 57–64, Jan. 2006.

[174] J.-S. Rieh, K. M. Watson, F. Guarin, Z. Yang, P.-C. Wang, A. J. Joseph, G. Freeman, and S. Subbanna, "Reliability of high-speed SiGe heterojunction bipolar transistors under very high forward current density," *IEEE Trans. Device Mater. Rel.*, vol. 3, no. 2, pp. 31–38, Jun. 2003.

[175] J. W. May, R. A. Alhalabi, and G. M. Rebeiz, "A 3 G-bit/s W-band SiGe ASK receiver with a high-efficiency on-chip electromagnetically-coupled antenna," in *IEEE Radio Freq. Integr. Circuits (RFIC) Symp. Dig.*, Anaheim, CA, May 2010, pp. 87–90.

[176] R. A. Alhalabi and G. M. Rebeiz, "Design of high-efficiency millimeter-wave microstrip antennas for silicon RFIC applications," in *Proc. IEEE Int. Symp. Antennas Propag. (APSURSI)*, Spokane, WA, Jul. 2011, pp. 2055–2058.

[177] J. B. Hacker, A. L. Sailer, B. Brar, G. Nagy, R. L. J. Pierson, and J. A. Higgins, "A high-power W-band quasi-optical frequency tripler," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 3, Philadelphia, PA, Jun. 2003, pp. 1859–1862.

[178] Y. Chung, C.-T. Cheung, M. P. DeLisio, and D. B. Rutledge, "W-band waveguide-packaged InP HEMT reflection grid amplifier," *IEEE Microw. Wireless Compon. Lett.*, vol. 16, no. 6, pp. 324–326, Jun. 2006.

[179] M. A. Gouker, J. T. Delisle, and S. M. Duffy, "A circuit-fed tile-approach configuration for millimeter-wave spatial power combining," *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 1, pp. 17–21, Jan. 2002.

[180] C.-M. Liu, E. A. Sovero, W. J. Ho, J. A. Higgins, M. P. DeLisio, and D. B. Rutledge, "Monolithic 40-GHz 670-mW HBT grid amplifier," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 2, San Francisco, CA, Jun. 1996, pp. 1123–1126.

[181] M. P. DeLisio, B. C. Deckman, C.-T. Cheung, S. C. Martin, C. J. Rollison, J. J. Rosenberg, G. Smith, and J. Eisenberg, "Power and spectral regrowth performance of 10-W and 16-W Ka-band power amplifiers with single-chip output stages," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Long Beach, CA, Jun. 2005, pp. 839–842.

[182] J. Hubert, J. Schoenberg, and Z. B. Popovic, "High-power hybrid quasi-optical Ka-band amplifier design," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Orlando, FL, May 1995, pp. 585–588.

[183] S. Mandegaran and A. Hajimiri, "A breakdown voltage multiplier for high voltage swing drivers," *IEEE J. Solid-State Circuits*, vol. 42, no. 2, pp. 302–312, Feb. 2007.

[184] G. Qin, N. Jiang, G. Wang, and Z. Ma, "Configuration dependence of SiGe HBT linearity characteristics," in *Proc. Eur. Microw. Integr. Circuit Conf. (EuMIC)*, Manchester, United Kingdom, Sep. 2006, pp. 107–110.

[185] G. Qin, G. Wang, N. Jiang, and Z. Ma, "Tradeoff between CE and CB SiGe HBTs for power amplification in terms of frequency-dependent linearity and power-gain characteristics," in *Proc. Topical Meeting Silicon Monolithic Integr. Circuits in RF Syst.*, Long Beach, CA, Jan. 2007, pp. 1–4.

[186] G. Niu, Q. Liang, J. D. Cressler, C. S. Webster, and D. L. Harame, "RF linearity characteristics of SiGe HBTs," *IEEE Trans. Microw. Theory Techn.*, vol. 49, no. 9, pp. 1558–1565, Sep. 2001.

[187] Y. Li, J. Lopez, D. Y. C. Lie, K. Chen, S. Wu, and T. Y. Yang, "Efficiency enhancement and linearity trade-offs for cascode vs. common-emitter SiGe power amplifiers in WiMAX polar transmitters," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Paris, France, May 2010, pp. 1915–1918.