

mm-Wave Silicon Technology 60 GHz and Beyond

Series on Integrated Circuits and Systems

Series Editor: Anantha Chandrakasan
Massachusetts Institute of Technology
Cambridge, Massachusetts

mm-Wave Silicon Technology: 60 GHz and Beyond
Ali M. Niknejad and Hossein Hashemi (Eds.)
ISBN 978-0-387-76558-7

Ultra Wideband: Circuits, Transceivers, and Systems
Ranjit Gharpurey and Peter Kinget (Eds.)
ISBN 978-0-387-37238-9

Creating Assertion-Based IP
Harry D. Foster and Adam C. Krolnik
ISBN 978-0-387-36641-8

Design for Manufacturability and Statistical Design: A Constructive Approach
Michael Orshansky, Sani R. Nassif, and Duane Boning
ISBN 978-0-387-30928-6

Low Power Methodology Manual: For System-on-Chip Design
Michael Keating, David Flynn, Rob Aitken, Alan Gibbons, and Kaijian Shi
ISBN 978-0-387-71818-7

Modern Circuit Placement: Best Practices and Results
Gi-Joon Nam and Jason Cong
ISBN 978-0-387-36837-5

CMOS Biotechnology
Hakho Lee, Donhee Ham and Robert M. Westervelt
ISBN 978-0-387-36836-8

SAT-Based Scalable Formal Verification Solutions
Malay Ganai and Aarti Gupta
ISBN 978-0-387-69166-4, 2007

Ultra-Low Voltage Nano-Scale Memories
Kiyoo Itoh, Masashi Horiguchi and Hitoshi Tanaka
ISBN 978-0-387-33398-4, 2007

Routing Congestion in VLSI Circuits: Estimation and Optimization
Prashant Saxena, Rupesh S. Shelar, Sachin Sapatnekar
ISBN 978-0-387-30037-5, 2007

Ultra-Low Power Wireless Technologies for Sensor Networks
Brian Otis and Jan Rabaey
ISBN 978-0-387-30930-9, 2007

Sub-Threshold Design for Ultra Low-Power Systems
Alice Wang, Benton H. Calhoun and Anantha Chandrakasan
ISBN 978-0-387-33515-5, 2006

High Performance Energy Efficient Microprocessor Design
Vojin Oklobdzija and Ram Krishnamurthy (Eds.)
ISBN 978-0-387-28594-8, 2006

Abstraction Refinement for Large Scale Model Checking
Chao Wang, Gary D. Hachtel, and Fabio Somenzi
ISBN 978-0-387-28594-2, 2006

Continued after index

Ali M. Niknejad • Hossein Hashemi
Editors

mm-Wave Silicon Technology 60 GHz and Beyond

 Springer

Editors:

Ali M. Niknejad
University of California, Berkeley
Berkeley, CA
USA

Hossein Hashemi
University of Southern California
Los Angeles, CA
USA

Series Editor:

Anantha Chandrakasan
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139
USA

Library of Congress Control Number: 2007938756

ISBN 978-0-387-76558-7

e-ISBN 978-0-387-76561-7

Printed on acid-free paper.

© 2008 Springer Science+Business Media, LLC

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use in this publication of trade names, trademarks, service marks and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

9 8 7 6 5 4 3 2 1

springer.com

*Dedicated to our students, who have boldly
gone where no one has gone before.*

Preface

Seven years ago research in the field of mm-wave silicon was virtually non-existent. Few people thought that operation at 60 GHz was even feasible in silicon technology. In the course of seven years the topic has transitioned from an obscure research topic to an exciting buzzword (60 GHz) that has generated much interest from industry and the venture community. To put things in historical perspective, seven years ago most commercial efforts were focused on the 1-10 GHz spectrum for voice and data applications for mobile phones and portable computers. Many people were actively seeking solutions to the “last mile” problem, or a way to deliver high speed data to users in their homes and offices through cable, telephone, or wireless infrastructure. At the same time, the explosive growth of wireless data such as WiFi spurred significant research into and development of new architectures for radio transceivers that could deliver very high data rates over short ranges, particularly for video and personal area networks. This problem can be viewed as the “last meter” or even the “last inch” connection that delivers high bandwidth multimedia content to devices. The growth of MP3 media devices, and now handheld video devices, and the rapid adoption of HDTV and flat screen televisions has created a healthy demand for technology that enables high speed wireless video transmission. For this reason, today we witness a very active interest in mm-wave silicon technology. Other important commercial applications include automotive radar for safety and improved driving experience. But these applications are only the tip of the iceberg.

Research in silicon mm-wave circuits and technology started at universities and research institutions and now continues in the commercial realm. Many of the researchers in the field have contributed to this book to chronicle these efforts. This book is written for practicing RF and analog circuit designers who wish to join this exciting and growing field. The chapters are self-contained and include a short tutorial on important concepts before delving into details. The audience for this book is assumed to be experienced in the field of analog/RF integrated circuits. The focus of each chapter is the key innovations and techniques that enable operation at mm-wave frequencies, which is close to the activity limits of silicon technology. Many of the chapters are focused around several key publications in the field. Rather than republish the original papers, the authors have gone to great lengths to expand the material

and provide more background and breadth than the original technical publication. As such, this book would complement a graduate microwave or mm-wave course based on silicon technology.

The book begins with the fundamental technology scaling and device-level changes that have allowed mm-wave silicon performance. This includes a detailed discussion of the design, modeling, and achievable performance of active and passive components. Next, front-end mm-wave building blocks are covered in detail, including amplifiers and mixers, voltage-controlled oscillators (VCOs) and dividers, and power amplifiers. Complete chipsets operating in the 60 GHz band are also covered. The book culminates in the discussion of phased-array systems capable of beam forming and exploiting spatial diversity for increased throughput or range. Throughout the book we focus on a design methodology and design tools that are optimized for silicon technology. Unique issues related to silicon, such as the lossy silicon substrate, are discussed at each stage of the book. Other important issues include the ever shrinking low supply voltage, which limits the dynamic range and output power capability of the technology.

We hope you find this book useful in your exploration of silicon mm-wave devices, circuits, and systems. There are many unique challenges to working with silicon mm-waves, but there are many rewards to reap from this technology. As the editors, we would like to thank the contributors to this book, including the graduate students and the contributing authors who have worked tirelessly to share their insights with you in this book. We also thank Carl Harris of Springer for making this project happen. To the reader our advice is simple: *think small!*

Berkeley and Los Angeles,
October, 2007

Ali M. Niknejad
Hossein Hashemi

Acknowledgements

The editors of the book would like to greatly acknowledge the work of several generations of graduate students at University of California, Berkeley and at the University of Southern California who have worked diligently to make our mm-wave dreams a practical reality. This research would not be possible without the support of our sponsors. The DARPA TEAM program was instrumental in enabling many of the research efforts. In particular, we thank Dan Radack and Bob Reuss for recognizing the important role that will be played by silicon in mm-wave applications and starting the TEAM program, and Michael Fritze for continuing the program. We also thank Barry Perlman, the Army Technical Monitor, who oversaw our research and helped us make many links in the field. We would like to acknowledge Boeing Phantom Works, Charles Lee Powell Foundation, and the Viterbi School of Engineering for their support of USC's silicon based integrated beam forming research. We also thank the BWRC member companies, in particular STMicroelectronics for donation of chip fabrication and for support. We also thank IBM for creating a wonderful and powerful SiGe process technology that has fueled many research efforts. The research was also supported through an NSF Infrastructure Grant 0403427, NSF grants ECCS-0702037 and ECCS-0636677, and through the C2S2 Focus Center.

Contents

1	Introduction to mm-Wave Silicon Devices, Circuits, and Systems	1
	Ali M. Niknejad and Hossein Hashemi	
1.1	Introduction	1
1.2	Why mm-Waves?	3
1.3	The Birth of Silicon mm-Wave	5
1.3.1	Why CMOS?	8
1.3.2	True Cost of Silicon mm-Wave	10
1.4	Communication in the 60 GHz Band	12
1.4.1	Beam Forming	13
1.5	Unique mm-Wave Applications	17
1.5.1	mm-Wave Spectrum	17
1.5.2	Automotive Radar	18
1.5.3	mm-Wave Imaging for Medical Applications	22
1.5.4	Collaborative Distributed MIMO	23
1.6	Overview of Book	24
	References	24
2	Silicon Technologies to Address mm-Wave Solutions	25
	Andreia Cathelin and John J. Pekarik	
2.1	Why Silicon?	25
2.1.1	Performance	26
2.1.2	Cost, Integration [3]	27
2.1.3	Manufacturing Capacity	27
2.2	Modern SiGe and CMOS Technology	28
2.2.1	Lithography	28
2.2.2	Low-K Dielectrics and Copper Wiring	30
2.2.3	Mobility and Strain Engineering	30
2.2.4	Metal Gates & High-K Dielectrics	32
2.3	Active Devices on Recent Bulk and SOI Technologies	32
2.3.1	Bipolar Devices	33
2.3.2	CMOS devices	34

2.3.3	SOI CMOS devices	41
2.3.4	Current Density Scaling for CMOS and Bipolar Devices	46
2.3.5	Comparison Between State-of-the-Art HBT and CMOS Devices	46
2.4	Impact of the Back-End of Line on mm-Wave Design	49
2.5	Conclusion	55
2.6	Acknowledgements:	55
	References	56
3	Design and Modeling of Active and Passive Devices	59
	Ali M. Niknejad, Sohrab Emami, Chinh Doan, Babak Heydari, Mounir Bohsali	
3.1	Passive Devices	59
3.1.1	Transmission Lines	59
3.1.2	Inductors	70
3.1.3	Capacitors	72
3.1.4	Transformers	75
3.1.5	Resonators	77
3.2	Active Devices	79
3.2.1	Modeling	79
3.2.2	Active Device Design	80
3.2.3	Small-Signal Model	89
3.2.4	Large-Signal Model	90
3.2.5	FET Noise Model	97
3.3	Conclusion	105
	References	107
4	Amplifiers and Mixers	109
	Ali M. Niknejad, Brian A. Floyd, Sohrab Emami, Babak Heydari, Ehsan Adabi, Bagher Afshar	
4.1	60 GHz Low-Noise Amplifiers: What's Different?	109
4.1.1	Transistors Closer to Cutoff	109
4.1.2	Small Wavelengths	110
4.1.3	Parasitics at 60 GHz	111
4.2	Low-Noise Amplifier Design Methodology	111
4.2.1	Input Match Optimization for Noise and Power	112
4.2.2	Transistor Noise Parameters	113
4.2.3	Common-Base vs. Common-Emitter	115
4.3	Low-Noise Amplifier Examples	117
4.3.1	Bipolar LNA (v1), Common-Base Input	117
4.3.2	Bipolar LNA (v2), Common-Emitter Input	121
4.3.3	CMOS Common Source Amplifiers	122
4.3.4	CMOS Common Gate Amplifiers	126
4.3.5	Differential Pair Amplifiers	127
4.3.6	Multi-Stage Amplifier Design	129
4.3.7	A Two-Stage 30 GHz Amplifier	135

4.4	Mixers and Frequency Translation	137
4.4.1	Single Transistor Mixers	137
4.4.2	Dual Gate Mixers	141
4.4.3	Gilbert Cell Mixers	144
4.5	Examples of Integrated Front-Ends	147
4.5.1	CMOS 130nm 60 GHz Front-End	147
4.5.2	SiGe Transceiver Chipset	150
4.6	Conclusion	155
	References	156
5	Voltage-Controlled Oscillators and Frequency Dividers	159
	Jri Lee	
5.1	Considerations of VCOs	159
5.2	Cross-Coupled Oscillators	162
5.3	Colpitts Oscillator	169
5.4	Other Topologies	173
5.4.1	mm-Wave Oscillators	174
5.4.2	Push-Push Oscillators	176
5.4.3	Distributed Oscillators	177
5.5	Considerations of Dividers	178
5.6	Static Dividers	180
5.7	Regenerative (Miller) Dividers	183
5.8	Injection-Locked Dividers	187
5.9	Case Study	190
5.9.1	52-GHz LO Signal Generator	191
5.9.2	60-GHz PLL in 0.15- μ m GaAs	192
5.9.3	A 75-GHz PLL in 90-nm CMOS	195
	References	199
6	Power Amplifiers at 60GHz and Beyond	201
	Ehsan Afshari and Abbas Komijani	
6.1	Motivation and Challenges	201
6.2	Passive Components	202
6.2.1	Substrate-Shielded Coplanar Waveguide Structure	202
6.2.2	Characterization of the Substrate-Shielded CPW Structure	204
6.2.3	Conductor-Backed Coplanar Waveguide as the Transmission Line Structure	207
6.2.4	Wirebond and Pad Parasitic Effects	209
6.3	Power Transistors	209
6.3.1	Single-Transistor Power Gain and Stability	210
6.3.2	Stability of the Cascode Pair	212
6.3.3	Relationship of Breakdown Voltage and Cut-off Frequency	214
6.4	Power Combining Techniques	215
6.4.1	Basic Principle	215

6.4.2	Distributed Active Transformer	216
6.4.3	Electrical Funnel	220
6.4.4	Circuit Implementation	228
6.5	Case Studies	229
6.5.1	A 77GHz Amplifier for Automotive RADAR Application	230
6.5.2	A Broadband Amplifier at 85GHz	233
6.6	Summary	238
	References	240
7	Integrated Beamforming Arrays	243
	Harish Krishnaswamy and Hossein Hashemi	
7.1	Introduction	243
7.2	What is a Phased Array?	245
7.2.1	Case Study: A 60GHz WPAN Link Budget	248
7.3	Phased Arrays versus Timed Arrays	249
7.4	Conventional Phased Array Architectures	253
7.4.1	RF Phase-shifting	254
7.4.2	LO Phase-shifting	263
7.4.3	Digital Arrays	269
7.4.4	Comparative View of the Conventional Architectures	269
7.5	The VPRO-PLL Phased Array Architecture	271
7.5.1	VPRO Concept	271
7.5.2	Transmit Mode	272
7.5.3	Receive Mode	273
7.6	The Effect of Mismatch in Phased Arrays	277
7.6.1	Beam-pointing Error	278
7.6.2	Sidelobe Rejection Ratio	280
7.6.3	Implications on Array Packaging	281
7.6.4	Array Calibration	282
7.7	Quantization Error in Phased Arrays	283
7.8	Multi-Beam Antenna Arrays	285
7.9	Antenna Arrays and Multiple Input Multiple Output (MIMO) Transceivers	290
7.10	Concluding Remarks	291
	References	293
	Index	297

List of Contributors

Ali M. Niknejad

Dept. of EECS, Berkeley Wireless Research Center, U.C. Berkeley, CA

Hossein Hashemi

Department of Electrical Engineering - Electrophysics, University of Southern California, Los Angeles, CA

Andreia Cathelin

STMicroelectronics, Crolles, France

John J. Pekarik

IBM Corporation, Essex Junction, Vermont

Chinh Doan

Sohrab Emami

SiBEAM, Sunnyvale, CA

Babak Heydari, Mounir Bohsali, Ehsan Adabi, Bagher Afshar

Dept. of EECS, Berkeley Wireless Research Center, U.C. Berkeley, CA

Brian A. Floyd

IBM T. J. Watson Research Center, Yorktown Heights, NY

Jri Lee

EE Dept., National Taiwan University, Taipei, Taiwan

Ehsan Afshari

Cornell University, Ithaca, NY

Abbas Komijani

Entropic Communications, San Jose, CA

Harish Krishnaswamy

Department of Electrical Engineering - Electrophysics, University of Southern California, Los Angeles, CA

Chapter 1

Introduction to mm-Wave Silicon Devices, Circuits, and Systems

Ali M. Niknejad and Hossein Hashemi

1.1 Introduction

Silicon-based RF and microwave technology has had a dramatic impact on the world of wireless technology. Today we can access voice/data and entertainment in virtually every corner of the globe, from short range Bluetooth and WiFi networks, to cellular and satellite networks, to meet different range and throughput requirements. A laptop computer without wireless capability is unthinkable today whereas ten years ago the technologies did not exist at all. What do the next ten years promise? What gaps in wireless technology exist even today? Perhaps the most obvious missing link is between the various peripherals that we carry with us, such as cellular phones or personal digital assistants (or “smart phones” if you prefer), digital cameras, music and video players (such as the ubiquitous iPod), laptops, peripherals such as external hard drives and monitors. The case of the mobile phone is particularly important since the existing wireless connectivity is either too slow and power hungry (Bluetooth) or designed and optimized for longer ranges (WiFi). What is missing is a wireless USB capability that can support high data rates demanded by large datarate multimedia applications. Wireless technology has been conspicuously absent from MP3 music players (such as Apple’s iPod), which are ideal candidates for downloading music and video. While nascent UWB technology is a potential solution, it has some shortcomings including problems with interference and a limited data rate. The 3–5 GHz spectrum is relatively crowded with many interferers appearing in the WiFi bands. A UWB system has very limited transmit power (average of about 0 dBm) which means that bandwidth has to be traded to overcome the limited SNR. In comparison, the 60 GHz band offers the same amount of spectrum (7 GHz), virtually no

Ali M. Niknejad
University of California, Berkeley, e-mail: niknejad@berkeley.edu

Hossein Hashemi
University of Southern California, Los Angeles, CA 90089-0271, e-mail: hossein@usc.edu



Fig. 1.1 A hypothetical incarnation of the “iPhone” with mm-wave capability can communicate with several other devices simultaneously using beam forming at a data rate of over 1 Gb/s.

interference¹, and up to 40 dBm of transmit power. Given this bandwidth, it is easy to envision a wireless link capable of supporting multi-Gb/s communication between devices transporting large volumes of media information (video) .

Consider a hypothetical 60 GHz enabled iPhone shown in Fig. 1.1, which uses the 60 GHz connection to download movies from a kiosk (perhaps at a train station or airport), transmits video to a larger screen for easier viewing, and connects to external peripherals such as hard disks and wired and optical networks. If such a device is realized with reasonable power consumption, we see that it can truly displace the laptop computer. Given that this 60 GHz spectrum can be exploited using inexpensive silicon technology, there is quite a lot of excitement and energy in academia and industry focused on silicon mm-wave technology. In Fig. 1.2 we show a plot of the number of papers published worldwide with the keyword “mm-wave”, “60 GHz”, or “77 GHz”. The graph is further separated into papers using CMOS and SiGe. As evident in the figure, interest has grown considerably in this technology in this decade.

Research in mm-wave electronics did not of course start with silicon and predates much of the current generation of work in the area. Active use of mm-wave includes

¹ There is no interference today, but we envision a near future crowded with 60 GHz devices. Nevertheless, the higher propagation loss at 60 GHz means that networks will be naturally spatially isolated.

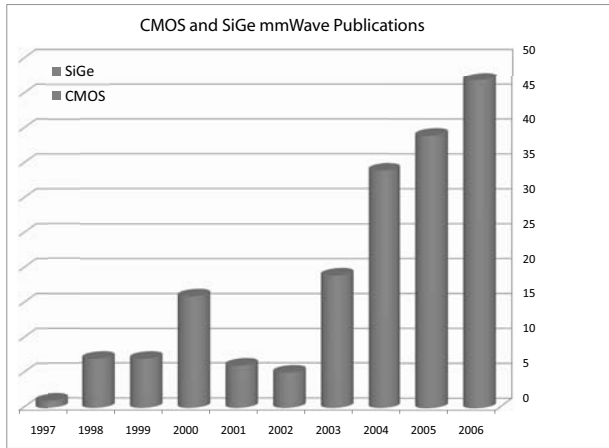


Fig. 1.2 The number of mm-wave papers published in the past decade. Search results obtained from IEEE Xplore database using “60 GHz” or “mm-wave” and “CMOS” or “SiGe” as keywords.

military radar systems, radio astronomy, and space programs. But commercialization efforts, though, have failed to materialize into mm-wave products, mainly due to cost and volume considerations. With the exception of automotive radar in a limited number of luxury vehicles, there are virtually no practical applications for mm-waves today. In fact, when the 60 GHz project was started at the Berkeley Wireless Research Center (BWRC) in 2000, most of our industrial members did not see any potential for the technology in the foreseeable future. Many thought it was good long range research that may materialize in ten years. The pessimists doubted that the technology could be realized with silicon technology, worse yet with CMOS. But in the course of six years, today most of the members of BWRC are actively interested in the developments in 60 GHz and many have begun their own investigations. Many academic institutions worldwide are also actively researching mm-wave technology and many of these efforts are highlighted in this book. Research projects have now explored SiGe and CMOS for mm-wave frequencies beyond 100 GHz.

1.2 Why mm-Waves?

Based on Shannon’s Theorem, the maximum data-rate of a communication channel, known as channel capacity, C , is related to the frequency bandwidth of the channel, BW , and the signal-to-noise ratio, SNR , in the following manner [1]:

$$C = BW \cdot \log_2(1 + SNR) \quad (1.1)$$

which shows that one way to increase the communication data rate is to use more bandwidth. The information bearing signal is usually modulated around a carrier

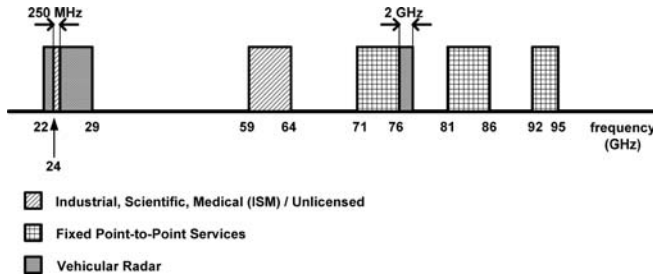


Fig. 1.3 The mm-wave band allocation in the United States.

frequency for proper propagation; therefore more bandwidth is available around higher carrier frequencies. For instance, optical fibers provide a low-loss and wide bandwidth medium, typically tens of GHz, around optical carrier frequencies of 250 THz for high speed wireline communications. The lower part of frequency spectrum is allocated to several narrowband applications such as radio, television, mobile phones, satellite communications, radio astronomy, wireless networking, and various military needs. The Federal Communication Commission (FCC) has allocated several frequency bands at millimeter waves for high data rate wireless communication. Figure 1.3 shows selected parts of the FCC-allocated frequency spectrum. Also shown in this figure are frequency allocations for automotive radar applications. The radar azimuth resolution (perpendicular to the radar wave) and range resolution (in the direction of radar wave) are inversely proportional with the carrier frequency and bandwidth, respectively, explaining the choice of such high frequencies and bandwidths. While the 22-29 GHz frequency band is allocated for short-range applications such as park assist, stop-and-go, and blind spot detection, the 77 GHz band is used for long-range automatic cruise control application.

The second term affecting the communication data rate as well as the radar maximum range is the overall SNR . Unfortunately, for a given distance, the received signal at higher carrier frequencies experiences more attenuation due to the following reasons: First, since the antenna size is normally inversely proportional to the carrier frequency; the higher the carrier frequency, the smaller the antenna size, resulting in less collected power. Higher absorption of air and most other material is the second cause of more signal attenuation at higher frequencies. In a multi-path environment, multiple replicates of the transmitted signal that are reflected from various objects reach the receiver at different times with different amplitudes and phases, causing unwanted signal fading. The amount of attenuation due to unwanted multi-path effects depends on the size of scattering objects relative to the carrier frequency as well as their type and location. A lower SNR reduces the data rate in communication systems for a given distance or reduces the range of wireless communication (e.g., in a radar application). Interference signals also have a noise-like behavior and reduce the received SNR . Fortunately, the larger attenuation at higher frequencies reduces the level of interferences as well as the multi-path components; the latter causes a

smaller delay spread, making it suitable for high speed wireless communication over a shorter range.

Over the years, a variety of modulation, coding, and diversity schemes have been employed to increase the SNR , hence increase the communication data rate. Spatial diversity, consisting of multiple antennas spaced apart, is one of the most attractive methods to increase the SNR . In a statistically fading environment, the received signal varies rapidly as the distance between the transmitter and receiver is altered by only a fraction of a wavelength. Spatially separated antennas can be used to extract more information from these large signal fluctuations, leading to an effectively higher data rate. They can also focus the signal energy into a narrow beam radiating into/from specific directions and place nulls in undesired directions. These beam forming schemes reduce the interference levels, increase the effective SNR , lower the transmitted signal energy to other directions, and are very suitable for radar and imaging applications. Similar to the receive antennas, in a statistically fading channel, transmitted signals from spatially separated transmitting antennas experience independent fades as they reach the receiver. By transmitting different signals from various antennas over time, a larger signal-to-noise ratio at the receiver can be achieved, increasing the overall capacity and performance. In addition to more available bandwidth at higher frequencies that results in higher capacity in data communication or better resolution in radar and imaging systems, the required antenna size and spacing in multiple-antenna systems are also reduced, making the system more compact and affordable.

1.3 The Birth of Silicon mm-Wave

Silicon technology has all but displaced GaAs and other technologies for RF applications in the low GHz regime. A few niche applications, such as power amplifiers, remain as a stronghold but are also under threat by several upstarts. For those with faith in Moore's Law, this was an inevitable consequence in scaling. Transistors became small enough, and consequently fast enough, to operate into the GHz frequencies, thus vying for countless communication applications in these frequency bands.

As the high-frequency capabilities of CMOS improve through scaling, the question is not *if*, but rather *when* will silicon become a viable alternative for mm-wave applications. To answer this question, it is fruitful to consider how silicon-based RF and microwave circuits came about. Even in the $0.25\ \mu\text{m}$ CMOS technology node, researchers and industrial start-up companies demonstrated low-cost radio solutions up to nearly 6 GHz for 802.11a applications. These demonstrations used relatively conventional circuit design and modeling techniques. And yet only ten years later, mm-wave silicon circuits have been widely demonstrated starting with the 130 nm technology node. Clearly the innovations to reach this speed did not come from scaling alone, since the process f_T has only increased by two-fold. Furthermore, increasing the operating frequency by a factor of 10 required a new design

methodology since the wavelength is on the order of the dimensions of the chip components (on-chip $\lambda = 2.5$ mm at 60 GHz).

Silicon was not the obvious choice for 60 GHz systems. Many non-silicon-based technology choices come to mind such as using GaAs MESFETs, PHEMT, InP HEMT, GaAs MHEMT, GaAs HBT, InP HBT. While these exotic technologies offer higher frequency of operation, they are expensive and have low manufacturing yields, and thus offer limited integration possibilities. Furthermore, these processes are not expected to scale in cost as rapidly as silicon (particularly CMOS) technology. Silicon technology enjoys steady scaling in large part due to the investments of industry and government which is tied to a healthy and vibrant multi-billion dollar market for digital, analog, and RF circuits.

Industry's choice to embrace Si rather than traditional III-V technologies was also a philosophical revolution. Engineers trained in the art and science of silicon analog design carried with them a design methodology based on compact models and SPICE-based transient computer simulation. The microwave community, on the other hand, relied on a rich theory based on two-port parameters and extensive measurements to characterize devices. Non-linear analysis was performed with harmonic balance. Active devices were often treated as black boxes with a given N -port response characterized by measurements or simple equivalent circuits. Which design methodology is most appropriate for emerging new applications in the microwave and mm-wave bands? In this book we present both design approaches and highlight where one is particularly more suitable or appropriate.

Economics was the ultimate factor in the decision to use silicon for applications up to 10 GHz. Silicon RF technology is relatively inexpensive, and in the case of CMOS, practically "free." The burden and major cost to develop silicon CMOS technology was provided by a large digital microprocessor market and analog and RF applications could leverage many of the innovations in the manufacturing of silicon circuits. On the other hand, early RF and analog designers using silicon and particularly CMOS were hampered by modeling problems, as the models were often extracted for digital applications. Today the situation is quite different and silicon is the technology of choice for many RF applications, including low-range low-cost radios, short-range high data-rate wireless LAN, cellular radios, and even power amplifiers. SiGe technology has given bipolar technology a second life, with advanced speed and low noise capability allowing first pass design success in demanding RF and video applications. Today communication equipment is a large and perhaps the largest fraction of the world's semiconductor manufacturing and thus RF simulation, compact models and design kits are more common. Design tools in the silicon oriented tool flows have improved considerably, allowing more accurate non-linear simulations using periodic steady-state and harmonic balance techniques. These same techniques allow more accurate prediction of noise in non-linear and time-varying circuits, particularly mixers and oscillators. Most commercial applications today require operation up to 6 GHz, but given that a plethora of new standards are emerging with applications up to 10 GHz for UWB, 24 GHz, and 60 GHz and beyond, there are many hurdles that need to be overcome in order to create the right design methodology for silicon-based circuits.

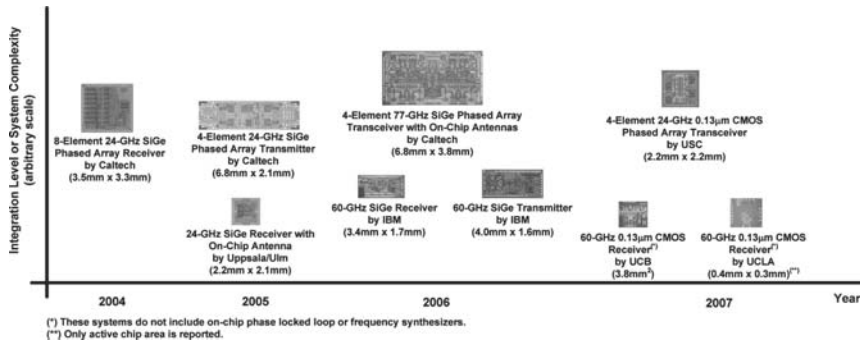


Fig. 1.4 Die photos of several published mm-wave systems integrated into silicon (from 2004-2007).

The main premise behind using silicon at millimeter waves is the higher level of integration offered at a high yield that leads into lower cost systems. Over the relatively short span of five years, several highly integrated and complex millimeter wave systems have been reported by academia and industrial research labs (Fig. 1.4)². These fully integrated chips consist of several thousand RF and digital transistors and on-chip passives in multi metal-layer silicon processes and include all the receiver, transmitter and even transceiver building blocks such as low noise amplifiers, mixers, voltage controlled oscillators, phase locked loops, power amplifiers, and in some cases on-chip antennas. Moreover, in many cases multiple receive and transmit paths are integrated in a single chip to realize fully integrated phased arrays. A few trends can be predicted for the near future. The first one is the realization of complete transceivers at 60 GHz and above using a standard CMOS technology. The second direction is realizing integrated CMOS beam-forming arrays at millimeter waves. The third trend is the incorporation of more functionality into millimeter wave transceivers such as mismatch, I/Q, and array calibration, modulation and demodulation of high constellation modulation formats, and possibly wideband channel equalizers. Finally, the issue of packaging and integration of antennas with silicon chips, especially in the case of arrays, is an open area for future research. Moving into higher frequencies, 94 GHz and the near THz region, are all extremely exciting research directions that can benefit not only the engineering and scientific community, but also several industries. Overall, we should anticipate an even higher level of research enthusiasm towards silicon-based integrated mm-wave systems in the future.

² This figure is intended to illustrate the advancements in the silicon integration of millimeter wave systems. It is not by any means comprehensive to all the published work in the area of silicon and CMOS based integrated millimeter wave systems.

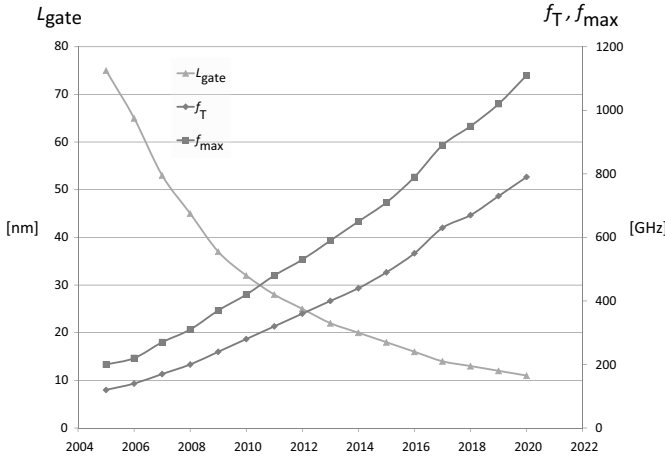


Fig. 1.5 The f_T/f_{max} trend with scaling for CMOS technology according to the ITRS 2006 [3].

1.3.1 Why CMOS?

As duly noted, CMOS technology is driven by a mass consumer market for high-speed digital microprocessors whereas specialized technologies are only needed in niche applications where the cost can be justified. For these reasons CMOS technology is a compelling choice, despite the added difficulties in doing mm-wave in a digital process.

A concerted effort by academia, industry, and worldwide research organizations has resulted in aggressive and steady technology scaling of silicon CMOS technology. This scaling was fueled by the demand for digital computation and memory and thus the technology has primarily evolved to serve these markets. The technology trend curve (f_T) shown in Fig. 1.5, shows the steady expected improvement in the operation speed of these devices. Even though specialized versions of bulk CMOS processes have been adapted to serve the communication sector for RF and microwave applications, the core devices are still digital in character. Silicon technology has been favored for scaling due to the simplicity of fabrication, which results in part from the native SiO_2 oxide and self-aligned gate formation. Scaling was initially guided by Dennard's Law [4], which called for constant field scaling, which in turn required lowering the supply voltage and also scaling the gate oxide thickness. Eventually constant field scaling was violated and high field effects, such as velocity saturation, led to additional innovations such as non-uniform doping (halo implants) and lightly doped drain contacts extensions (LDD). Despite many dire predictions about the scaling of bulk CMOS technology, the shrinking has continued down to the 45 nm technology node.

But today's CMOS is not "pure" in the traditional sense as the technology has evolved from requiring a handful of materials (Si, O, Al, and a few dopants and

rare earth metals) fabricated using optical lithographic techniques to a sophisticated process using a wide variety of materials, high-K gate dielectrics, metal gates, low-K interconnect dielectrics, seven or more metal layers incorporating Cu, and various technologies that introduce strain into the channel for enhanced mobility (SiGe drain/source or capping layers). Today devices with $L_g < 32\text{nm}$ (sub optical) are routinely fabricated for high volume production. Beyond this gate length, new device structures have been proposed, such as multi-gate transistors (FinFET) and SOI transistors. In this book we limit our focus to the bulk CMOS transistor with experimental results down to the 90nm node. These devices are commercially available, relatively low cost in volume production, and perhaps represent the “sweet spot” for operating in the mm-wave band.

The technology scaling of the channel gate length L_g has resulted in raw performance benefits, particularly in the speed of the transistor. But a big penalty has been paid in other performance metrics, in particular the device output conductance g_o which translates into low intrinsic gain ($A_v = g_m/g_o \sim 10$) and extremely low supply voltages. Fortunately the supply scaling has stopped at 1V, but even at this level there are severe limitations in the achievable dynamic range of amplifiers and other key building blocks. Naturally this also translates into poor output power capability. Finally, in all technology nodes the bulk substrate is conductive which results in higher losses in passive components such as inductors and transmission lines. Fortunately, in many mixed-mode process nodes the substrate is only moderately conductive ($\rho \sim 10\Omega\text{-cm}$) to minimize substrate coupling, which is quite tolerable in many applications. Despite these shortcomings, CMOS technology has transitioned from a low performance digital process to the prevalent technology for consumer RF applications. The low gate leakage has also been a key for realizing switched capacitor analog circuits, making CMOS the technology of choice for mixed-signal circuits. In the RF regime CMOS has proved viable as a low cost alternative to SiGe and GaAs for consumer applications up to 5 GHz. Niche applications such as power amplifiers still favor technologies with higher breakdown voltages but cost considerations often dominate performance in consumer applications. For example, in many cost sensitive products, given the choice between two power amplifiers, the difference between the cost will lead to the adoption of a CMOS or silicon PA with considerably lower efficiency.

As Moore’s Law has recently pushed the f_t and f_{max} of CMOS transistors above 100 GHz, an all-CMOS solution at 60 GHz is feasible. It should be noted that the major reason behind Moore’s Law and CMOS technology scaling is a reduction of the cost per function. However, for more than forty years, CMOS technology scaling has also given us a speed improvement, leading to mm-wave CMOS designs. But the die area of a mm-wave chip is dominated by passive devices rather than active devices, so there is no inherent area advantage in using scaled technology nodes. And while Moore’s Law has enabled CMOS mm-wave circuit design, CMOS is by far not optimized for mm-wave performance. Although the success of RF-CMOS products has steered the focus of CMOS technology to consider RF design, CMOS remains in the first place a digital technology with a low supply voltage (due to breakdown), and a lossy substrate. Furthermore, operating in the 90nm node means operating

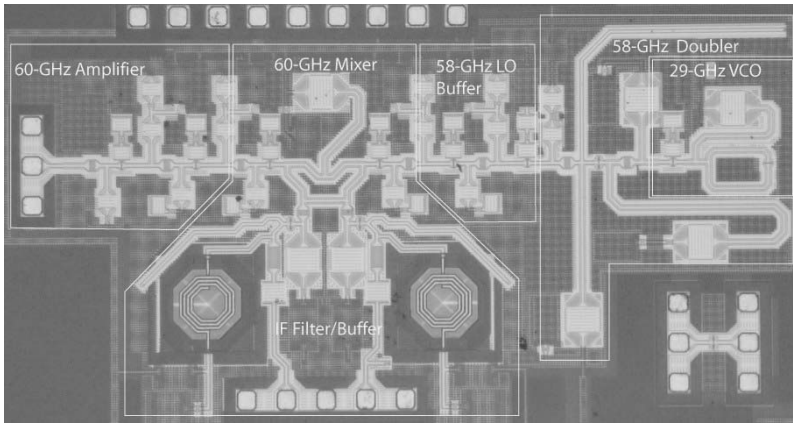


Fig. 1.6 A front-end 60 GHz receiver realized in a CMOS 130nm process technology node.

relatively close to the limits of activity ($f_T \sim 100$ GHz, $f_{max} \sim 200$ GHz). Therefore, circuit techniques are mandatory to achieve acceptable mm-wave performance from a CMOS technology.

While today's silicon technology is capable of operating at 60 GHz and beyond, there is still incentive to continue to use newer scaled technology since smaller transistors will provide higher performance at constant or even lower power levels. But this begs the question as to the true cost of CMOS technology, in particular if the die area continues to be dominated by the passive components.

1.3.2 True Cost of Silicon mm-Wave

It is important to briefly justify silicon technology for mm-wave applications based on economic arguments. It is now clear that the technology is capable of operating in these frequencies, but is there really a cost advantage? In particular, we have already noted that the die area of a mm-wave chip is dominated by passive devices. This is clearly visible in the die photo shown in Fig. 1.6, where the transmission lines and capacitors occupy most of the area. In this example, scaling from 130nm to 90nm would not reduce the die area noticeably. Also, the fabrication costs of silicon technology below 100nm are prohibitively high due to the mask costs, which means that silicon is only a viable solution in volume applications, where mask costs are shared over millions of parts. This also means that there should be additional investments in the modeling and tools of these devices to ensure "first pass" success in the fabrication, since any errors in the design will translate into millions of dollars of lost revenue for new masks. These problems are not new and are faced by RF and microwave circuit designers in the cellular and WiFi markets. In particular, the use of SiGe over CMOS is sometimes justified in large part because of more

reliable models and tools rather than performance benefits. Moreover, since the cost of older generation lithography tools can produce a SiGe transistor on par with a more expensive cutting-edge CMOS process, the cost difference between CMOS and SiGe is somewhat artificial. Advanced packaging makes multi-chip solutions viable and attractive. But if the entire design can be done in CMOS, then there is clearly a benefit in offering a smaller footprint product. Furthermore, the investment in modeling and designing in a cutting-edge CMOS process will pay back in the long run, since future products can leverage this knowledge and these skills as the technology cost reduces over the years.

Die costs are only a fraction of the overall cost to produce integrated circuits. Costs are often dominated by packaging and testing parts. This is particularly important for RF circuits, which require expensive test equipment and longer testing to validate functionality. Special packages with low inductance are also needed in power amplifier and ultra high frequency applications, making the die cost in some cases negligible in the overall equation. But this situation changed quickly when CMOS and SiGe designers learned to leverage automatic tuning and digital calibration techniques, particularly with the aid of advanced digital signal processing available either for “free” (shared with digital baseband) or at marginally higher costs. Today many aspects of an RF transceiver are tuned automatically, allowing lower performance technologies such as digital CMOS to compete effectively with more expensive SiGe or higher performance processes. Examples include I/Q gain/phase matching, filter passband tuning, and VCO/PLL frequency calibration. Many of these settings are done “live” to compensate for temperature and lifetime aging variations. Fully integrated transceivers are tested in loop-back mode to avoid expensive RF test equipment. With only DC probing, a transceiver chip can be tested from “bits *in* to RF *out* and then to bits *out*” from the baseband DAC to the baseband analog and RF transmitter to the RF front-end receiver blocks all the way back through the receiver analog and baseband sections. This end-to-end functionality is tested with the aid of a digital baseband.

Communication algorithms today make intensive use of signal processing to deal with multi-path fading in increasingly broader bandwidths of spectrum. The use of OFDM has become common in many existing WiFi chips (802.11a/g) and is proposed for future 4G cellular networks and short range personal area networks (OFDM-UWB adopted by Zigbee and Bluetooth). Furthermore, sophisticated Turbo codes and low-density parity codes (LDPC) are used and proposed in future systems. This means that end-to-end testing will require a very powerful DSP engine to assist in the testing, which make it more difficult to perform this level of testing in a multi-chip solution. If testing has to be performed after packaging, the yield will suffer since a faulty RF section means that the digital baseband is wasted or vice versa.

As we look at the emerging mm-wave applications, it is clear that CMOS and SiGe will only offer an advantage over traditional mm-wave technologies if packaging and testing costs can be reduced. The same approach to testing a 5 GHz part can be applied to a fully integrated 60 GHz transceiver. And while low cost packaging incurs a heavy penalty in routing signals on and off the chip, about a 1dB of loss in transmit power and 1dB higher noise figure, many low to medium performance applications will

tolerate this inefficiency in favor of lower costs. Low cost packaging has already been demonstrated for mm-wave systems [5] and many existing testing procedures for RF applications can be adopted for mm-wave. In fact, a mm-wave transceiver can be viewed as nothing but an RF transceiver with a simple front-end to translate the signals from say 60 GHz down to 1 GHz where traditional signal processing techniques prevail. Of course there are new challenges, such as the processing of extremely wideband signals (compare 2 GHz bandwidth for a 60 GHz WLAN to 200 kHz bandwidth of a cellular system). Also the incorporation of multiple antennas in mm-wave systems necessitates innovative and inexpensive ways of testing mm-wave components, especially for radar and antenna array applications. One should not doubt that these innovations will come and the ultimate cost of mm-wave silicon will reduce dramatically, enabling many new and exciting applications.

1.4 Communication in the 60 GHz Band

Many applications require or benefit from high data rates far exceeding the capability of existing WiFi and UWB technology. High quality video signals require data rates exceeding several Gb/s. This is because sending uncompressed data greatly reduces power overhead for encoding and decoding video. Set-top boxes and digital video cameras are obvious applications for this technology. In general, the need for bandwidth is insatiable, much like the demand for CPU speed, static and dynamic RAM, flash memory, and external hard disk capacity. While new spectrum is available in the low-GHz bands, these bands are likely to be overly congested in the near future. Moving up to higher frequency also provides natural isolation from fast switching digital circuitry typical of today's microprocessors, already operating at several GHz clock speed. Furthermore, the only way to extract more information from a fixed bandwidth at lower frequencies is the application of more complicated modulation schemes. To extract 1 Gb/s from 100 MHz of bandwidth obviously requires 10 bits per Hz, but only 1 bit per Hz from a 60 GHz solution with 1 GHz bandwidth. The 60 GHz system uses a relatively narrowband signal, and low order constellations can be used to transmit and receive the data. The lower GHz system, on the other hand, must use sophisticated signal modulation, often placing stringent demands on the phase noise and power amplifier linearity (particularly for OFDM), and this translates into a system with less overall sensitivity. Much energy must be consumed in the baseband of these systems to provide FFT and equalization functionality, which will end up consuming more energy per bit than a mm-wave solution, despite the higher power consumption of the front-end blocks at 60 GHz.

The key motivation for the exploration of the 60 GHz spectrum is the availability of 5–7 GHz of unlicensed bandwidth, with numerous obvious advantages and applications. This is not only true in the US but also globally. There is thus great incentive to design wireless systems capable of exploiting this bandwidth. In July 2003, the IEEE 802.15.3 working group for WPAN began investigating the use of the 7 GHz of unlicensed spectrum around 60 GHz as an alternate physical layer

(PHY) to enable very high data-rate applications such as high-speed internet access, streaming content downloads, and wireless data bus for cable replacement [1]. The targeted data-rate for these applications is greater than 2 Gb/s. Although the excessively high path loss at 60 GHz, due to oxygen absorption, precludes communications over distances greater than a few kms, short-range WPANs actually benefit from the attenuation, which provides extra spatial isolation and higher implicit security. Furthermore, due to the oxygen absorption, the FCC regulations allow for up to 40-dBm EIRP transmit power, which is significantly higher than what is available for the other WLAN/WPAN standards. The wide bandwidth and high allowable transmit power at 60 GHz enable multi-Gb/s wireless transmission over typical indoor distances (~ 10 m). Moving to higher frequencies also reduces the form factor of the antennas, as antenna dimensions are inversely proportional to carrier frequency. Therefore, for a fixed area, more antennas can be used, and the antenna array can increase the antenna gain and help direct the electromagnetic energy to the intended target. As previously noted, the directive antenna pattern improves the channel multipath profile by limiting the spatial extent of the transmitting and receiving antenna patterns to the dominant transmission path. Consequently, the delay spread and Rician K-factor of an indoor wireless channel can be significantly improved. However, directivity is both a blessing and a curse. Omni-directional antennas simplify the system design for point-to-multipoint systems, whereas directive antennas are best suited for point-to-point applications. A system employing antenna arrays with adaptive electronically-steerable beams will allow for mobility and ease the setup of the device compared to one using a fixed high-gain directional antenna.

1.4.1 Beam Forming

First to see that high antenna gain is a necessity in a mm-wave link, consider the simple link budget shown in Table 1.1, where we assume that we wish to communicate over a 1 GHz bandwidth given a transmit power of 10 dBm and a receiver noise figure of 10 dB, both reasonable numbers for silicon technology. The receiver noise floor is already -74 dBm for 1 GHz of bandwidth ($kTBF$), which means that we can only tolerate a path loss of about 74 dB if we assume a required SNR of 10 dB ($P_{TX}\mathcal{L} > kTBF$). This is a very small range for an omni-directional antenna, even for line of sight propagation (about a meter). As evident in the link budget calculation, there is no link margin under the assumption of 10dB shadowing loss.

To overcome the challenge of high path loss, we must either boost the transmit power, very difficult to achieve in a low voltage process at mm-wave frequencies, or use a different antenna technology. The high path loss results from the small capture area of an antenna at 60 GHz. Assuming that the radiation flows omni-directionally, then a “ball” of radius R has an average energy density of $P_t/(4\pi R^2)$, which is simply the transmitted power divided by the surface area of the ball. We know that an antenna’s effective cross section or capture area is proportional to its area. For a simple dipole or loop antenna, this is on the order of λ^2 , which means that operating

Table 1.1 Link budget for 1 Gp/s 60 GHz wireless system at 1m communication distance.

Component	Contrib.	Running Total	Comment
Tx power	+10 dBm	+10 dBm signal	PA at P_{-1dB}
Tx Ant Gain	+2 dB	+12 dBm signal power	
Path loss	-68 dB	-56 dBm signal	path loss at 1 m
Shadowing loss	-10 dB	-66 dBm signal	
Rx Ant Gain	+2 dB	-64 dBm signal	
Background noise	-174 dBm/Hz	-174 dBm/Hz noise	kT at room temp
Noise BW	+90 dB	-84 dBm noise	1 GHz noise BW
Noise figure	+10 dB	-74 dBm noise	NF of receiver
SNR at input		10 dB	Signal power / noise power
SNR required		10 dB	Coherent FSK
System Margin		0 dB	BER = 10^{-3}

at higher frequencies will incur a penalty if we use the smallest possible antennas. This result also comes using Friis' equation, the ratio of the power received P_r to the power transmitted P_t for a line-of-sight link at a distance r from the source is given by

$$\frac{P_r}{P_t} = \frac{D_1 D_2 \lambda^2}{(4\pi r)^2}$$

Here we account for the antenna directivity at the transmitter and receiver. We know for simple dipoles, the antenna directivity factor is on the order of unity and fixed for a simple antenna structure. The path loss is therefore proportional to the wavelength λ squared. This limits the range of a wireless communication system, especially when we consider the poorer noise performance of silicon at mm-wave and the limited output power capabilities. Highly directive antennas can be used to decrease the path loss, say in the form of dishes or horns, since these structures have larger capture areas. But these options are not suitable or appropriate for consumer applications. A more promising and exciting alternative is an antenna array.

A different perspective comes about if we consider that any consumer application will limit the antenna footprint to a fixed area. For a fixed antenna aperture size A , the directivity is simply

$$D = \frac{4\pi A}{\lambda^2}$$

and it can be seen that there is actually an improvement in the received power by moving to higher frequencies for a fixed antenna form factor. For example, a 60 GHz system with a 16-element antenna array occupies about the same area as a dipole antenna at 5 GHz, while providing much higher directivity.

Antenna arrays, shown schematically in Fig. 1.7, are a collection of antennas that can be used to cover a larger physical area. If variable gain and electronic phase

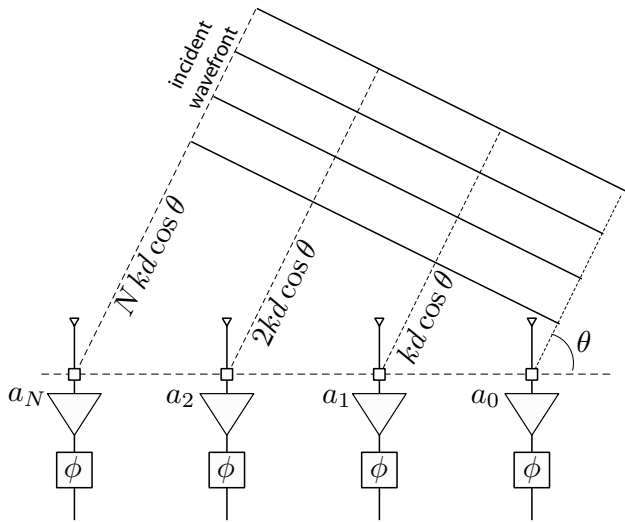


Fig. 1.7 A generic beam forming system employs arrays of antennas and transceivers to boost gain, power, and sensitivity.

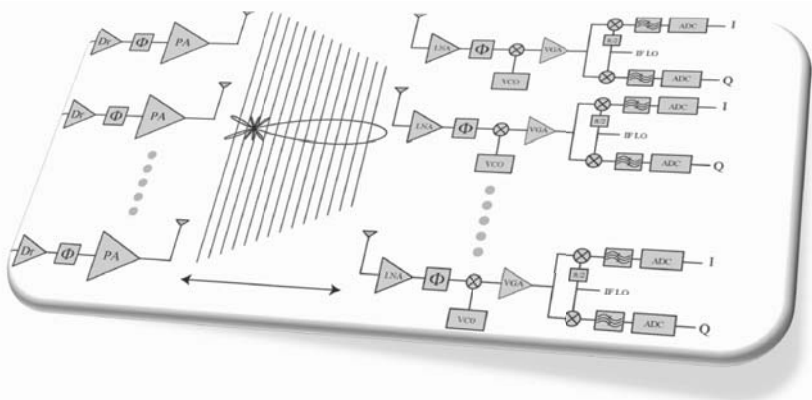


Fig. 1.8 A multiple antenna transceiver architecture employing antenna steering for maximum gain.

shifters (or time delay elements) are also incorporated into the array, then any antenna pattern can be devised, only limited by the number of antennas and a discrete Fourier transform (over the spatial dimension). A particularly important case is a “beam” pattern which is synthesized by adding enough phase shift at each element to receive an incident signal at an angle of θ coherently. Due to the small wavelength, antenna arrays can be realized in the area of the package or printed circuit board. A generic transceiver, therefore, takes the form of a multi-antenna and multi-transceiver array shown in Fig. 1.8. Beam forming improves the antenna gain while also providing spatial diversity and thus resilience to multi-path fading. The main benefit of the

Table 1.2 Link budget for phased-array 1 Gp/s 60 GHz wireless system at 1m distance.

Component	Contrib.	Running Total	Comment
Tx power	+10 dBm	+10 dBm signal	PAs at P_{-1dB}
Tx Ant Gain	+2 dB	+12 dBm signal power	
Array Gain	+9 dB	+21 dBm signal power	8-fold antenna array
Path loss	-68 dB	-47 dBm signal	path loss at 10 m
Shadowing loss	-10 dB	-57 dBm signal	
Rx Ant Gain	+2 dB	-55 dBm signal	
Array Gain	+9 dB	-46 dBm signal	8-fold antenna array
Background noise	-174 dBm/Hz	-174 dBm/Hz noise	kT at room temp
Noise BW	+90 dB	-84 dBm noise	1 GHz noise BW
Noise figure	+10 dB	-74 dBm noise	Projected NF
SNR at input		28 dB	Signal power / noise power
SNR required		10 dB	Coherent FSK
System Margin		18 dB	BER = 10^{-3}

multi-antenna architecture used here is the increased gain that the directional antenna array pattern provides. This gain is needed in order to support data rates approaching 1–10 Gb/s at typical indoor distances. Such an antenna array also allows spatial power combining which greatly simplifies the design of the transmitter. Automatic power control can be realized with no additional burden or efficiency loss by simply turning off some of the transmit paths.

To realize an antenna array, the second major challenge is providing low-cost circuit building blocks to realize the transceivers. Since multiple transceivers are potentially needed in many architectures, the power consumption of these blocks is important. Since package and board parasitics are prohibitive at 60 GHz, these transceivers should be fully integrated solutions. Given the lossy substrate is the limiting factor in low-GHz designs, it may seem that scaling to higher frequencies should be even more difficult, but in fact we shall show that relatively good passive devices are possible, even in mm-wave frequencies using inexpensive digital CMOS technology.

Table 1.2 is the link budget for a 60 GHz system communicating with 1 GHz of bandwidth. As already noted, the path loss is the dominant factor, requiring a directive antenna array to compensate for the loss of gain. Also note that the large input bandwidth increases the noise floor considerably compared to a narrowband system. Under similar conditions as the single transceiver, we see that the array allows for a substantial increase in the link margin, which allows a longer communication range or increased robustness over multi-path fading.

The third major challenge is the realization of a baseband system to filter, digitize, equalize, and detect wideband signals marred by high path loss and multi-path fading. The 802.15.3c committee has identified several channel model scenarios including

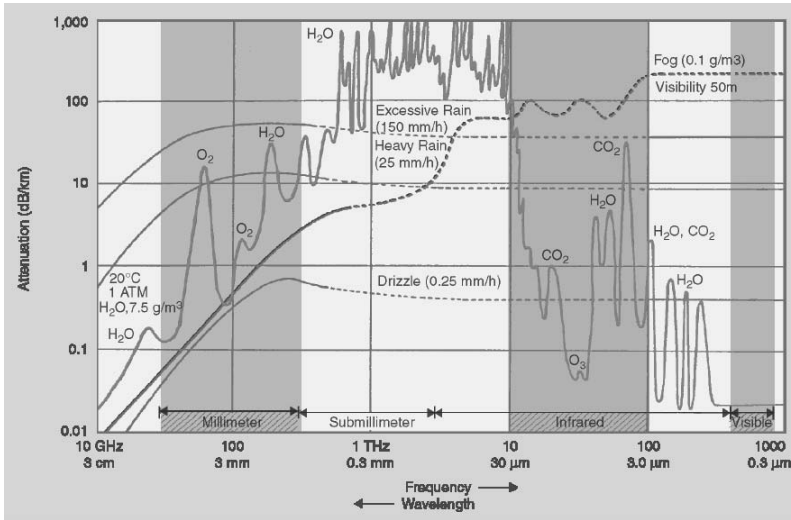


Fig. 1.9 The propagation attenuation characteristics [dB/km] versus frequency (wavelength) for Earth's atmosphere under various conditions [7] (© IEEE 2003).

indoor and outdoor, line of sight (LOS) and non-LOS (NLOS), and many variations including desktop, office, and media environments. In high data rate systems, the choice of baseband architecture can have a large impact on the overall system complexity and power consumption of the mobile transceiver. In typical “mostly digital” wireless receivers or transmitters, high resolution interface circuitry (ADCs and DACs) are required to convert the signal waveform between the analog and digital domain so that the subsequent or preceding digital processing is not limited by accuracy of these interface circuits. In multiple-antenna systems, where there may be several instantiations of the baseband circuitry, the aggregate power consumed by these high-speed, high-resolution interface circuits may become prohibitively large. It is therefore important to draw the boundary between digital and analog carefully. Similar to high speed links, analog equalization can help mitigate multi-path effects, lowering the required resolution of the ADCs. Digital estimation and control of the analog circuitry is important in realizing a robust and adaptive system [6].

1.5 Unique mm-Wave Applications

1.5.1 mm-Wave Spectrum

The availability of inexpensive small footprint silicon mm-wave transceivers leads to the possibility of higher complexity mm-wave systems incorporating dense arrays of transceivers incorporating sophisticated multi-antenna signal processing. With the

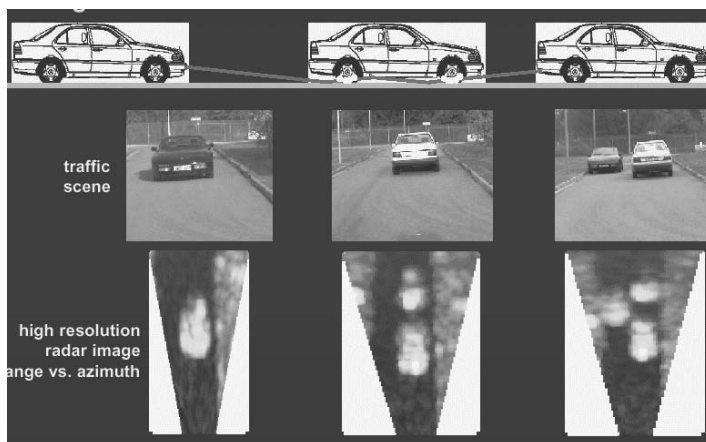


Fig. 1.10 Images produced by an automotive radar system at 77 GHz. [8] (© IEEE 2002)

continued scaling of nanoscale CMOS technology and the resulting increase in intrinsic device f_T , there are new opportunities for circuitry operating above 60 GHz. New potential applications include mm-wave imaging and sub-THz chemical detectors, with applications in astronomy, chemistry, physics, medicine, and security. Important frequencies of interest will include 90 GHz, 140 GHz, and frequencies above 300 GHz or in the so called THz regime. The reason to focus on these frequencies is evident in Fig. 1.9, which shows the attenuation of signal propagation in air for various frequencies. Clearly there are windows of opportunities where the attenuation is either minimal or maximal. The 60 GHz band is particularly lossy due to oxygen absorption, making it suitable for short range networks with good spatial isolation. But other bands such as 90 GHz are ideal for long range imaging. Compare the performance of 90 GHz to the visible spectrum and you can see why plants and animals evolved to use this spectrum for energy scavenging and imaging. But also compare the attenuation in the case of fog, and you can see that while a mm-wave system can outperform an optical system under (by definition) low visibility conditions³.

1.5.2 Automotive Radar

An important application for imaging is the automotive radar operating at 24 GHz and 77 GHz. Today only luxury automobiles are equipped with mm-wave radar technology. This can aid in driving in low visibility conditions, especially in fog, and in automatic cruise control and even automated driving on a future freeway [9]. In the U.S. 42,000 people die annually as a result of automotive accidents, while

³ That is low visibility for animals that rely on the optical spectrum for their visual needs.

1.5 million people are injured, making radar technology very attractive to realize in every automobile. Advanced video imaging, radar, GPS, and gyroscopes can be used in conjunction to give the driver a much richer set of data and ultimately a better and safer driving experience. An example image captured by a radar is shown in Fig. 1.10, where the visual image is compared to the radar image. It's clear that the radar image can look under vehicles and provide even more data than possible otherwise, especially with the abundance of larger passenger vehicles (SUVs and mini-vans) on today's roads. The key development has been the dramatic reduction in the area, cost, and power to perform the related signal processing for an automotive radar system.

Of the more than 42,000 fatalities caused by automobiles every year, about 5,000 are pedestrian fatalities. Every year, there are over 3,000,000 crashes, 1,500,000 injuries, and 9100 fatalities that occur at U.S. intersections, costing us about \$124 billion. This is one-third of all crashes and 17% of all highway fatalities. Recent accident statistics indicate that $\sim 56\%$ of all crashes in 2005 occurred at or less than 40 mph speeds. In addition, while crashes remained flat, there was an increase in the accidents suffered by pedestrians and cyclists. Sensors that detect objects in near ranges (approximately 30 meters with large field of view) will enable many active safety applications that can significantly reduce crashes at these speeds. In Fig. 1.11 we see a host of sensors that provide object detection and parking assist, side-impact pre-crash detection, blind spot assistance, and long range object detection capability. Each sensor requires a different and custom tailored technology to meet the required specifications.

One promising technology for short range vehicular sensing applications is radar. Compared to lidar and video technologies, radar is an all-weather sensor with a sufficient resolution ($< 5\text{cm}$) and operates in real time. In a typical radar, a radio frequency (RF) signal, usually in the form of a time-limited pulse, is transmitted towards the target(s) of interest. Information regarding the shape, distance, and speed of the target(s) is embedded in the arrival time and shape of the reflected or scattered signal(s). The radar azimuth resolution is inversely proportional to the carrier frequency. This is due to the smaller beam size of a given radiating or receiving aperture at higher frequencies. Therefore, the FCC has allocated the 76-77 GHz frequency band for long range (100m) automatic cruise control (ACC) automotive radar applications. These radars are currently realized using compound semiconductor technologies and limited to higher end cars. Radar range resolution is inversely proportional to the bandwidth of the transmitted pulse. Therefore, the FCC has allocated a wide frequency spectrum around 24 GHz (22-29 GHz) for short range automotive radar applications. Many other countries including the European Union have also approved, albeit temporarily, this frequency band for commercial vehicular use. The FCC allocated frequency band allows using the ultra wideband (UWB) technology to achieve a higher resolution for short range vehicular sensing applications such as blind spot detection, side and rear impact sensing, blind spot detection, and stop-and-go. One desirable objective is pedestrian detection and protection. The FCC defines ultra wideband (UWB) as signals having more than 500 MHz of instantaneous bandwidth or exceeding 20% fractional bandwidth.

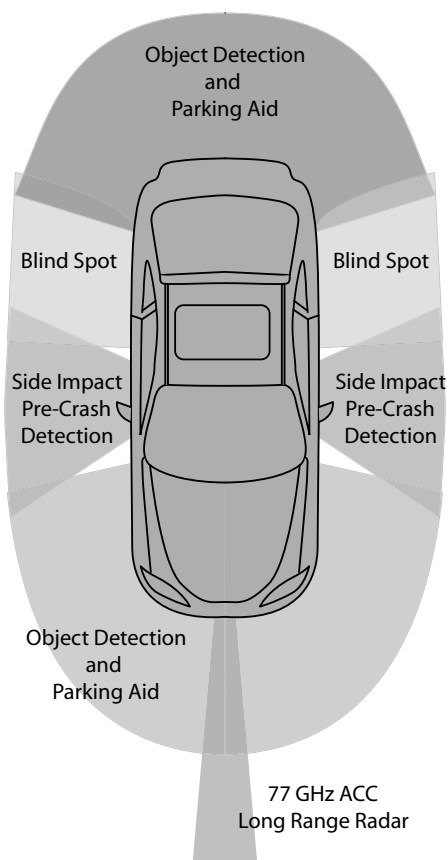


Fig. 1.11 Top view of an automobile and the desired sensors.

For instance, an UWB radar that utilizes pulses with 3 GHz of bandwidth within the FCC allocated 22-29 GHz frequency spectrum for automotive radar applications will have a range resolution of 5 cm. As the vision of the automotive industry is to incorporate several of such short-range sensors around all cars to provide 360° awareness for the driver, cost and power consumption of these sensors are crucial. Commercial SiGe UWB short-range radars at 24 GHz have already entered the market. It should be only a matter of time before CMOS versions become industry standards.

One critical system metric in all radars is the rate of successful detection and false alarms. This is even more important in automotive radars since the data collected from multiple sensors will be used to assist the driver and in some cases take the control away from the driver in critical situations. False alarms are common in typical driving situations that involve metallic bridges, ground holes, underground pipes, and parking structures. The sensor sensitivity suffers even more in a dynamic driving environment with several stationary and moving objects. The primary reason for the high rate of false alarms is the wide field of view of sensors; the radar transmitted pulse

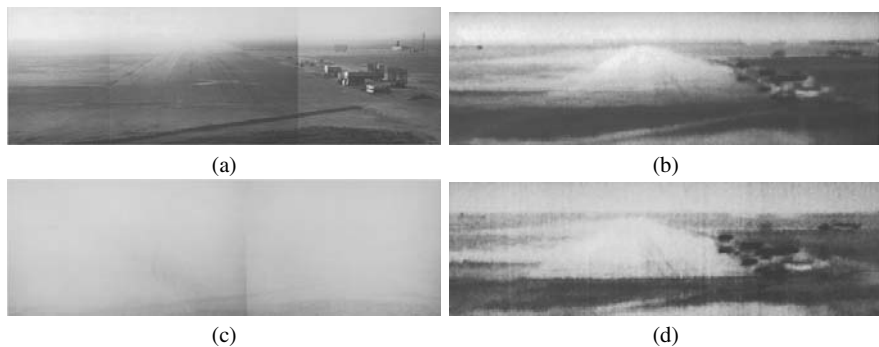


Fig. 1.12 (a) Images of an airport runway produced by an (a) optical and (b) passive mm-wave imager. Same scene under low visibility (fog) in (c) and (d) [7] (© IEEE 2003).

reaches and reflects off all the objects that are within the field of view. Therefore, the radar receives numerous pulses with varying amplitudes and shapes at different times. Advanced signal processing algorithms should be used to discriminate various targets. However, even the most advanced algorithms that are known are limited in their ability to reliably identify and track all objects in real time. Scanning radars with a limited instantaneous field of view can greatly mitigate this problem and are commonly used in existing military systems. Mechanically scanning radars are not suitable for vehicular systems due to the slow scanning rate, higher cost, difficulty in automotive sensor integration, and higher failure rate due to mechanical weariness and damages.

Phased array is a well-known technique that allow for electronic steering of the electromagnetic wave. A phased array imitates the behavior of a mechanically scanned antenna whose bearing can be adjusted electronically. A phased array transmitter can form a narrow beam that can be steered towards intended angles in a narrow field of view. A phased array receiver is only sensitive to the reflected signals that arrive at a specific angle that can be electronically controlled. Therefore, phased arrays, also known as electronically scanned arrays or beam-forming arrays, reduce the undesired effect of multiple reflections and interference while allowing for a full spatial coverage. In addition, phased array receivers enhance the radar SNR and hence its range. RF phased arrays have been extensively used in military applications for high resolution ground- and ship-based as well as airborne radars. Silicon technology, CMOS in particular, allows for the integration of an entire phased array radar with a typical size of 4 to 16-elements in a single chip. Silicon integration also improves the radar robustness and reliability through on-chip signal processing and calibration that is offered at little incremental cost. It is likely that RF and electrical engineers will play a central role in the safety of future road transportation systems.



Fig. 1.13 A passive mm-wave image can easily denude a person to reveal a hidden weapon. See optical image for a comparison [7] (© IEEE 2003).

1.5.3 mm-Wave Imaging for Medical Applications

Another potential application for mm-wave technology is passive mm-wave imaging. By detecting only the natural thermal radiation of objects in the mm-wave band, images of objects can be formed in a very similar fashion as in an optical system. Either a group of receivers or a movable mechanical antenna is required to scan the area of interest. Unless special techniques are employed, due to the relatively large wavelength, the resolution of this approach is limited to objects on the order of a *mm*. But this resolution is sufficient for many applications. In Fig. 1.12 we see the optical and mm-wave image seen from an airplane in good and bad visibility conditions. The mm-wave image is clearly able to penetrate through the fog and rain and provide a clear image. In security applications, passive (or active) mm-wave images of a person can be used to find hidden weapons. In Fig. 1.13 we see that a hidden gun or knife is easily visible due to the difference in emissivity of the body and the object. While metal objects are easy to detect by other means, non-metallic objects are also visible, which is a great advantage of the technology. Unlike X-ray based imaging systems, which can only be used with limited dosage with living organisms, passive mm-wave imaging does not use any additional radiation than what is naturally present. Even active imaging systems use photos with milli-eV energies compared to k-eV necessary for X-ray systems.

Other emerging applications for mm-wave technology include medical imaging for tumor detection, temperature measurements, blood flow and water/oxygen content measurements. These applications were under intense exploration in the past two decades but much of the research has discontinued due to the fact that these traditional systems were not able to compete with existing MRI or X-ray CAT scan systems. Due to the much larger wavelength, these systems have relatively poor resolution. As silicon technology allows larger arrays of transceivers to be realized in a small area at a low cost, we believe that many of these applications will re-emerge due to the immense potential for size and cost reduction. Furthermore, as we push into higher frequencies above 100 GHz, the wavelength becomes smaller and new application domains emerge. In particular, as we pass 300 GHz and enter the

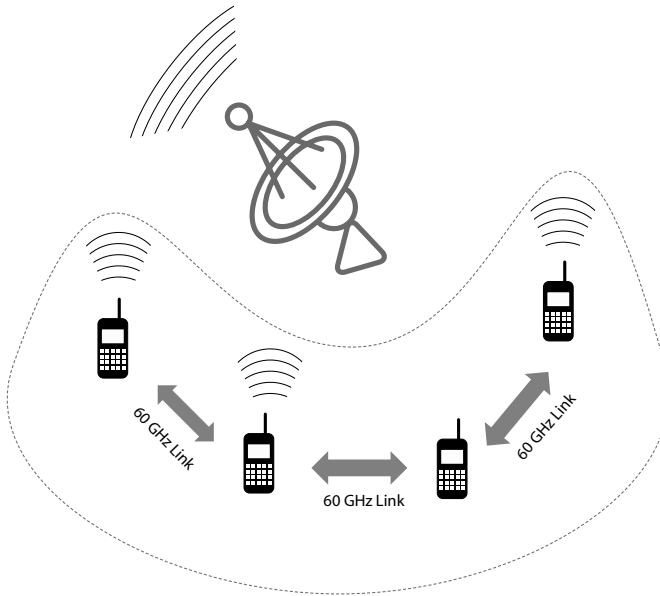


Fig. 1.14 A mm-wave collaborative distributed MIMO system.

THz regime, there are numerous applications in chemical spectroscopy due to THz vibrational resonances in molecules. One of the outstanding problems in the THz community is the lack of a sufficiently powerful transmitter. Silicon technology may offer a low power and efficient transmitter (1 mW) in the form of an oscillator and power amplifier which work directly at THz frequencies. This is in contrast to today's optical systems which use non-linear effects in crystals to produce THz emissions as a by-product.

1.5.4 Collaborative Distributed MIMO

With the wide availability of mm-wave technology, we have to fundamentally rethink the way we do communications. For instance, if a short range mm-wave link can be established with under 50 mW of power with a data rate of 2 Gb/s, then the energy consumption of such a system is about 25 pJ/bit, which compares favorably to a cellular or WiFi system consuming about 10-100 nJ/bit. Moreover, if we imagine a scenario where many portable devices with mm-wave capability are within a room or within a vicinity large enough to explore the multi-path diversity at lower frequencies, then the mm-wave links can be used to set up a collaborative and distributed MIMO system shown in Fig. 1.14. A single user forwards the data traffic to nearby users using a high bandwidth mm-wave link and then each user transmits the

same (or encoded) data to the recipient. Effectively the users together form a large aperture MIMO radio. Spatial coding is used to increase the capacity of the system greatly over a signal antenna system. This is better than a MIMO system within one device (such as 802.11n) due to the size constraints and the number of antennas one can fit into a handset footprint. Also, it is much more likely that distant radios will have higher diversity in their spatial links, allowing one to use the spectrum in an opportunistic fashion.

1.6 Overview of Book

This book develops the science and art of silicon-based mm-wave design beginning with a review of CMOS and SiGe technology considerations in Chapter 2. Next, Chapter 3 covers details of mm-wave active and passive modeling, with an emphasis on CMOS active technology. Amplifier and frequency conversion building blocks are covered in Chapter 4 in both CMOS and SiGe technology, with many examples from recent publications in conferences and journals. The design of power combining and power amplifiers is covered in Chapter 5. Voltage controlled oscillators (VCOs), frequency dividers, and other key frequency synthesizer building blocks are covered in Chapter 6. Using these foundational building blocks, the book covers system aspects of mm-wave silicon design by addressing the realization of a beam-forming array in Chapter 7.

References

1. C. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, pp. 379-423, 623-656, July-October 1948.
2. S. Emami, C. H. Doan, A. M. Niknejad, and R. W. Brodersen, "Design of CMOS for 60 GHz applications," *ISSCC Digest of Technical Papers*, 2004, pp. 440 - 538.
3. <http://www.itrs.net/>.
4. R. Dennard, et al., "Design of ion-implanted MOSFETs with very small physical dimensions," *IEEE Journal of Solid State Circuits*, vol. SC-9, no. 5, pp. 256-268, Oct. 1974.
5. http://domino.research.ibm.com/comm/research_projects.nsf/pages/mmwave.sixtygig.html
6. D. Sobel, *Berkeley Wireless Research Center Retreat*, <http://bwrc.eecs.berkeley.edu>.
7. L. Yujiri, M. Shoucri, P. Moffa, "Passvie mm-Wave Imaging," *IEEE Microwave Magazine*, vol. 4, issue 3, pp. 39-50, Sept. 2003.
8. "Automotive Radar and Prospective Circuit/Antenna Technologies - From Car Collision Avoidance to Autonomous Driving," *IMS 2002 Workshop*, <http://www.mtt.org/symposia/ims/2002/workshops.html>.
9. <http://www.path.berkeley.edu/>

Chapter 2

Silicon Technologies to Address mm-Wave Solutions

Andreia Cathelin and John J. Pekarik

2.1 Why Silicon?

There are strong reasons not to consider silicon technologies for mm-wave applications. Silicon comes up short in many comparisons to III-V semiconductors. Silicon carrier mobility is relatively low and so device-level FOMs of raw performance appear to be inferior. The silicon bandgap is relatively small and so voltage tolerance tends to be lower. Furthermore, highly-resistive or semi-insulating silicon substrates are difficult to achieve resulting in poorer isolation and higher losses in interconnects and passive devices. Each of these presents serious challenges to implementing mm-wave functions.

However, advances in silicon technology driven by high-performance digital applications, offer advantages to the mm-wave designer that might not be apparent on first consideration. Performance, quantified by f_T , f_{max} or NF_{min} for example, has dramatically increased with geometry scaling and technology enhancements in both CMOS and SiGe HBTs [1]. Both CMOS and BiCMOS technologies have been used to demonstrate circuit functioning at frequencies in and above the K-band. Now, these silicon technologies are, by virtue of nanometer-scale design rules, able to implement staggering amounts of digital logic in a given area thereby enabling the on-chip integration of sophisticated control logic for performance tuning and/or digital signal processing. Furthermore, the worldwide manufacturing capacity of silicon technologies driven by consumer applications like gaming and personal electronic appliances assures low-cost. This will certainly provide an impetus for the evolution of mm-wave consumer applications. The combination of mm-scale wavelengths, low cost and the ability to integrate begs the consideration of array-based transceiver topologies being implemented on a single die or package.

Andreia Cathelin
STMicroelectronics, Crolles, France e-mail: andreia.cathelin@st.com

John J. Pekarik
IBM Corporation, Essex Junction, Vermont, USA e-mail: pekarik@us.ibm.com

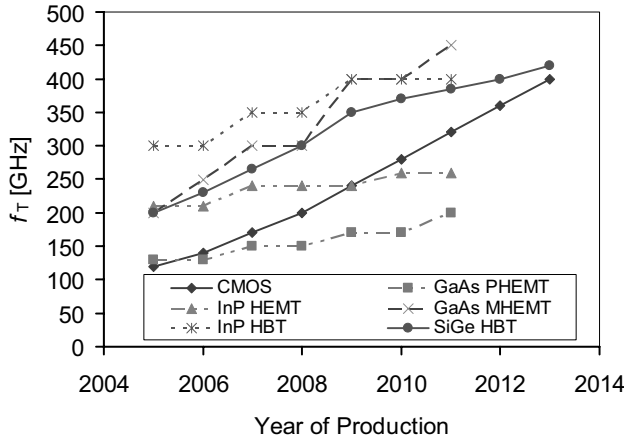


Fig. 2.1 Cutoff frequency by year of production comparing silicon and III-V compound semiconductor devices.

2.1.1 Performance

The International Technology Roadmap for Semiconductors (ITRS) [1] hopes to “ensure cost-effective advancements in the performance of the integrated circuit and the products that employ such devices” by compiling and publishing roadmaps which “identify critical challenges” and “encourage innovative solutions” with “focus on technologies that are crucial to the design, materials, and manufacture of semiconductors, as well as the factory sciences, process control, metrology, and environmental aspects.” One chapter of the ITRS focuses on Radio frequency and analog/mixed-signal (RF and AMS) technologies for wireless communications. In tables therein, the demands of a wide range of applications (cellular phones, wireless local area networks, wireless personal area networks (PAN), phased array RF systems, and other emerging wireless communication, radar, and imaging applications) operating between 0.8 GHz and 100 GHz are used to define needed capabilities of silicon and III-V compound semiconductor devices.

Fig. 2.1 shows the roadmap of the cutoff frequency (f_T) comparing a number of III-V semiconductor devices with the silicon CMOS NFET and SiGe HBT as taken from the 2006 ITRS [2]. Inasmuch as f_T is an adequate measure of transistor performance, the ITRS shows that silicon transistors are at least competitive with transistors made in III-V semiconductors. Furthermore, the RF & AMS chapter of the ITRS tabulates the f_T of the NFET from the Low Standby Power digital CMOS roadmap whereas the NFET from the High Performance digital CMOS roadmap is shown as having an intrinsic switching speed a factor of three higher. We have noted some of the challenges inherent with the use of silicon transistors and will discuss more below. However, it is evident that silicon technology currently exhibits small-signal gains that are competitive with those of III-V transistors and are predicted to scale at least as quickly in the near-term future.

2.1.2 Cost, Integration [3]

If silicon technology has adequate performance to implement the front-end portions of the transceiver, the ability to integrate digital logic in CMOS at increasing densities offers the opportunity to drastically lower overall system cost. Lower cost could be the prime motivator for the use of BiCMOS or CMOS over III-V technologies. Again considering f_T as a measure of performance, the SiGe BiCMOS HBT has comparable performance to the NFET at roughly twice the minimum feature size. For stand-alone RF functions, where area is dominated by passive devices and I/O pads, BiCMOS may be the lower-cost option despite the approximately 20% additional process complexity required to form the HBT. III-V transistor performance at substantially relaxed lithography dimensions is comparable with leading edge CMOS. So, again for purely RF devices, III-V implementations may be lower cost especially when utilizing existing designs and time-to-market is considered. However, when even modest amounts of digital logic are to be integrated, CMOS has a clear advantage as circuit density and chip size scale with the square of the minimum lithographic dimension.

This can be illustrated by comparing integration densities of BiCMOS and CMOS at equivalent f_T . The square of the second metal wiring level (M2) pitch is a measure of the integration density of the digital CMOS available at each technology node as illustrated in Fig. 2.2. Any digital content comes at that node's CMOS density (and performance). For a hypothetical chip that is 50% digital CMOS and 50% RF implemented in 0.35 μm BiCMOS that is being migrated to 0.18 μm CMOS, if we assume that the RF portion, being dominated by passive devices, does not scale, we see that the overall chip area shrinks by $\sim 25\%$ due to the dramatic increase in the density of the digital circuitry. By fitting proportionally more chips on a wafer that costs about the same, the economic benefit is obvious. Such an analysis needs to consider all aspects of product cost such as masks, packaging and volume discounts and needs to use current pricing information as competitive pressures usually force prices lower with time. Furthermore, many other considerations will factor into a decision on which technology to use for a given product IC. Designer expertise or the existence of verified circuit designs could drastically affect the time to market. Package, board and system assembly costs could well dominate the contribution of the IC die to the total product cost.

2.1.3 Manufacturing Capacity

The worldwide silicon manufacturing capacity for semiconductor wafers is staggering. It has increased by about 67% from the beginning of 2001 through the first half of 2007, and during that time enough wafer area was manufactured to cover roughly a fourth of the land area of Manhattan. Fig. 2.3 shows data from Semiconductor International Capacity Statistics (Sicas) [4]. The graph shows MOS capacity which includes BiCMOS. Bipolar-only capacity is not shown but that represents less than

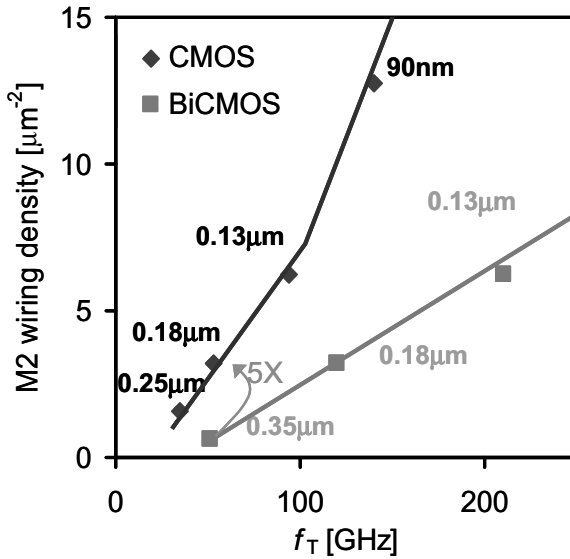


Fig. 2.2 Integration density (illustrated by second-level metal pitch) vs. unity current gain frequency for CMOS and BiCMOS technology nodes (© IEEE 2005).

5% of the total silicon area being produced. It is evident that the growth is almost exclusively in the smallest lithography node. Also, the capacity of older nodes tends to stabilize to a value slightly below its peak and remain on line for a long time. These two observations are interesting in that most of the demonstrated millimeter-wave circuits have been in the 130nm and smaller nodes. The capacity at 130nm appears to have stabilized and we can now expect commodity pricing pressures to come to bear. Anticipating historical trends will be repeated, the next major node at 90nm may have already reached peak capacity with rapid growth occurring at 65nm and 45nm capacity beginning to emerge. Given the added transistor performance at these smaller nodes, over half of the worldwide silicon manufacturing capacity is ready for the production of millimeter-wave devices.

2.2 Modern SiGe and CMOS Technology

2.2.1 Lithography

The performance trend illustrated in Fig. 2.1 is the continuation of a three-decade long trend primarily driven by advances in optical lithography. As illustrated in Fig. 2.4, the wavelength used in lithography has evolved to shorter and shorter lengths but, for some time now, on-wafer dimensions have been below the wavelength of light used to print them. A multitude of techniques are employed to allow printing sub-

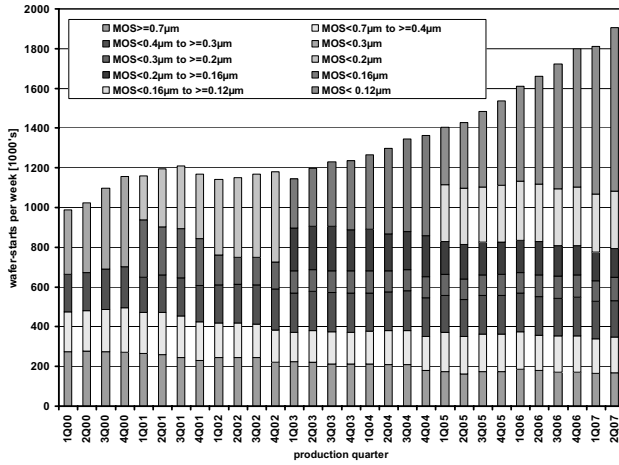


Fig. 2.3 Worldwide MOS manufacturing capacity in 8-inch wafer equivalent areas.

wavelength features. Design rules evolved from restrictions on minimum allowed linewidth, space and pitch to restrictions on pitch and orientation with further restrictions on feature placement and shape being considered. On masks, the use of optical proximity corrections (OPC) and sub-resolution assist features (SRAF) aids in the creation of as-printed patterns that more closely resemble as-design shapes. For critical mask levels, the gate, contacts and first metal, the use of phase-shifting is typically employed. The increasing complexity of calculating the optimum deployment of these on-mask techniques is driving the field of computational lithography [5]. Photoresist layer thicknesses have necessarily shrunk with the decrease in feature size. Thinner resist layers are unable to withstand etches or block ion implants to the extent of thicker resists. To aid in these functions, hardmask layers, for example silicon dioxide which selectively withstands silicon etches, are introduced into the process as a necessary complication. Softmasks, non photo-active polymer films, are used when additional resist thickness is needed and are sometimes included as anti-reflection coatings (ARC) to reduce optical interference effects. Photo resist can also be etched after images are developed to achieve smaller printed images, albeit without pitch reduction. Advances in lithography tooling include the use of off-axis illumination to improve depth of focus and automatic dose adjustment either by feedback from prior-lot measurements or in-situ measurements [2]. Major lithography technology advancements include immersion lithography now being introduced into production [6] and extreme ultraviolet (EUV) lithography still in development [7].

The dimensional shrinking of the physical structures on the wafer has led to a series of dramatic changes in materials and structures used in aggressively scaled CMOS process technologies. The following sections briefly touch on three of these that are likely to have particular impact to mm-wave design, the introduction of low-K dielectrics in copper wiring levels, the use of strain engineering to enhance carrier mobility, and the use of metal gates and high-K dielectrics in FETs.

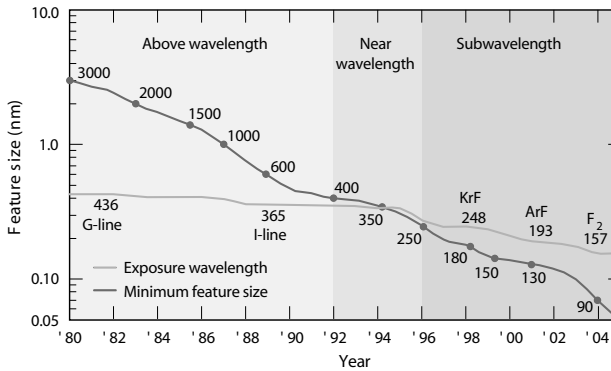


Fig. 2.4 Reduction in minimum feature size and exposure wavelength over time [8].

2.2.2 Low-K Dielectrics and Copper Wiring

Copper wiring was introduced in the mid-1990's to reduce RC-delay in narrow signal lines and to improve robustness to electromigration. The addition of low-K dielectric materials provides an additional reduction in wiring delay. A comparison of representative circuit delay, segregating wire and gate delays is shown in Fig. 2.5. A reduction of wiring delay of almost $3.5\times$ is illustrated and this full benefit is realized at the smallest dimensions where wiring delays dominate transistor delays. Recent utilization of porous low-K material, with relative dielectric constants as low as 2.4, provides additional incremental improvement as illustrated in Fig. 2.6. In order to assure uniformity in wire dimensions both horizontally, limited by lithography, and vertically, limited by etch and chemical-mechanical polishing (CMP), ground rule restrictions on wiring density are established. Furthermore, foundries often introduce perforations in wide metal shapes and introduce small tiles of metal to fill sparse areas. The impact of these is discussed in Section 2.4.

2.2.3 Mobility and Strain Engineering

For CMOS, through the 90nm node, the scaling of gate length and gate oxide thickness was roughly proportional. Further scaling of oxide thickness is limited by tunneling current and therefore, since the transconductance (g_m) is inversely proportional to the oxide thickness, any scaling of g_m must be accomplished through an increase in carrier mobility. Mechanical strain, which distorts the semiconductor crystal lattice, also distorts the energy band structure resulting in higher or lower carrier mobility depending on the carrier type (electrons or holes), whether the strain is compressive or tensile, whether the strain is uniaxial or biaxial, and the magnitude of the strain. Biaxial tensile strain in silicon and compressive or tensile biaxial strain in SiGe can

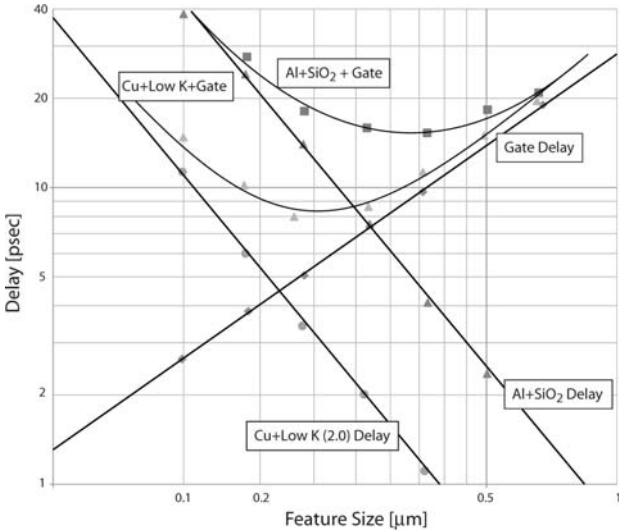


Fig. 2.5 Gate delay and interconnect delay vs. feature size comparing Al/SiO₂ and Cu/low-K wiring/dielectric schemes. Adapted from data in [9](© IEEE 1999).

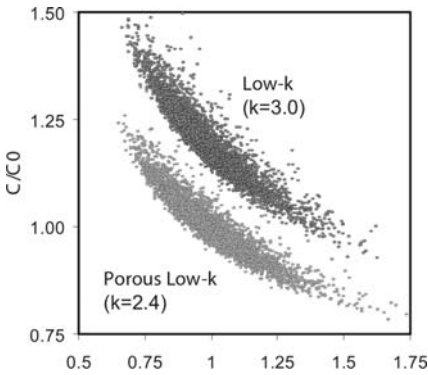


Fig. 2.6 Normalized capacitance versus resistance demonstrating reduced RC parasitics [6](© IEEE 2006).

be induced with epitaxial growth on SiGe buffer layers yielding enhanced electron and hole mobilities and enhanced NFET and PFET performance[10]. Uniaxial strain can be induced by depositing stressed films on top of the completed FET prior to contacts [11] and if the stressed film is deposited prior to the final anneals, the stress is preserved in the device structure even after removal of the film [12]. These techniques are very effective in imparting tensile strain in the channel which thereby improves NFET performance. Compressive uniaxial strain to improve PFET performance has been achieved by selective epitaxial growth of SiGe in the source/drain

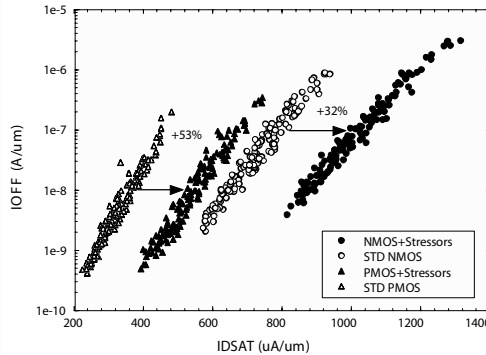


Fig. 2.7 Increase in FET drive current due to mobility enhancement techniques [14](© IEEE 2005).

region adjacent to the gate [13]. Combining these techniques can lead to dramatic simultaneous improvement in both NFET and PFET performance [14].

Carrier mobilities are dependent on crystal orientation for example, hole mobility is higher in the (110) direction while electron mobility is higher in the (100) direction. Silicon wafers using hybrid orientation technology (HOT) allow the simultaneous use of preferred crystal directions for NFETs and PFETs with a single gate orientation. Silicon wafers used in CMOS are conventionally oriented with gates in the (100) crystal direction [15]. Orienting the gates in the (110) direction and applying stressor films is another way to achieve simultaneous NFET and PFET improvement [16].

2.2.4 Metal Gates & High-K Dielectrics

In order to avoid the high tunneling currents resulting from thin gate oxide, dielectric materials with higher effective permittivity (high-K) are being introduced. For an equivalent or lower effective oxide thickness these materials can be physically thicker and have dramatically lower tunneling current [17]. The introduction of metal gate electrodes addresses the problem of polysilicon depletion in the high vertical electric fields now inherent in scaled CMOS and the problem of high gate resistance from polysilicon sheet resistance and contact resistivity. Coupling high-K dielectric materials with the appropriate choice of metal gates leads to FETs with high drive current [18] and improved high-frequency performance [19]. These technology elements are to be employed in high-performance 45nm digital CMOS [20].

2.3 Active Devices on Recent Bulk and SOI Technologies

This section gives an overview of several silicon technologies compatible with very high frequency operation: SiGe:C BiCMOS and bulk and thin SOI deep submicron CMOS. It also presents some active and passive device insights very useful for the millimeter-wave designer [21].

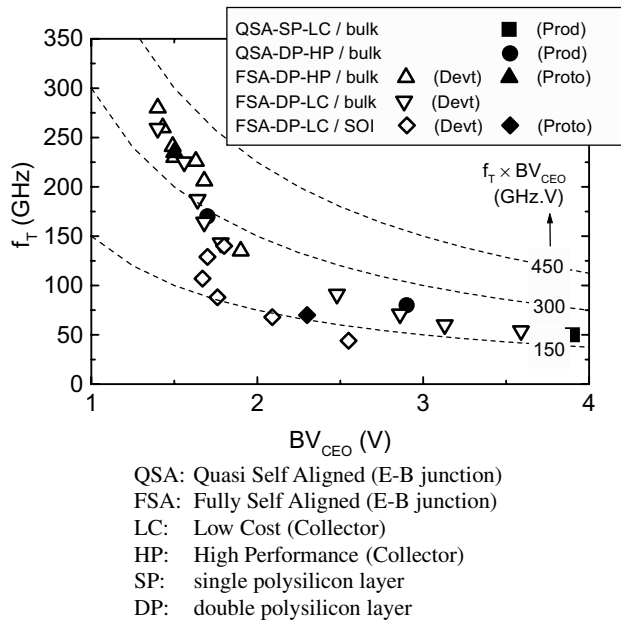


Fig. 2.8 $f_T - BV_{CEO}$ chart built with various Si/SiGeC HBTs available in 130-nm CMOS node. Different architectures with different maturities are compared (technology trials at STMicroelectronics) (© IEEE 2006).

2.3.1 Bipolar Devices

Silicon Heterojunction Bipolar transistors offer some advantages compared to CMOS devices such as lower $1/f$ noise, higher output resistance and higher voltage capability for a given speed [22], [16]. The range of technologies on the market today offers HBTs with $f_T > 200\text{GHz}$ and sometimes $f_{max} > 300\text{GHz}$, as depicted in Fig. 2.8 and Fig. 2.9. Different types of E-B structures are presented, but those obtaining $f_T > 200\text{GHz}$ and $f_{max} > 300\text{GHz}$ possess fully self aligned architectures (FSA) and a high performance collector. And finally, the f_{max} parameter for an HBT, has reduced sensitivity to layout parasitics when compared to a MOSFET.

Fig. 2.10 and Fig. 2.11 show f_T and f_{max} data gathered from several foundries, together with the very aggressive ITRS road-map (these two figures should be read together, as the given performances correspond to a given technology node). Differences may come from the differentiation between pure bipolar and BiCMOS processes.

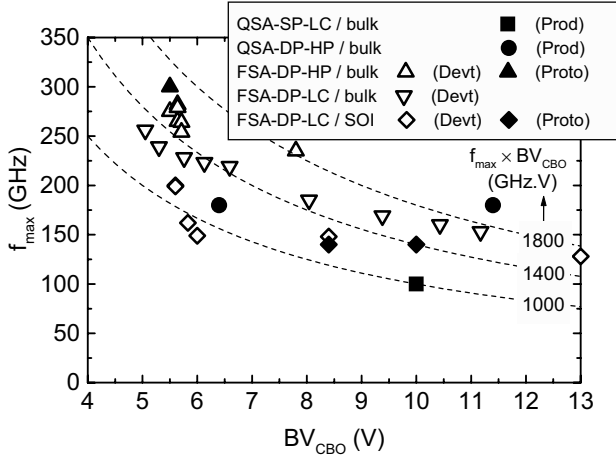


Fig. 2.9 f_{max} - BV_{CBO} chart built with various Si/SiGeC HBTs available in 130-nm CMOS node. Different architectures with different maturities are compared (© IEEE 2006).

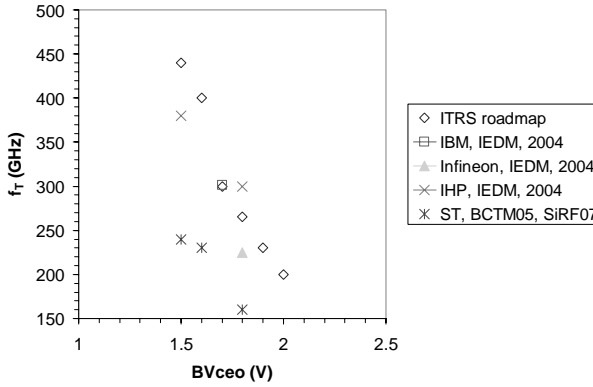


Fig. 2.10 f_T - BV_{CEO} chart for different technology nodes.

2.3.2 CMOS devices

CMOS transistors follow the well-known Moore's Law of scaling, thus leading to always increasing functional integration. The 65nm node still uses polysilicon gate, but the carrier mobility is sometimes increased by using several technological solutions as described previously. For MOS devices, $f_T \propto 1/L_g^\alpha$, where ($\alpha \sim 1$) and, as a first order approximation, is independent of the gate oxide thickness, as depicted in Fig. 2.12. f_T as high as 150GHz and 200GHz are reached in the 65nm node for Low Power (LP) and General Purpose (GP) devices, respectively. Data gathered

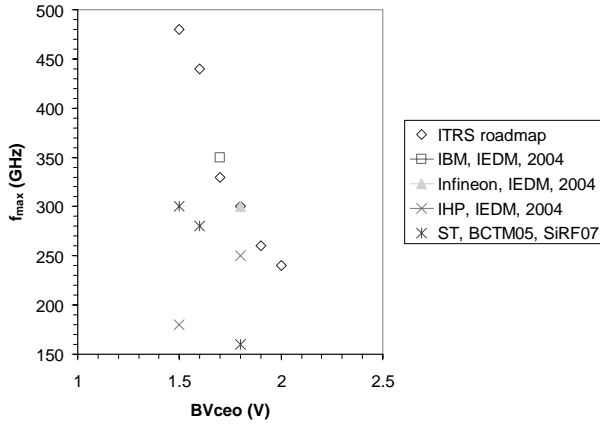


Fig. 2.11 f_{max} - BVCBO chart for different technology nodes.

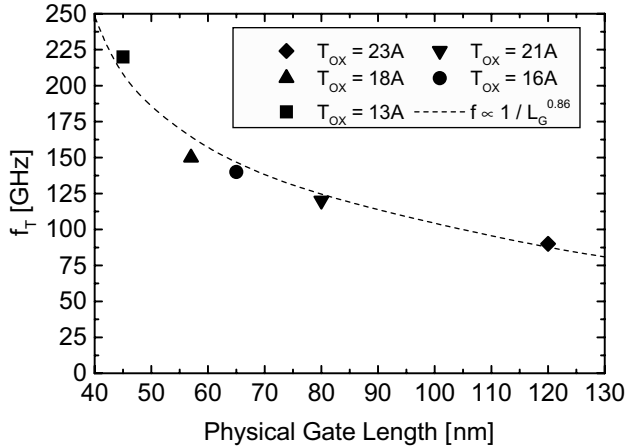


Fig. 2.12 Evolution of f_T with physical gate length for different NMOS devices (LP and GP) of 130-nm, 90-nm and 65-nm CMOS nodes (© IEEE 2006).

from several major semiconductor foundries show good conformity with the ITRS road-map, as depicted in Fig. 2.13.

The following concerns apply to both bulk and SOI devices, as long as they show the same gate length and the same carrier mobility. Equation 2.1 presents one way of calculating the transition frequency f_T [23] for deep submicron technologies:

$$f_T \approx \frac{g_m}{2\pi C_{gin} \sqrt{1 + 2 \frac{C_{Miller}}{C_{gin}}}} = \frac{f_c}{\sqrt{1 + 2 \frac{C_{Miller}}{C_{gin}}}} \quad (2.1)$$

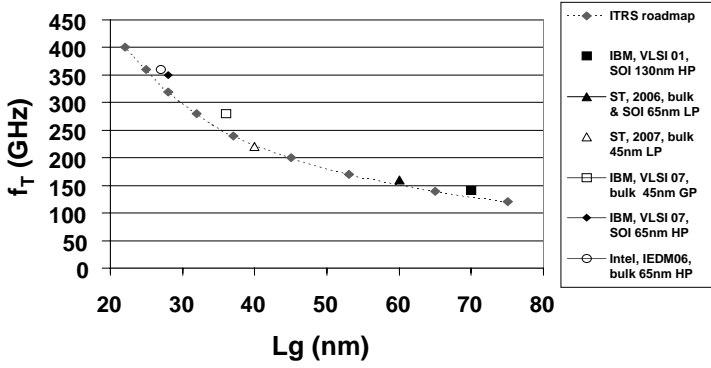


Fig. 2.13 f_T ITRS road-map and several foundries' performances.

With g_m , the gate transconductance, and as also depicted in Fig. 2.14

$$C_{gin} = C_{gsi} + C_{overlap} + C_{fringing} \quad (2.2)$$

$$C_{Miller} = C_{gdi} + C_{overlap} + C_{fringing} \quad (2.3)$$

and the intrinsic cut-off frequency

$$f_c = \frac{g_m}{2\pi C_{gin}} \quad (2.4)$$

Where:

- C_{gsi} , C_{gdi} = the equivalent capacitance induced by the source / drain field effect into the channel (see Fig. 2.15)
- $C_{overlap}$ = the equivalent capacitance given by the LDD (low doped drain / source regions) diffusion under the gate (see Fig. 2.15)
- $C_{fringing}$ = the parasitic capacitance depending on the gate height and on the contact to gate distance (see Fig. 2.16).

As it can be seen, f_T increases for reduced gate length devices thanks to higher transistor transconductance [24]. Concerning the capacitive part, f_T is degraded by a high gate to drain capacitance (called also the Miller capacitance).

From a theoretical point of view, the transistor's transconductance is increasing with the reduction of the effective gate length. Nevertheless, for advanced deep submicron technologies this trend is not straightforward. The features limiting this evolutionary trend are the following:

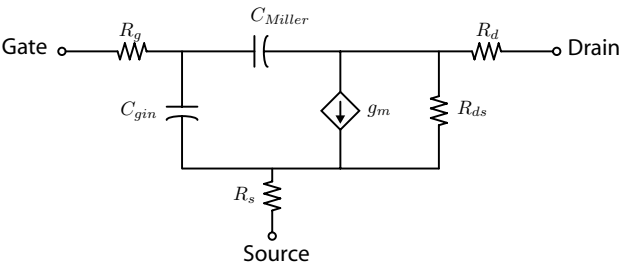


Fig. 2.14 MOS transistor small-signal equivalent schematic for f_T and f_{max} calculations.

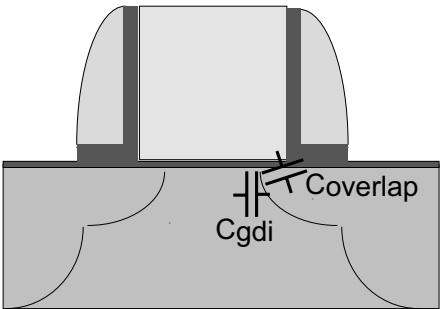


Fig. 2.15 Schematic illustration of the C_{gdi} and $C_{coverlap}$ equivalent capacitances, C_{Miller} for a MOS device.

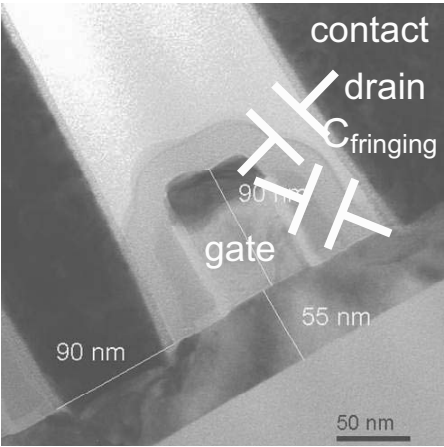


Fig. 2.16 Illustration of the $C_{fringing}$ parasitic capacitance, C_{Miller} for a MOS device.

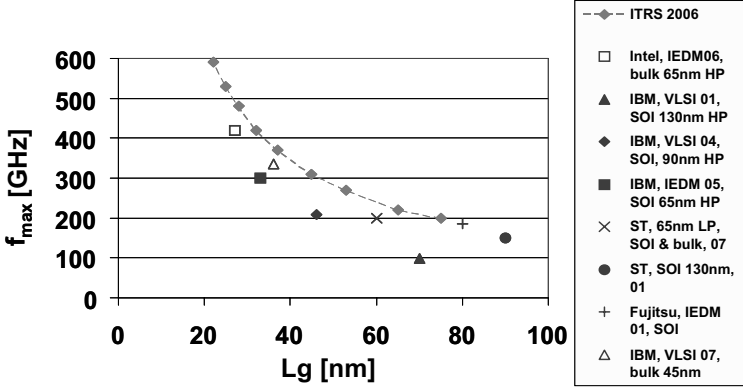


Fig. 2.17 f_{max} ITRS road-map and several foundries' performances.

- Reduced gate oxide thickness. This feature induces gate tunneling current, which is a critical point for Low Power technologies.
- Active zone doping increase. This feature generates mobility degradation.
- Low doped drain regions (LDD) . This feature increases directly the source and drain equivalent series resistance.

As discuss in Section 2.2, several process techniques have been hence developed in order to cope with these transconductance increase limitations. As presented in [16], the hole mobility in PMOS transistors is increased by using 45° rotated devices and for the electron mobility in NMOS, tensile liner films are used. In order to limit the gate leakage and to further decrease the effective electrical gate length, high-K dielectrics are used as a part of the gate material. High f_{max} values have also been reported for CMOS devices, such as 200GHz on 65nm LP node. Fig. 2.17 presents f_{max} evolution with respect to the effective gate length, for several deep submicron technologies coming from different silicon manufacturers.

Equation 2.5 [23] gives one calculation method for the maximum device frequency f_{max} , all the constants having the well-known significance (see also Fig. 2.14). R_i is the equivalent non-quasistatic resistance.

$$f_{max} \approx \frac{g_m}{2\pi C_{gin}} \frac{1}{2\sqrt{(R_g + R_s + R_i) \left(g_d + g_m \frac{C_{Miller}}{C_{gin}} \right)}} \quad (2.5)$$

$$= \frac{f_c}{2\sqrt{(R_g + R_s + R_i) \left(g_d + g_m \frac{C_{Miller}}{C_{gin}} \right)}} \quad (2.6)$$

The f_{max} of CMOS devices, which should correlate to the performance of large signal operation blocks such as mixers, oscillators and power amplifiers, is very sensitive to layout parasitics and also to the choice of the transistor's finger width. An optimal finger width is chosen for the majority of millimeter wave circuits, all

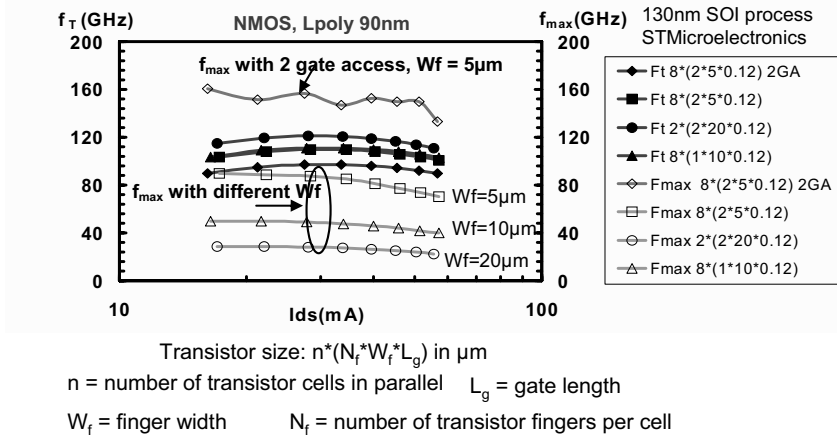


Fig. 2.18 Influence of MOS transistor layout over f_T and f_{max} parameters.

the solutions seem to converge towards values lower than $5\mu\text{m}$, with two sided gate contacts. A study of this trend is illustrated in Fig. 2.18 [24], where the influence of transistor's layout on the f_T and f_{max} parameters is evaluated on several transistors with the same total equivalent size, in the 130nm CMOS node. This technology trial study has been held on a SOI technology from STMicroelectronics, but it may of course be extrapolated to any CMOS process. The transistor is divided into several identical cells, and each cell is an interdigitated device with a given unitary finger width.

Several transistor layout topologies are used in order to minimize the extrinsic parasitic elements added to the intrinsic transistor core [25], [26]. The f_{max} parameter is extrapolated from very sensitive high frequency S parameter measurements, and may sometimes suffer from the de-embedding technique. The transistor layout structure presented in [26] is depicted in Fig. 2.19. In this case the transistor is divided into two equal parts with a double contacted gate. This allows minimizing parasitic capacitances and gate resistance while using coplanar access [27]. Source contacts are located on the transistor periphery making easier the contact to the coplanar ground plane. Finally, this structure allows an impedance matched access towards and from the active device to the rest of the circuit.

Table 2.1 presents a comparison between the presented transistor layout and a classical in-line transistor structure [27]. The comparison is done for a CMOS 65nm LP bulk transistor with $W_{total} = 60\mu\text{m}$, $L_g = 60\text{nm}$ and with $W_{finger} = 1\mu\text{m}$ embedded in a 6 metal layers back-end (last metal layer is thick). Both structures have 2 sided gate contacts. The extrinsic parasitic elements in this table correspond to an electrical model on the top of a core BSIM4 electrical model presented here as to be the intrinsic NMOS transistor model (see also Fig. 2.20). The intrinsic electrical core model comprises the front-end device plus the contacts and Metal 1 layer on the top of it.

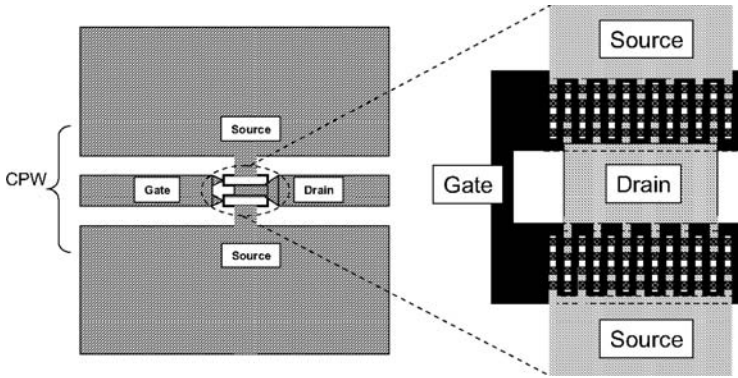


Fig. 2.19 Layout of n-MOS transistors to maximize mm-wave performances and to limit discontinuity [6] (© IEEE 2007).

Table 2.1 Comparison of extrinsic small-signal equivalent circuit parameters for an in-line transistor and the structure presented in Fig. 2.19. LP CMOS 65nm transistor with intrinsic $f_T \sim 160\text{GHz}$ and $f_{max} \sim 200\text{GHz}$, $W_{total} = 60\mu\text{m}$, $L_g = 60\text{nm}$, $W_{finger} = 1\mu\text{m}$, 6 Metal BE .

Extrinsic parasitic elements In-Line Device Fig. 2.19 Device		
C _{gs} (fF)	6.4	3.7
C _{gd} (fF)	0.3	3.2
C _{ds} (fF)	5.3	4.4
R _s (Ω)	0.1	0
R _d (Ω)	0.1	0.1
R _g (Ω)	4	0.35
L _s (pH)	4	0
L _d (pH)	4.5	2
L _g (pH)	4	2
Extrinsic f_T (GHz)	118	138
Extrinsic f_{max} (GHz)	125	194

The noise parameter NF_{min} is also an important marker for the RF and mm-wave performances of an active device. The very low noise figure values, presented by deep submicron CMOS devices nowadays, are very difficult to measure with an excellent accuracy, as the de-embedding is very complex in the high frequency range. For the MOS transistor, the following formula can be used to calculate the NF_{min} parameter (see also Fig. 2.14).

$$NF_{min} \approx 1 + \frac{f}{f_T} \sqrt{g_m(R_g + R_s)} \quad (2.7)$$

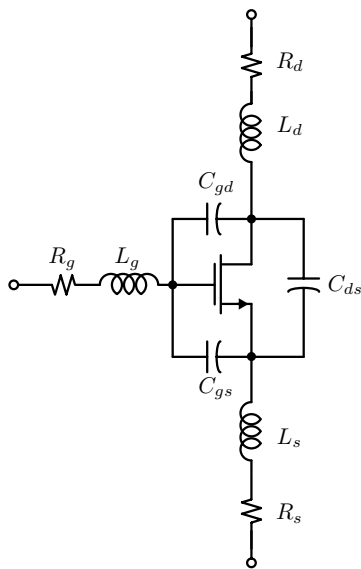


Fig. 2.20 Extrinsic parasitic elements extracted in Table 1.1 for an NMOS transistor (with respect to a BSIM4 intrinsic model).

As it can be easily recognized, this parameter is strongly degraded by high gate and source parasitic resistances, thus being layout dependent. Fig. 2.21 presents some NF_{min} data, gathered from two silicon foundries and presented with respect to the ITRS road-map. One may observe that similar RF noise behavior may be obtained from different technology nodes transistors, depending on the transistor's architecture. Nevertheless, both bulk and SOI processes from the same technology node perform the same noise behavior, given that the transistors show the same transconductance and the same parasitic resistances.

Figures 2.22 and 2.23 present the frequency evolution of NF_{min} for several total widths of a given NMOS transistor (W in μm), data coming from bulk and SOI FB 65nm LP CMOS devices from STMicroelectronics [28].

2.3.3 SOI CMOS devices

As seen from the previous section, moving from bulk to thin SOI does not change the HF performances of CMOS devices significantly [24].

Two kinds of CMOS transistors are available on SOI [29], floating body (FB) and body contacted (BC) devices (see Fig. 2.24). FB devices show the same layout as a bulk CMOS transistor. Thin SOI Partially Depleted (PD) technology induces a floating body effect in the transistor's channel, which is beneficial for digital operation. Nevertheless, this device suffers from Kink effect (see Fig. 2.25) that degrades

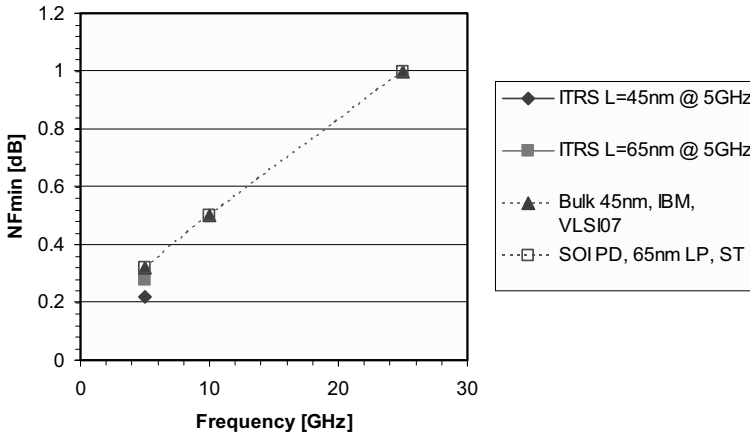


Fig. 2.21 NF_{min} ITRS road-map and several foundries' performances.

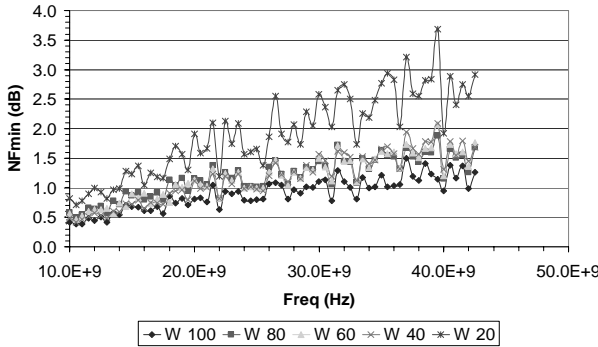


Fig. 2.22 Evolution of NF_{min} with frequency; bulk and SOI 65nm LP CMOS; $L_g=60\text{nm}$, 2 gate access, $W_{finger}=1\text{ }\mu\text{m}$, $I_{ds}=150\text{ }\mu\text{A}/\mu\text{m}$, $V_{ds}=1.2\text{V}$ (data from STMicroelectronics).

both voltage gain (high g_{ds} , see Fig. 2.26) and $1/f$ noise (see Fig. 2.27). This excess low frequency noise occurs only in the Kink region and only for the floating body transistors. The body contact efficiency depends on the body resistance restricting access to the channel, thus a maximum finger width limit should be applied for BC devices to maintain low body resistance. Nevertheless, the low frequency noise behavior of a BC device is identical to the one of a similar bulk transistor. Given all this information, it is obvious that BC transistors are well-suited for the analog part of an RF/mm-wave design.

However, the Kink effect has no impact on HF characteristics and SOI FB MOS-FETs exhibit f_T and f_{max} at least identical to those of bulk devices with identical

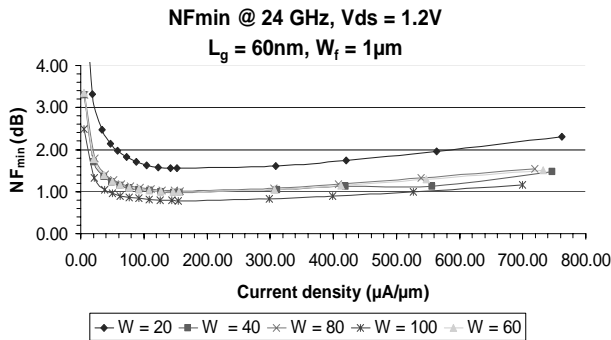


Fig. 2.23 Evolution of NF_{min} @ 24GHz with I_{ds} ; bulk and SOI 65nm LP CMOS (data from STMicroelectronics).

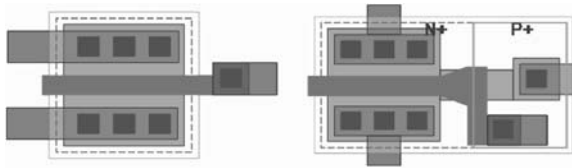


Fig. 2.24 Layout difference between Floating Body (FB) - left, and Body Contacted (BC) - right, devices in SOI technologies (130nm technology node).

layout. On the contrary, HF performance of BC devices is penalized by their specific layout (higher gate resistance and gate to source capacitance). The example depicted in Fig. 2.28 shows 65nm LP SOI NMOS transistors that are well suited for mm-wave design:

- for the FB NMOS device: $f_T \sim 160\text{GHz}$, $f_{max} \sim 200\text{GHz}$
- for the BC NMOS device: $f_T \sim 108\text{GHz}$, $f_{max} \sim 126\text{GHz}$. In this case this is a worst case value, as because of the particular layout architecture of BC devices, the optimum W_{finger} for those devices is different from the one of FB devices.

As depicted in Fig. 2.21, SOI FB devices show the same noise performance as the bulk ones. Obviously, the BC transistors show larger noise figures, given the increased gate resistance and some capacitive coupling to the body contact. Fig. 2.29 shows measured NF_{min} for FB and BC NMOS devices in a 65nm LP SOI CMOS technology, together with the associated gain.

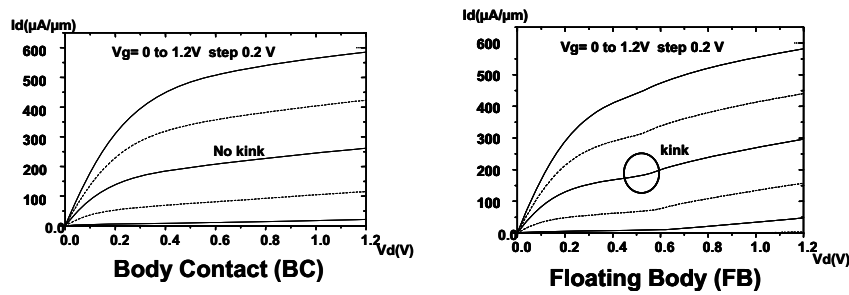


Fig. 2.25 DC curves for FB and BC devices in PD SOI CMOS.

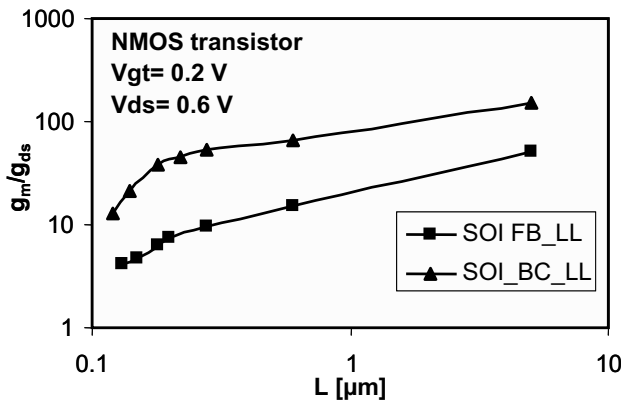


Fig. 2.26 Comparison of voltage gain for FB and BC devices in 130nm SOI CMOS (data from STMicroelectronics).

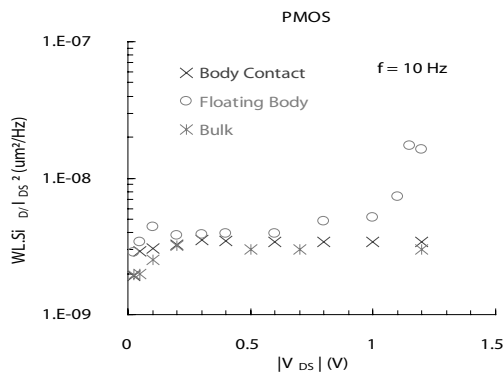


Fig. 2.27 Low Frequency noise power spectral density; comparison between FB, BC and bulk transistors in 130nm SOI CMOS (data from STMicroelectronics).

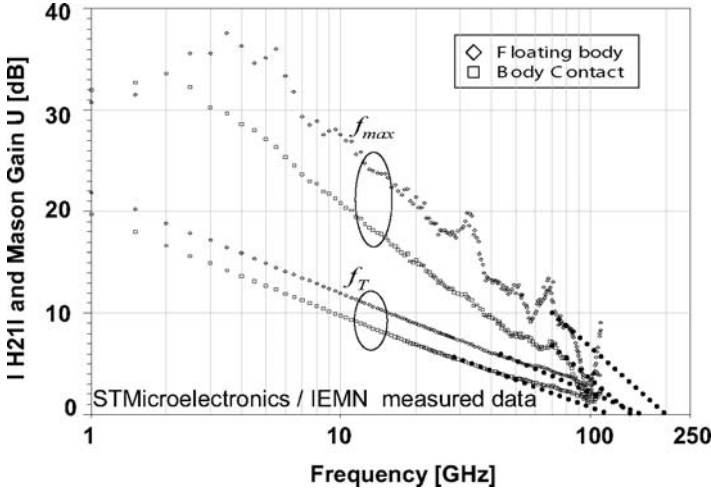


Fig. 2.28 f_T and f_{max} behavior of FB and BC NMOS devices in a 65nm LP SOI CMOS technology (data from STMicroelectronics, IEMN measurement data); $V_{ds}=1.2V$, $I_{ds}=0.3mA/\mu m$, $L_g=60nm$, $W_{finger}=1\mu m$, $W_{total}=64\mu m$ (© IEEE 2007).

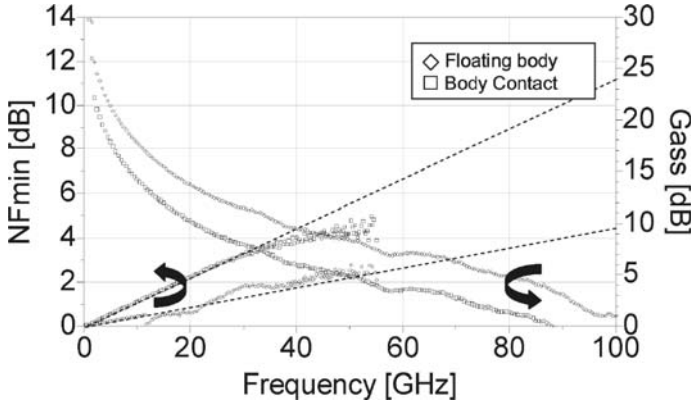


Fig. 2.29 NF_{min} behavior of FB and BC NMOS devices in a 65nm LP SOI CMOS technology (data from STMicroelectronics, IEMN measurement data); $V_{ds}=1.2V$, $I_{ds}=0.1mA/\mu m$, $L_g=60nm$, $W_{finger}=1\mu m$, $W_{total}=64\mu m$ (© IEEE 2007).

2.3.4 Current Density Scaling for CMOS and Bipolar Devices

A very interesting scaling feature has been pointed out in [30], regarding GP bulk and SOI NMOS transistors from different technology nodes. It has been shown that, as a result of constant-field scaling, the peak f_T ($\sim 0.3\text{mA}/\mu\text{m}$), peak f_{max} ($\sim 0.2\text{mA}/\mu\text{m}$) and optimum noise figure NF_{min} ($\sim 0.15\text{mA}/\mu\text{m}$) current densities are unchanged from one technology node to another (see Fig. 2.30, Fig. 2.31). This has been proven by different foundries' measurements with technology nodes from $0.25\mu\text{m}$ down to 90nm . This feature is very convenient when porting CMOS design from one technology node to another. Having in mind these data, constant current-density biasing schematics may be applied to CMOS circuit designs for analog, RF/mm-wave and also high speed digital functions. This way, the impact of statistical process variation, temperature and bias current variation may be drastically reduced.

For HBT devices from different technology nodes, this constant current density for peak f_T is not observed, as depicted in [31] and in Fig. 2.32. Indeed, one efficient way to increase the transistor's f_T by delaying the onset of the Kirk effect is to increase the collector doping. Hence, the collector current density at peak f_T rises under these circumstances.

2.3.5 Comparison Between State-of-the-Art HBT and CMOS Devices

In general, bipolar devices are specially developed for high frequency / high speed operation modes, while the MOS devices are implemented first for digital applications and then slightly optimized for analog/RF purposes. Fig. 2.33 and Fig. 2.34 attempt to realize a comparison between state of the art contemporary 130nm node SiGe:C HBT and 65nm LP node NMOS devices (data from STMicroelectronics). For this comparison, the dimensions chosen for the two devices are suited geometries for mm-wave design (mostly for low power linear operation). The bipolar device is a 130nm based FSA Si/SiGe:C HBT with 3 emitter fingers of $2.5\mu\text{m}$ each. The NMOS device consists of 10 cells in parallel, each cell of 9 gate fingers of $W_{finger} = 1\mu\text{m}$ each and the gates are contacted on both sides.

The first observation is the current difference for the peak value of f_T (see Fig. 2.33). The MOS device has its peak f_T value at a drain current density per unit width of $0.3\text{mA}/\mu\text{m}$, while the equivalent value for the bipolar device is reached at a collector current density per emitter length of $1.9\text{mA}/\mu\text{m}$. Fig. 2.33 presents also the minimum noise figures measured at 40GHz for both devices. NF_{min} of 1.1dB and 0.95dB are measured for the HBT and the NMOS devices, respectively. Fig. 2.34 presents the NF_{min} evolution with the frequency for both devices. Performing very accurate noise figure measurements at such high frequencies with a tolerance lower than 0.1dB is very difficult. When reading the cited figure, for frequencies below 40GHz , the MOS device performs a better noise figure than the

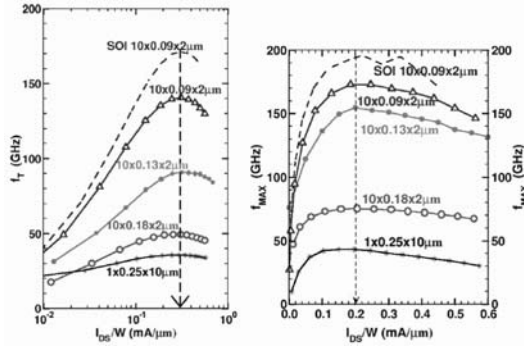


Fig. 2.30 Measured f_T and f_{max} as a function of drain current per micron of gate width for n-channel MOSFETs fabricated in different bulk and SOI technology nodes (different foundries) (© IEEE 2006).

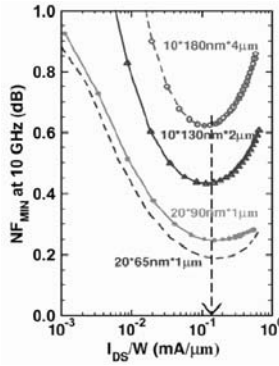


Fig. 2.31 Simulated NF_{min} at 10 GHz as a function of drain current per micron of gate width in nMOSFETs fabricated in different technology nodes (© IEEE 2006).

bipolar, but this trend should be reversed for frequencies above 60GHz. The noise figure of the NMOS device can be extrapolated to high frequencies with a good degree of confidence, which would give a value of about 2.7dB at 80GHz. It is to be noticed that NF_{min} is reached for all the frequencies at a constant value of V_{GS} . A linear fit can be applied to HBT too, but with a lower level of confidence, as NF_{min} shifts towards a higher V_{BE} when the frequency increases. Nevertheless, the HBT curve in Fig. 2.34 shows a nice fit between measurement and the linear fitted curve, up to 65GHz. An extrapolation using the small-signal equivalent circuit would give $NF_{min} \sim 2.2\text{dB}$ at 80GHz, and this value is coherent with the one obtained from the linear extrapolation ($\sim 2\text{dB}$). As a fair comparison sum-up, it may be admitted that

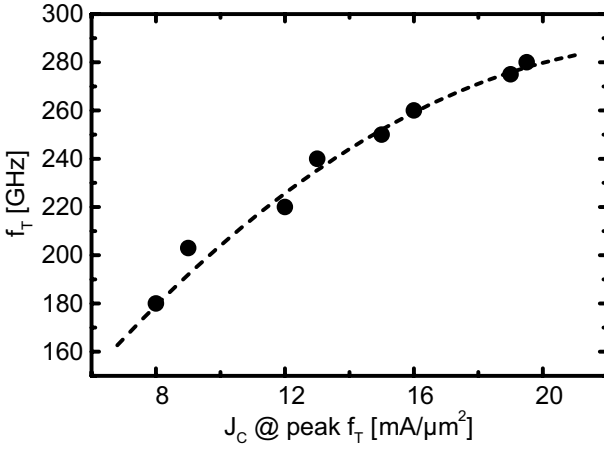


Fig. 2.32 Evolution trend of the current gain cut-off frequency f_T with the collector current density J_C (technology trials at STMicroelectronics) (© IEEE 2006).

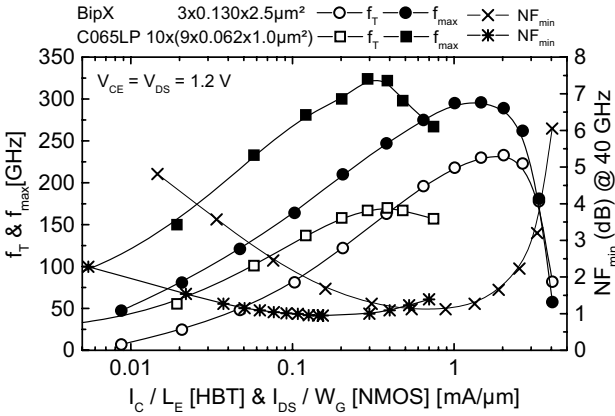


Fig. 2.33 f_T , f_{max} and NF_{min} at 40GHz vs. current density for 0.13 μm BiCMOS and 65-nm LP NMOS (data from STMicroelectronics) (© IEEE 2006).

both the bipolar and the NMOS device exhibit rather similar noise performances. For a given f_{max} , the higher current densities and voltage swing, the lower collector to bulk capacitance, along with a higher transconductance, permit the HBT to acquire significant advantage over the MOS transistor for large signal operation blocks such as power amplifiers. Nevertheless, and also to conclude this sub-section, both devices exhibit RF features compatible with millimeter wave circuit integration with industrial margin. The technology choice is given at the end of the day by the type of application, the technology availability and by the SiP or SoC approach chosen for the product.

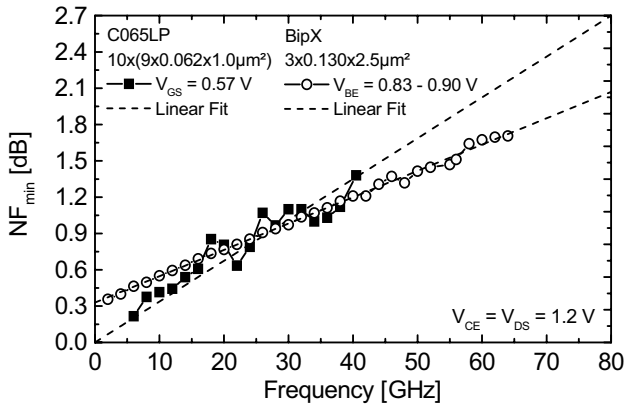


Fig. 2.34 Evolution of NF_{min} with the frequency for $0.13\mu\text{m}$ BiCMOS and 65-nm LP NMOS (data from STMicroelectronics) (© IEEE 2006).

2.4 Impact of the Back-End of Line on mm-Wave Design

Contrary to what usually happens in analog (and in a given perimeter in RF) design, information on the active device is not sufficient for the millimeter wave designer in order to choose the most appropriate technology. For frequencies above 10GHz , each μm of back-end strip has a significant influence on the electrical behavior of such a circuit. The trend when moving from one CMOS node to the next is the vertical shrink of the BEOL together with the decrease of the metal and dielectric thicknesses and of the metal pitch in order to increase integration density. Diminishing the lateral dimensions (width) of a metallic path imposes also a shrink on the vertical dimensions (height). The generally used rule of thumb in this case gives a shape factor of about three between a metal conductor height and its minimal width [32]. The Back End of Line (BEOL) evolution through technology nodes from $0.35\mu\text{m}$ down to 65nm is depicted in Fig. 2.35. One may observe that the general trend is somehow in antinomy with higher and higher working frequencies. In order to insure a very (and ultra) large scale of integration for the active devices, the first metal layers become thinner as do the corresponding Inter Metal Dielectric (IMD) layers. The increase of the integration scale imposes also the use of a larger number of metal levels and also implies some changes in the nature of the metallic materials used. The first technology generations developed for RF design such as $0.35\mu\text{m}$ to $0.18\mu\text{m}$ had Aluminum layers. Starting from the $0.13\mu\text{m}$ generation node, the metallic layers are made of copper which permits, for an equivalent metallic strip conductivity, to diminish the height of the deposited layer.

The connecting line on Fig. 2.35 represents the total dielectric height for a microstrip transmission line built with the last metal layer (for deep submicron technologies, not all the metal layers are supposed to be used as for a digital design). One can observe that the total height of dielectric is diminishing from one technology node to another, thus increasing the influence of the substrate losses on the propagation

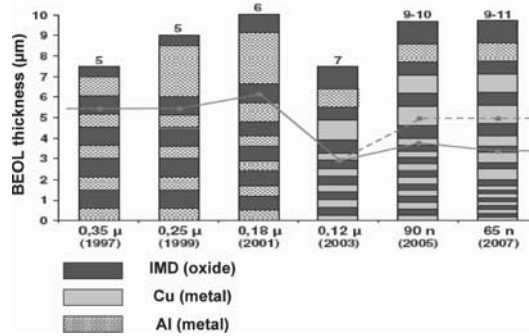


Fig. 2.35 BEOL evolution for several technology nodes (CMOS based).

constant. The dielectric oxide permittivity tends to diminish from one technology node to the next one, in order to limit the coupling effects between two conducting layers. Nevertheless, the use of thinner layers imposes an increased contribution of silicon nitride (in order to limit the copper diffusion into the oxide), which presents a higher permittivity. With these considerations we can conclude that the total Inter Metal Dielectric height diminishes with the technology nodes and the equivalent dielectric permittivity remains mostly constant.

For localized passive elements such as inductors, the small metal line pitch together with the thin dielectric layers induce larger ohmic and substrate losses. For lumped elements such as transmission lines, the patterned ground shields are not sufficient to limit the attenuation constant degradation. In order to cope with these design difficulties for mm-wave circuits, different trends came out during the latest years. Two families of back-ends may be distinguished: while the main stream is held by the (deep) sub-micron digital processes BEOL, the second trend is presenting millimeter-wave dedicated back-end. Fig. 2.36 illustrates a mm-wave dedicated BEOL for a 130nm node high performance BiCMOS process from STMicroelectronics [31].

It is obvious that, for millimeter-waves dedicated BEOL, the important number of thick metals (3 typically) and the use of thick Inter Metal Dielectrics provides for low losses in the transmission lines. In counterpoint, it imposes the development and the use of an application dedicated process, which sometimes may present higher development costs with respect to standard back end solutions. For a digital BEOL, a large number of metal levels lowers the losses of last metal level transmission lines, but in counterpoint one may look on the economic impact of processes with up to 10 metal levels.

The availability of High Resistivity (HR) SOI substrates for CMOS VLSI integration proposes an alternative to these two solutions, by drastically reducing all

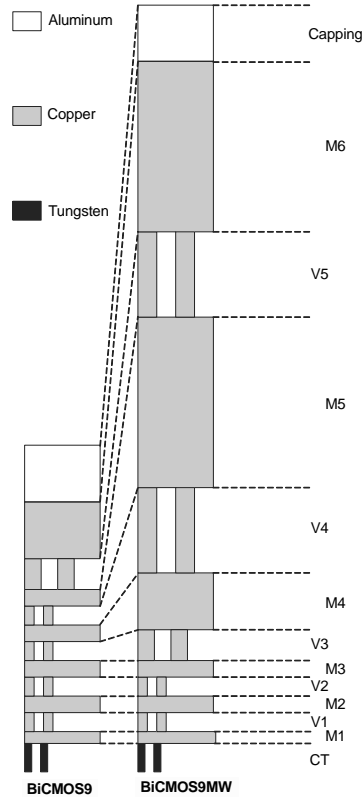


Fig. 2.36 Schematic comparison of the BEOL cross-sections of a standard 130nm BEOL and a mmw dedicated BEOL in the same technology node (both feature 6 copper metal layers but with twofold height difference). (data from STMicroelectronics)(© IEEE 2007).

parasitic losses towards the substrate [29], [33] while using standard digital back-end of line. Figure 2.37 presents, as an example, a cross section view on a 65nm CMOS SOI technology from STMicroelectronics. Like other SOI technologies, this one is distinguished by the presence of a buried oxide layer allowing the use of a high resistivity (HR) substrate underneath. The substrate resistivity value is superior to $3\text{k}\Omega\cdot\text{cm}$. The back end of line is composed of 6 or 7 copper metal layers (depending on the option) and a last encapsulated aluminum layer.

Fig. 2.38 illustrates the impact of the several types of back-end and substrate resistivity on mm-wave transmission lines [34]. Table 2.2 gives detailed explanations on the BEOL and substrate type, as well as implementation data for the compared transmission lines in Fig. 2.38. (Note: HR SOI CMOS measurements have been performed up to 200GHz with a very good confidence level, and gave good concordance with simulated data [33].)

The first comment is that, as expected, the mm-wave dedicated BEOL should allow the lowest attenuation constants for such 50Ω transmission lines, such as

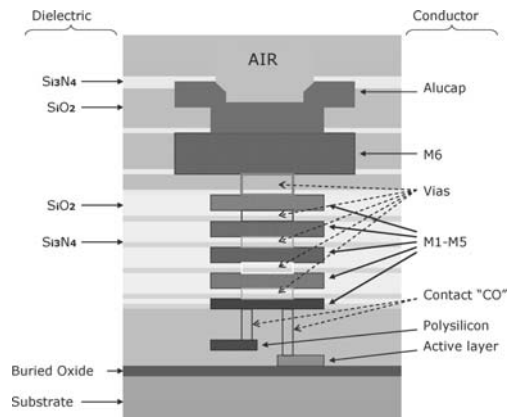


Fig. 2.37 Cross section view (BEOL zoomed) of a CMOS 65nm SOI technology.

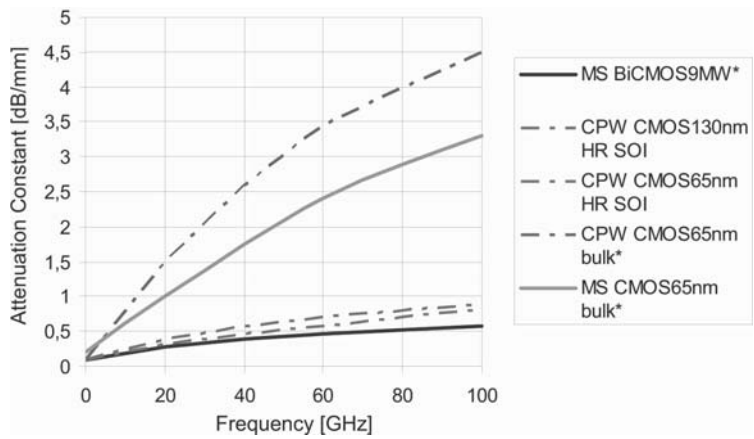


Fig. 2.38 Measured and simulated(*) attenuation constants for 50Ω transmissions lines in different technologies (all lines are compliant with manufacturing design rules) (data from STMicroelectronics).

0.5 dB/mm at 60GHz. On the other hand, both coplanar and microstrip transmission lines in a pure digital BEOL from the 65nm node show more losses, such as 2.5 to 3.5 dB/mm at 60GHz. The good surprise comes from the same 65nm node BEOL used this time in a HR substrate SOI technology, where one may observe attenuation constants almost as good as those from the mm-wave dedicated back end from a 130nm node. And finally, there is only 0.1dB/mm of difference between a 130nm and a 65nm BEOL transmission line on HR SOI processes.

Table 2.2 50Ω transmission lines characteristics for several BEOL and substrate resistivities (data from STMicroelectronics).

Reference	MS BiCMOS9MW	CPW CMOS130nm HR SOI	CPW CMOS65nm HR SOI	CPW CMOS65nm	MS CMOS65nm
Technology	BiCMOS (mmW dedicated): - 6 metal levels - 3 thick	SOI CMOS 130nm(digital): - 6 metal levels - 1 thick	SOI CMOS 65nm(digital): - 6 metal levels - 1 thick	CMOS 65nm (digital): - 6 metal levels - 1 thick	CMOS 65nm (digital): - 6 metal levels - 1 thick
Substrate	Low Resistivity	High Resistivity	High Resistivity	Low Resistivity	Low Resistivity
Line type	Microstrip	Coplanar	Coplanar	Coplanar	Microstrip
Size	W=5 μm (M6+Alu) Distance to ground (M1+M2), h=8.4 μm	W= 26 μm (M1 to Alu) Space to ground (M1 to Alu) s=22 μm	W= 12 μm (M1 to Alu) Space to ground (M1 to Alu) s=5 μm	W= 12um (M6 + Alu) Space to ground (M1 to Alu) s=7um	W=3.6um (M6) Distance to ground (M1) h=2.17um

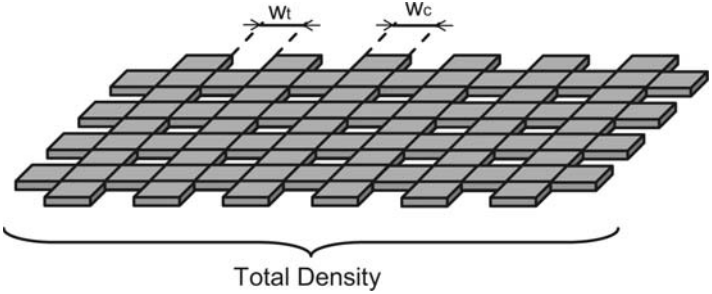


Fig. 2.39 Example of metal layer design rules .

Vertical shrink of the BEOL together with the use of copper layers impose also very drastic layout design rules in terms of metal densities per very tiny areas and also electromigration rules especially for high temperature. The deposition method for copper layers is damascene, based on the principle of dielectric trenches which are filled with metal followed by the CMP step which gives the desired level uniformity. However this uniformity operation is possible only if the underneath metal layers densities are homogeneous all over the wafer. This implies that either full metal drawing or the absence of metals on a large wafer surface is not allowed. Thus, the metallization levels must be perforated if their area is too large and some small tiles of dummy metal must be placed to preserve the homogeneity if the initial density is too low. It should be noticed that these rules of densities follow the projection of the technologies in the future, when these constraints will be more and more severe (smaller control windows). Fig. 2.39 illustrates some of these design rules on one metal level.

Among the rules to be respected for each metal level, let us quote inter alia:

- The minimum and maximum density per a given control window (in general between 20 and 80%)
- The minimum size of an enclosed area in a metallic layer (W_t , see Fig. 2.39)
- The minimum and maximum width of a metal stripe (W_c , see Fig. 2.39)

All these constraints have to be taken into account at the very beginning of every millimeter wave design. The ground planes contribute a lot to the metal density filling of the total area, but also the gap between the transmission line and these ground planes must respect the metal density rules. The maximum line width is also controlled by DRC rules, thus limiting the theoretical choices for the transmission lines design. On the other hand, the central signal stripe width is determined also by electromigration rules depending on the maximum current density to flow without damage.

An example of such transmission line on a digital BEOL in the 65nm node is presented in Fig. 2.40 and Fig. 2.41 [35]. Classical transmission line theory would have imposed that the ground planes should be built at the same level as the central conducting path, but in the presented case, in order to respect the metal density rules, the ground planes have also a vertical dimension in order to fill the space also for the lower level metals. These extra levels in the ground plane impose extra losses by fringing coupling. 3D electromagnetic field solvers [36], [37] offer good help for these studies.



Fig. 2.40 Coplanar transmission line in a 65nm CMOS digital BEOL (6 Metal layers) (© IEEE 2007).

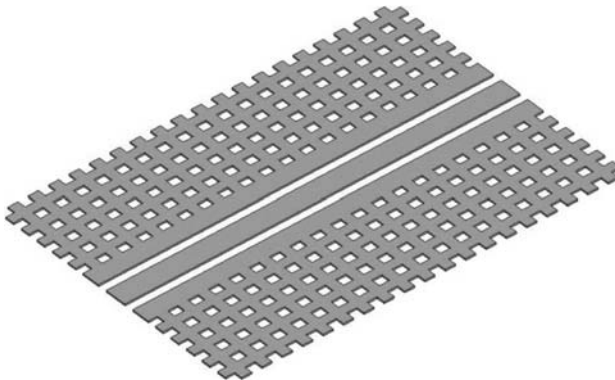


Fig. 2.41 Top view for the transmission line in Fig. 2.40.

The Back-End of Line of the recent technology nodes permits the integration of passive devices compliant with the design of mm-wave integrated products. A digital back-end on standard VLSI CMOS technologies implies rather lossy solutions for the passives, which have to be overcome by a higher overall current consumption and sometimes figures of merit (such as Noise Figure). In counterpoint, the use of a standard VLSI technology permits a good manufacturing flexibility. On the other hand, mm-wave dedicated back-end permits passive integration with very low losses, thus lower power consumption, but the overall production cost is raised by the use of an application dedicated process. An excellent compromise between these two alternatives may come from the use of VLSI CMOS SOI solutions using High Resistivity substrate, permitting a high integration level with reduced power consumption, for almost the same production cost as standard CMOS.

2.5 Conclusion

Silicon technologies are well-suited for implementing many millimeter wave applications. With the enormous and increasing worldwide silicon manufacturing capacity, reasonably-priced supply is guaranteed assuring the ability to satisfy demand on the consumer scale. Silicon transistor performance, measured by the conventional FOMs, is at least on par with that exhibited by III-V compound semiconductor devices. Furthermore, with the continued, intense focus of the commercial electronics industry, the relentless trend of increasing performance and integration density will doubtlessly continue. Indeed there are challenges to the designer. Design rule stricture and complexity, especially those involving metal pattern density, while certainly inconvenient to digital designers, will prove limiting to conventional millimeter design techniques. The lower voltage tolerance of silicon also presents a serious challenge to conventional design requiring innovation as illustrated by some excellent examples in later chapters. The extent to which the design community can overcome these challenges will determine the degree to which silicon dominates the emerging millimeter wave market.

2.6 Acknowledgements:

The following people are kindly acknowledged for their technical support to this material: from STMicroelectronics: B. Martineau, N. Seller, Ch. Raynaud (also with CEA/LETI), S. Pruvost, D. Gloria, F. Giancesello, P. Scheer, B. Sautreuil, A. Chantre, P. Chevalier, D. Belot, J.P. Schoellkopf and A. Artieri; from IBM: B. Jagannathan and A. Joseph.

References

1. International Technology Roadmap for Semiconductors, <http://www.itrs.net/>
2. ITRS 2006 Update, <http://www.itrs.net/Links/2006Update/2006UpdateFinal.htm>
3. Pekarik, J., et al., "Enabling RFCMOS solutions for emerging advanced applications", Euro. Gallium Arsenide and Other Semic. Appl. Symp., EGAAS, 2005, p. 29.
4. Semiconductor International Capacity Statistics; <http://www.sicas.info/>
5. J. Wiley, "Future challenges in computational lithography," *Solid-State Technology*, http://sst.pennnet.com/Articles/Article_Display.cfm?Section=HOME&ARTICLE_ID=254908&VERSION_NUM=2&p=5
6. Narasimha, S., et al., "High Performance 45-nm SOI Technology with Enhanced Strain, Porous Low-k BEOL, and Immersion Lithography," *Electron Devices Meeting*, 2006, pp. 1-4.
7. B. Santo, "Plans for Next-Gen Chips Imperiled," <http://www.spectrum.ieee.org/aug07/5394>
8. Rothschild, M., et al., "Recent Trends in Optical Lithography," *The Lincoln Lab Journal*, vol. 14, no. 2, 2003.
9. Kikkawa, T., "Advanced interconnect technologies for ULSI scaling," *6th International Conference on VLSI and CAD*, 1999, pp. 202-207.
10. Rim, K., et al., "Characteristics and device design of sub-100 nm strained Si N- and PMOS-FETs," *VLSI Technology*, 2002. Digest of Technical Papers, pp. 98-99.
11. Ito, S., et al., "Mechanical stress effect of etch-stop nitride and its impact on deep submicron transistor design," *Electron Devices Meeting*, 2000, IEDM Technical Digest, pp. 247-250.
12. Ota, K., et al., "Novel locally strained channel technique for high performance 55nm CMOS," *Electron Devices Meeting*, 2002, IEDM Technical Digest, pp. 27-30.
13. Ghani, T., et al., "A 90nm high volume manufacturing logic technology featuring novel 45nm gate length strained silicon CMOS transistors," *Electron Devices Meeting*, 2003, IEDM Technical Digest, pp. 11.6.1 - 11.6.3.
14. Horstmann, M., et al., "Integration and optimization of embedded-sige, compressive and tensile stressed liner films, and stress memorization in advanced SOI CMOS technologies," *Electron Devices Meeting*, 2005, IEDM Technical Digest, pp. 233-236.
15. Yang, M., et al., "On the integration of CMOS with hybrid crystal orientations," *VLSI Technology*, 2004, Digest of Technical Papers. pp. 160-161.
16. P. Chevalier, et al., "Advanced SiGe BiCMOS and CMOS platforms for Optical and Millimeter-Wave Integrated Circuits," *IEEE CSICS* 2006.
17. E. Gusev, et al., IBM J. Res. & Dev. vol. 50 No. 4/5 July/September 2006
18. R. Chau, et al., "High-k/Metal-Gate Stack and Its MOSFET Characteristics," *IEEE Electron Device Letters*, vol. 25, no.6, June 2004, p. 408.
19. S. Nuttinck, "Ultrathin-Body SOI Devices as a CMOS Technology Downscaling Option: RF Perspective," *IEEE Trans. On Electron Device*, vol. 53, no. 5, May 2006, p. 1193.
20. <http://www-03.ibm.com/press/us/en/pressrelease/20980.wss>,
<http://www.intel.com/pressroom/archive/releases/20070128comp.htm>
21. A. Cathelin, et al., "Design for Millimeter-wave Applications in silicon Technologies," *ESSCIRC 2007*, Munich, Sept. 2007.
22. P. Chevalier, "STMicronelectronics Foundry Services for mmW Frequencies," *ESSCIRC2006 Wireless Communications Workshop: "1Gbit/s+ wireless communications at 60GHz and beyond"*, Montreux, Sept. 2006.
23. G. Dambrine, et al., "What are the Limiting Parameters of Deep-Submicron MOSFETs for High Frequency Applications?," *IEEE Electron Device Letters*, vol. 24, no. 3, March 2003.
24. C. Raynaud, "Advanced SOI Technology for RF Applications," *2007 IEEE International SOI Conference Short Course*
25. B. Heydari, et al., "Low Power mm-Wave Components up to 104 GHz in 90nm CMOS," *ISSCC 2007*, Dig. of Tech. Papers.
26. B. Martineau, et al., "80 GHz Low Noise Amplifiers in 65nm CMOS SOI," *ESSCIRC 2007*, Munich, Sept. 2007.

27. B. Martineau, A. Cathelin, "Optimized MOS topology on CMOS process for millimeter wave design," *FR Patent Application*, September 2007.
28. F. Giancesello, "Evaluation de la technologie CMOS SOI Haute-Résistivité pour applications RF jusqu'en bande millimétrique," PhD Thesis presented at the Institut Polytechnique de Grenoble, France, October 2006
29. C. Raynaud et al., "Is SOI CMOS a promising Technology for SOCs in High Frequency range?," in *Proc. 207th ECS*, 2005, pp. 331-344.
30. T.O. Dickson, et al., "The Invariance of Characteristic Current Densities in Nanoscale MOS-FETs and Its Impact on Algorithmic Design Methodologies and Design Porting of Si(Ge) (Bi)CMOS High-Speed Building Blocks," *IEEE J. Solid-State Circuits*, vol. 41, no. 8, Aug. 2006.
31. P. Chevalier, et al., "High-Speed SiGe BiCMOS Technologies: 120-nm Status and End-of-Roadmap Challenges," 7th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems (SiRF07), Long Beach, Jan. 2007.
32. S. Pruvost, "Etude de faisabilité de circuits pour systèmes de communication en bande millimétrique, en technologie BiCMOS SiGeC 0.13 μm ," PhD Thesis presented at the Université des Sciences et Technologies de Lille, Nov. 25 2005.
33. F. Giancesello, et al., "65nm HR SOI CMOS Technology: emergence of Millimeter-Wave SoC," *RFIC 2007*, Honolulu, June 2007.
34. B. Martineau, "Potentialités des technologies CMOS 65nm SOI LP pour des applications en bande millimétrique," PhD Thesis to be presented at the Université des Sciences et Technologies de Lille, December 2007.
35. N. Seller, et al., "A 10GHz Distributed Voltage Controlled Oscillator for WLAN Application in a VLSI 65nm CMOS Process," *RFIC 2007*, Honolulu, June 2007.
36. http://eesof.tm.agilent.com/products/momentum_main.html
37. <http://www.ansoft.com/products/hf/hfss/>

Chapter 3

Design and Modeling of Active and Passive Devices

Ali M. Niknejad, Sohrab Emami, Chinh Doan, Babak Heydari, Mounir Bohsali

3.1 Passive Devices

3.1.1 *Transmission Lines*

Transmission lines play a critical role at mm-wave frequencies. Due to the relatively small wavelength, significantly long structures such as quarter wave can be realized on-chip. Transmission lines are suitable for high frequencies since there is no ambiguity in how one defines reference planes – since the signal and ground are always co-located, it's easy to connect a transmission line structure at any point in the circuit and predict the resulting reactance. Furthermore, the close physical proximity of the ground return current creates a dipole (multipole) radiation pattern that couples less energy to the substrate, which improves the quality factor of these devices. The well-defined ground return path also significantly reduces magnetic and electric field coupling to adjacent structures¹.

At mm-wave frequencies, the reactive elements needed for matching networks and resonators become increasingly small, requiring inductance values on the order of 50-250 pH. Given the quasi-transverse electromagnetic (quasi-TEM) mode of propagation, transmission lines are inherently scalable in length and are capable of realizing precise values of small reactances. Additionally, interconnect wiring can be modeled directly when implemented using transmission lines.

Four real (or two complex) parameters are needed to completely capture the properties of any quasi-TEM transmission line at a given frequency, ω_0 . A transmission line can be characterized by its equivalent frequency-dependent *RLGC* distributed circuit model (Fig. 3.1), which can be related to the characteristic impedance (Z_0) and complex propagation constant (γ) by [2]

Ali M. Niknejad, Sohrab Emami, Chinh Doan, Babak Heydari, Mounir Bohsali:
Berkeley Wireless Research Center, University of California, Berkeley

¹ Portions of this text are taken from [3], (© IEEE 2005)

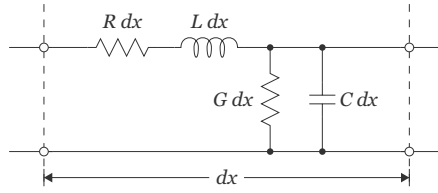


Fig. 3.1 Distributed RLGC lossy transmission line model [3] (© IEEE 2005).

$$Z_0 = \sqrt{\frac{R + j\omega_0 L}{G + j\omega_0 C}} \quad (3.1)$$

$$\gamma = \sqrt{(R + j\omega_0 L)(G + j\omega_0 C)} = \alpha + j\beta \quad (3.2)$$

For low-loss lines, the attenuation and phase constants, α and β , can be approximated as

$$\alpha \approx \frac{R}{2Z_0} + \frac{GZ_0}{2} \quad (3.3)$$

$$\beta = \frac{2\pi}{\lambda_g} = \omega_0 \sqrt{LC} \quad (3.4)$$

There is a major deficiency with the standard representation of Eq. 3.1 and 3.2. As seen in Eq. 3.3, the two different loss mechanisms are combined into one parameter, making it difficult to discern the relative importance of R and G . This is particularly relevant for transmission lines implemented on low-resistivity silicon, since they have a non-negligible G . To address this issue, the following four real parameters are proposed to characterize the line [3]

$$Z = \sqrt{L/C} \quad (3.5)$$

$$\lambda = \frac{2\pi}{\omega_0 \sqrt{LC}} \quad (3.6)$$

$$Q_L = \omega_0 L / R \quad (3.7)$$

$$Q_C = \omega_0 C / G \quad (3.8)$$

Notice that the two loss mechanisms are completely decoupled in Eq. 3.7 and 3.8. The first-order Taylor series expansions can be used to relate the characteristic impedance in Eq. 3.1 to the quantities in Eq. 3.5 - 3.8.

$$Z_0 \approx Z \left(1 + \frac{j}{2} \left(\frac{1}{Q_C} - \frac{1}{Q_L} \right) \right) \quad (3.9)$$

From Eq. 3.9, the sign of $\Im(Z_0)$ reveals which loss mechanism (Q_L or Q_C) is dominant for low-loss lines. If Z_0 is assumed to be real, this implicitly requires that $Q_L = Q_C$, which is not generally true.

Unlike transmission lines implemented on GaAs, where G is essentially zero, transmission lines implemented on low-resistivity silicon often have low capacitive quality-factors (Q_C) due to the substrate coupling. For transmission lines that store mostly magnetic energy, the inductive quality factor (Q_L) is the most critical parameter when determining the loss of the line, as opposed to the resonator quality factor or the attenuation constant.

3.1.1.1 Quality Factor Metric

Analogous to the case for lumped inductors [4], there are many different definitions for Q of a transmission line, where the applicability depends upon the intended function of the transmission line circuit. The most commonly used definition is the Q of the line when used as a resonator (Q_{res}),

$$Q_{res} = \omega_0 \frac{\text{avg. energy stored}}{\text{avg. power loss}} = \frac{\omega_0(W_m + W_e)}{P_L} \quad (3.10)$$

where ω_0 is the resonance frequency, W_m and W_e are the average magnetic and electric energy stored, and P_L is the average power dissipated in the line. Q_{res} can be related to the quantities introduced in Eq. 3.3 and 3.4 by [5]

$$Q_{res} \approx \frac{\beta}{2\alpha} = \frac{\pi}{\alpha\lambda_g} = \frac{\pi\sqrt{\epsilon_{eff}}}{\alpha\lambda_0} \quad (3.11)$$

where λ_0 is the free-space wavelength and ϵ_{eff} is the effective dielectric constant. Since ϵ_{eff} is determined mostly by the dielectric properties, and not the transmission line structure or dimensions, maximizing Q_{res} is roughly equivalent to minimizing the attenuation constant α . Using Eq. 3.11 along with Eq. 3.3 and 3.4, Q_{res} can be related very simply to the parameters in Eq. 3.7 and 3.8 by

$$\frac{1}{Q_{res}} \approx \frac{1}{Q_L} + \frac{1}{Q_C} \quad (3.12)$$

If the transmission line is forming a resonator, this is indeed a good definition to use. The Q_{res} quality factor is related to the half-power bandwidth of a quarter/half wave resonator. This is in direct analogy with lumped parallel/series resonant circuit, which is a good model for the resonator near resonance. But if the transmission line is employed for matching or interconnect, a different definition or metric is needed.

Another common metric is to simply compare the propagation loss versus frequency. While this is a fast and simple way to look at a transmission line, it ignores the fact that different lines can have a different rate of phase accumulation (such as slow wave structure) and a different characteristic impedance, which is important

in the design of matching networks. To first degree, then, the propagation constant should be compared versus β , rather than to frequency. Since β is proportional to frequency

$$\beta = \frac{2\pi}{\lambda_0} = \frac{\omega_0}{v} \quad (3.13)$$

the plot of α versus ω has the same shape with a frequency scaling factor related to the phase propagation velocity $v \leq c_0$. This definition is useful in cases where we desire a certain phase shift through a transmission line, and we are comparing two different transmission lines to see which one is more suitable.

When transmission lines are employed as inductors, or equivalently if $\ell \ll \lambda_0$, then we can approximate the impedance of a shorted section of a short transmission line

$$Z_i = Z_0 \tanh(\gamma \ell) \approx \gamma \ell Z_0 = (\alpha + j\beta) \ell Z_0 \quad (3.14)$$

which shows the line impedance is proportional to the characteristic impedance Z_0 . Assuming low-loss conditions so that $Z_0 \approx Z$ is real, then we have

$$Q = \frac{\Im(Z_i)}{\Re(Z_i)} \approx \frac{\beta}{\alpha} \quad (3.15)$$

For lumped circuits, a common metric is to compare the product of Q factor and the inductance value, since it's increasingly difficult to realize large reactance with high Q

$$Q \times L \approx \frac{\beta \ell \beta Z}{\alpha \omega} \quad (3.16)$$

since β is proportional to frequency, we have

$$Q \times L \approx \frac{\beta \ell}{\alpha v} Z = \frac{\beta}{\alpha} t_d Z \quad (3.17)$$

which shows that in addition to a large value of β/α , we also desire a transmission line with a large value of Z and a slow phase velocity, or equivalently a larger inductance per unit length. This argument can be further refined by defining more precisely an inductance quality factor.

Inductive Quality Factor

Transmission lines are often used to resonate with the intrinsic capacitance of the transistors (e.g., when used in matching networks). In this case, the line stores mostly magnetic energy, and it is therefore most appropriate to consider the power lost for a given amount of net reactive energy stored in the line, as opposed to the total stored energy [4]. Thus, for matching networks, the most meaningful metric is

$$Q = 2\omega_0 \frac{\text{net energy stored}}{\text{avg. power loss}} = \frac{2\omega_0(W_m - W_e)}{P_R + P_G} \quad (3.18)$$

where W_m and W_e are the average magnetic and electric energy stored, and P_R represent the “series” resistive losses and P_G the “shunt” conductive losses. Defining the series inductive and shunt capacitive Q we have [7]

$$Q_L = 2\omega_0 \frac{W_m}{P_R} \quad (3.19)$$

$$Q_C = 2\omega_0 \frac{W_e}{P_G} \quad (3.20)$$

we can express the overall Q as

$$\frac{1}{Q} = \frac{1}{\eta_L Q_L} + \frac{1}{\eta_C Q_C} \quad (3.21)$$

where

$$\eta_L = 1 - \frac{W_e}{W_m} \quad (3.22)$$

and

$$\eta_C = \frac{W_m}{W_e} - 1 \quad (3.23)$$

For a shorted transmission line, under the assumption of low loss, one can show that

$$W_m \approx \frac{1}{2} \frac{LV^{+2}\ell}{Z_0^2} \left(1 + \text{sinc}\left(\frac{4\pi\ell}{\lambda}\right) \right) \quad (3.24)$$

and

$$W_e \approx \frac{1}{2} CV^{+2}\ell \left(1 - \text{sinc}\left(\frac{4\pi\ell}{\lambda}\right) \right) \quad (3.25)$$

Thus we have

$$\frac{1}{\eta_L} = \frac{1}{2 \text{sinc}(\frac{4\pi\ell}{\lambda})} + \frac{1}{2} \quad (3.26)$$

and

$$\frac{1}{\eta_C} = \frac{1}{2 \text{sinc}(\frac{4\pi\ell}{\lambda})} - \frac{1}{2} \quad (3.27)$$

If the line is inductive (i.e., $W_m \gg W_e$), then $\eta_C \gg \eta_L$ and $Q_{net} \approx \eta_L Q_L$. The loss in the line is therefore almost completely determined by Q_L . For example, consider a shorted transmission line with $\ell < 0.1\lambda$. In this case, it can be shown that $\eta_C > 7.2\eta_L$, which greatly reduces the impact of the shunt losses on the inductive line. This is particularly important for integrated transmission lines on silicon, where the low-resistivity substrate causes Q_C to be non-negligible.

3.1.1.2 Microstrip and Coplanar Waveguide

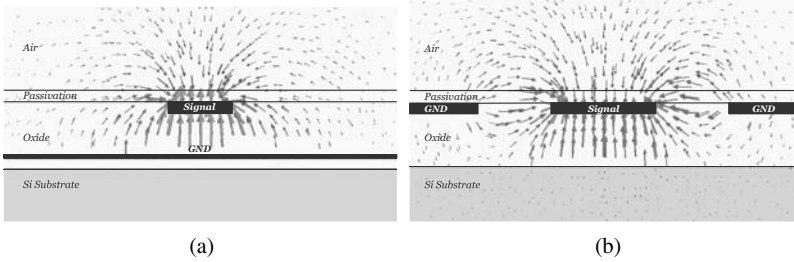


Fig. 3.2 Electric field distributions from 3-D EM simulations of (a) microstrip and (b) coplanar waveguide transmission lines [3] (© IEEE 2005).

Microstrip lines on silicon are typically implemented using the top-layer metal as the signal line, and the bottom-layer metal for the ground plane. Fig. 3.2a illustrates the effectiveness of the metal shield, with essentially no electric field penetration into the substrate. The shunt loss, G , is therefore due only to the loss tangent of the oxide, yielding a capacitive quality factor, Q_C , of around 30 at mm-wave frequencies (Fig. 3.3b). The biggest drawback to microstrip lines on standard CMOS is the close proximity of the ground plane to the signal line ($\sim 4\mu\text{m}$), yielding very small distributed inductance, L . This significantly degrades the inductive quality factor, Q_L (Fig. 3.3a).

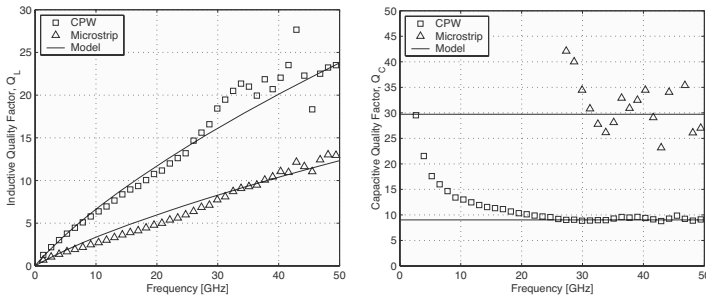


Fig. 3.3 Measured (markers) Q_L and Q_C for a coplanar waveguide and a microstrip line. The solid lines are for an empirical first-order model where the R variation is due only to skin effect and G is caused by a constant dielectric loss tangent [3] (© IEEE 2005).

Another option for on-chip transmission lines is the use of coplanar waveguides (CPWs) [10], [13], which are implemented with one signal line surrounded by two adjacent grounds (Fig. 3.2b). The signal width, W , can be used to minimize conductor

loss, while the signal-to-ground spacing, S , controls the Z_0 and the tradeoff between Q_L and Q_C . As an example, a CPW with $W = 10\ \mu\text{m}$ and $S = 7\ \mu\text{m}$ has a Z_0 of $59\ \Omega$ and a Q_L measured to be about double that of the microstrip (Fig. 3.3a). Therefore, CPW transmission lines are preferred in design of amplifiers (tuning and matching networks) due to their considerably higher Q_L compared to microstrip lines.

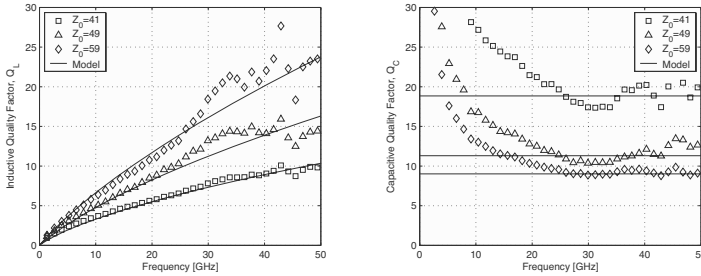


Fig. 3.4 Measured (markers) Q_L and Q_C for coplanar waveguides with varying geometries. The solid lines are for an empirical first-order model where the R variation is due only to skin effect and G is caused by a constant dielectric loss tangent [3] (© IEEE 2005).

By varying the signal-to-ground spacing, it is possible to design CPW lines to have either large Q_L and high-impedance ($S = 7\ \mu\text{m}$) or large Q_C and low-impedance ($S = 2\ \mu\text{m}$) (Fig. 3.4). On the other hand microstrip lines have, to first-order, constant Q_L and Q_C regardless of geometry. The larger the gap spacing S , the more fields penetrate into the substrate and cause additional shunt losses which lower Q_C . To first order, though, the series losses which contribute to Q_L remain relatively constant since these losses are dominated by the resistivity of the signal and ground plane. Second order effects, such as eddy current induced losses, change since the proximity of the ground plane determines the magnetic vector potential in the substrate, and hence the induced current magnitude. In a microstrip, to change the Z_0 , thinner lines must be used, which increases the series losses substantially.

Another important issue when designing with CPWs is the unwanted odd CPW mode, which arises because CPW lines inherently have three conductors. To suppress this parasitic propagation mode, the two grounds should be forced to the same potential [10]. In MMICs, this requires the availability of air bridge technology, which is costly and not supported by all foundries. Underpasses using a lower metal level in a modern CMOS process can be used to suppress this mode.

3.1.1.3 Transmission Line Modeling

Transmission lines are essentially two-dimensional structures, and TEM fields can in theory be predicted from static solutions in the transverse plane. In practice, though, intentional non-uniformity in the structure is introduced to dampen the presence of higher order TE/TM modes. Furthermore, the dielectric is non-uniform with different

low-K dielectric inter-layers in the typical stack-up. The abundance of metal layers also allows much more complicated structures to be realized, and this begs the question of the optimal structure that one should realize. For these reasons, prediction of performance requires extensive 3D simulation.

To achieve the highest level of accuracy for the transmission line models, a design-oriented modeling methodology, similar to [14], has been chosen for this work. The modeling approach is based on measured transmission line data and the models are optimized to fit most accurately at mm-wave frequencies. Scalable (in length) electrical models, which capture the high-level behavior of the lines, have been used and are supported in most simulators such as SpectreRF, ADS, and Eldo. The model parameters are easy to obtain from measured data or physical EM simulations since only a relatively small number of parameters are required to model the broadband performance of each transmission line: characteristic impedance, effective dielectric constant, attenuation constant, and loss tangent. A first-order frequency-dependent loss model is used. The model assumes that the conductor loss is only caused by the skin effect losses, and the shunt loss is due to a constant loss tangent. From Fig. 3.3, it can be seen that the losses are well-modeled.

Using simple electrical models has many advantages. The simulation time is very fast, and the models can be easily integrated into circuit simulators and optimizers. The transmission line models assume no coupling to adjacent structures. This assumption is justified since well-defined ground return path helps confine the magnetic and electric fields, and the close proximity of the adjacent grounds to the signal line help to minimize any second-order effects.

Full-wave electromagnetic simulation of transmission lines is useful for the design of new structures, especially when the line properties are not uniform with length. This occurs, for instance, when dummy fills are incorporated into the air structure or if regular bridging layers are used. Slow wave structures also employ non-uniform filling with filaments. In these cases full-wave 3D simulation is very time consuming. Accurate simulations are also difficult to obtain due to the high frequency material properties which are not typically characterized by the foundry. This includes dielectric loss of the various layers of dielectrics used in a modern CMOS process, the metal resistivity and roughness, and the precise substrate profile. The designer should strive to control the layout as much as possible by placing dummy fill blocking layers around the transmission lines and by carefully controlling the layout where the electromagnetic fields have an appreciable magnitude. Attempts to simulate these structures using commercial tools directly using only knowledge of the substrate conductivity and assuming lossless dielectrics has been unsuccessful. In order to match measurements and simulations, a custom tailored substrate profile with various dielectric loss parameters is needed in order to fit a wide array of measured data. Given a predictive simulation setup allows much greater flexibility in the design of mm-wave passive structures.

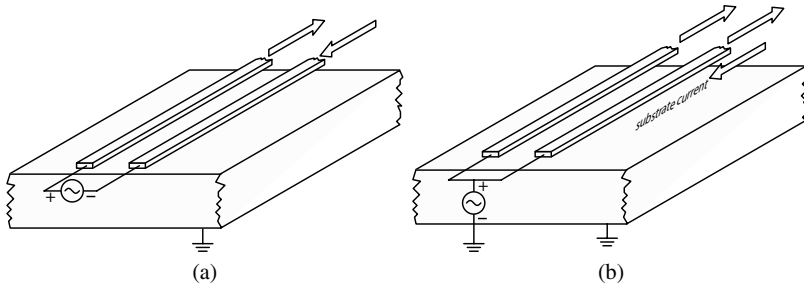


Fig. 3.5 A differential transmission line excited (a) in odd mode and (b) in even mode. Arrows indicate direction of current flow.

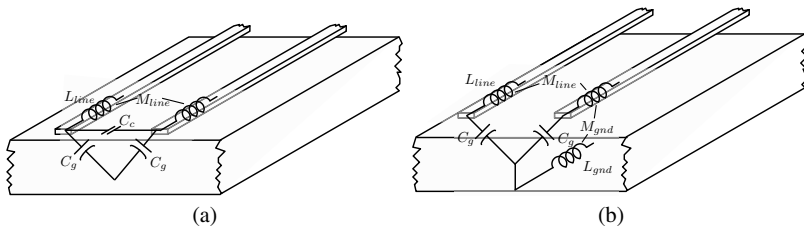


Fig. 3.6 Distributed circuit models for the (a) odd and (b) even modes of propagation.

3.1.1.4 Differential Lines

Since differential circuits are common in ICs, differential transmission lines are very convenient structures. As shown in Fig. 3.5, a differential pair can excite equal and opposite currents on a differential two-wire transmission line. Clearly there are two modes of propagation on the line, the “odd” mode which is usually the desired mode, where the voltage on each line is out of phase with the other voltage, and the “even” mode, where the voltages have equal magnitude. In this case the voltage is excited with reference to the substrate potential and the return current path flows through the substrate. We may therefore think of the even mode as a microstrip occurring primarily between the two signal lines (effectively connected in parallel) and the substrate ground. The two modes have different characteristic impedance and different phase velocity. The characteristic impedance of the odd mode is given by analyzing the equivalent half circuit shown in Fig. 3.6a

$$Z_{oo} = \sqrt{\frac{L_o}{2C_c + C_g}} \quad (3.28)$$

where L_o is the inductance per unit length defined by integrating the magnetic field between the lines, and C_c and C_g are the ground and mutual capacitance per unit

length. The odd mode inductance can also be obtained from the partial inductance of each line (assumed identical) by including the mutual inductance between the lines, $L_o = L_{line} - M_{lines}$, which subtracts due to the opposite direction of current flow. We are implicitly assuming that no current flows in the ground plane for this configuration due to the balanced excitation. Even though induced currents flow in the ground, the net ground current is zero. The capacitance C_c is doubled due to Miller effect. For the even mode, the return current flows in the substrate so the inductance per unit length L_e is calculated by integrating in the vertical plane from the signal line to the substrate. Since the two lines are at the same potential, the capacitance C_c has no effect and the equivalent half circuit can be calculated as shown in Fig. 3.6b. The characteristic impedance is given by

$$Z_{oe} = \sqrt{\frac{L_e}{C_g}} \quad (3.29)$$

where L_e is the even mode inductance per unit length or calculated through the partial inductance terms by $L_e = L_{line} + M_{line} + L_{gnd} - M_{gnd}$. We see that if $L_e \sim L_o$, the common mode impedance is higher. In fact, we implicitly assumed that the return current flows through the substrate, which is not necessarily the case. To understand this point, imagine that there is an alternative path to ground through metallization between the source and load, such as other ground planes, interconnect, or the substrate. Since current flows the path of least impedance, at low frequencies the current may flow through the distant return path increasing the even mode inductance substantially, or $L_e \gg L_o$, which makes the even mode impedance large. At high frequencies, though, the current will flow underneath the structure – even though the path is more resistive, it has lower inductance. This makes the even mode very lossy as well, which is reflected in a higher even mode propagation loss $\alpha_e > \alpha_o$.

The phase velocity of the two modes can differ due to the different field configurations. In the odd mode the fields reside substantially between the conductors with very little leakage into the substrate and air, which results in $\epsilon_o \approx \epsilon_{oxide}$. On the other hand, for the even mode, there is significant leakage into the substrate, which decreases the phase velocity (since $\epsilon_{Si} \approx 3\epsilon_{ox}$).

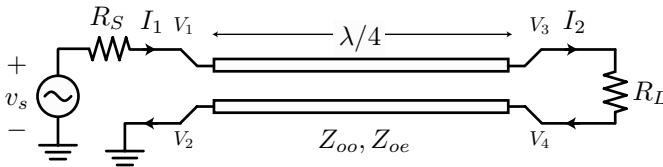


Fig. 3.7 A transmission line converts an unbalanced signal to a balanced signal.

One particularly good application for a differential transmission line is to convert a single-ended signal into a differential signal. The circuit shown in Fig. 3.7, is known as a transmission line. Nominally the circuit should be a quarter wavelength.

Intuitively we can see that the single ended port excites an equal component of the even and odd modes. To see this suppose that $V_e = V_o = V_0/2$, then

$$V_1 = (V_e + V_o) = V_0 \quad (3.30)$$

and

$$V_2 = V_e - V_o = 0 \quad (3.31)$$

At the load end of the line, we wish to have a fully balanced signal, or $V_e(\ell) = 0$. A balanced load is an open circuit boundary condition for the even mode, which means the line is highly imbalanced. Since the line is excited with a common mode voltage at the source end, a standing wave pattern emerges on the line due to the open circuit reflection. Since the line is exactly $\lambda/4$, the voltage profile on the line is monotonic. Given that the source is driven with a non-zero common mode, the common mode voltage must decay to zero at the load, resulting in a perfect transformer. To propagate all the energy of the source into the odd mode, we should terminate the line in the characteristic impedance of the odd mode $Z_L = Z_{oo}$ to avoid reflections.

3.1.1.5 Slow Wave Structures

One way to avoid substrate leakage in a transmission line is to place a solid shield under the conductors. For the differential mode, the return current flows through the the conductors, and not the ground, but due to the changing magnetic field, eddy currents are induced on the ground plane which lowers the inductance per unit length. The net current in the ground plane is zero since the induced currents have opposite direction, and in essence we see the ground current flows in a loop (typical of “eddy currents”). Commensurate with the decrease in inductance, the capacitance per unit length also increases due to the increase in C_g , and in fact this increase must exactly balance the decrease in inductance to maintain constant phase velocity.

If we break the shield in such a way as to limit current flow along the length of the line, then we see that the inductance per unit length remains intact while the capacitance per unit length increases. This is a “slow wave” transmission line, since the phase velocity is lowered. One way to accomplish a slow wave structure shown in Fig. 3.8. Here, conductive strips are placed beneath the two wires to increase the capacitance per unit length without altering the inductance per unit length. In a modern CMOS process, a very high density of filaments can be used to increase the capacitance significantly. As a result, the quality factor of a resonator improves since the dielectric losses are minimized in the transmission line. Researchers have demonstrated quality factors as high as 40 for such integrated resonators in CMOS [28].

Another slow wave structure is a varactor loaded line, where MOS capacitors are placed periodically along the line. The transmission line characteristic impedance and phase velocity are now a function of the bias voltage, which has application in phase shifters or dynamic tuning elements.

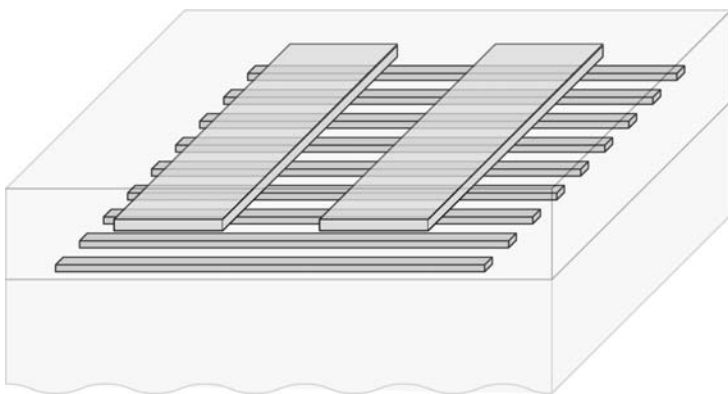


Fig. 3.8 A balanced slow wave two-conductor transmission line.

3.1.2 Inductors

Inductors are used extensively to tune out parasitics and for matching at RF frequencies and take the form of spiral inductors or short sections of transmission lines. While lumped inductors are generally smaller than distributed circuit equivalents, the advantages of distributed circuit elements include better prediction, scalability, and improved isolation. The value of the inductance of a shorted transmission line depends only on the length, Z_0 , and propagation constant $\gamma = \alpha + j\beta$. If a transmission line is well characterized, then any arbitrary inductance can be synthesized by varying the length of the line. Furthermore, there are no “leads” that add parasitic inductance. A common example is the inductance degeneration employed in amplifiers, which places an inductor from the source to ground. The value of the degeneration inductance is ambiguous, though, as the terminals of the inductor are widely separated. The “return path” flows through ground and through the gate (or base) elements, forming a series resonant circuit. The proximity of the ground and the placement of the gate (base) inductors therefore have an appreciable impact in the value of inductance, requiring careful analysis. Since ground return currents flow intrinsically in the transmission line, there is no ambiguity as to how currents flow in the structure and the proximity of the ground plane minimizes induced currents in the substrate.

In many situations, though, a lumped inductor can be realized with higher quality factor. For typically small inductors with $L \sim 100\text{pH}$, simple ring inductors, as shown in the inset of Fig. 3.10, are sufficient. In fact, we can draw an analogy between a ring inductor and a differential transmission line. To realize high Z_0 transmission line, we increase the gap spacing to increase the inductance per unit length while reducing the odd mode capacitance per unit length. But in order to short the line with low inductance, we prefer to bend the end of the line and reduce the gap spacing. Furthermore, to present close leads to connect to a capacitor or transistor, we prefer

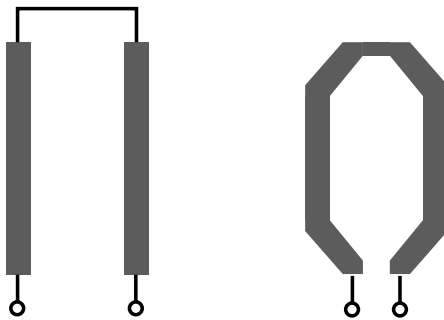


Fig. 3.9 (a) A high Z_0 line is formed by using a large gap spacing in the differential line. (b) A differential transmission takes the form of a loop when practical leads are formed.

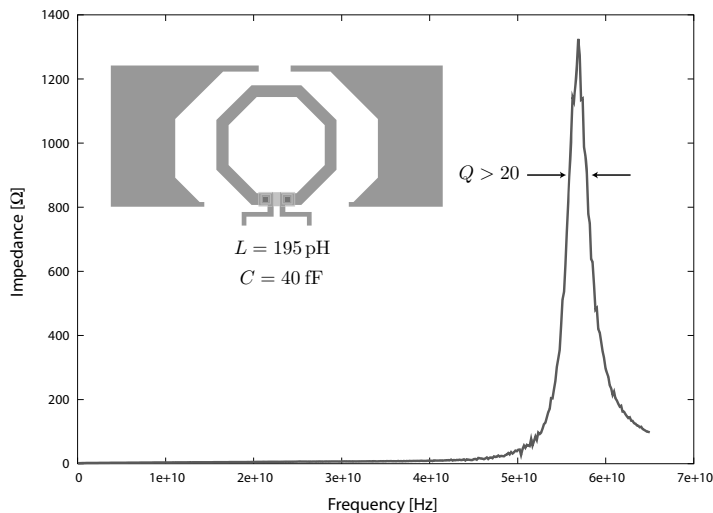


Fig. 3.10 Measured resonance of a mm-wave tank formed by a ring inductor and MIM capacitors.

to also bend the front of the line, as shown in Fig. 3.9b. Clearly this structure very much resembles a ring inductor, showing a close connection between transmission lines and ring inductors.

A ring inductor has been designed and fabricated in a Si RF process. To accurately measure the Q factor, a parallel resonant circuit is formed with MIM capacitors. In this way, the Q factor is insensitive to small lead inductance and resistance since at resonance the circuit forms a “open” or large impedance circuit. The measured resonance curve is shown in Fig. 3.10. The resonator occupies an area of $150^2 \mu\text{m}^2$ and has a loaded measured quality factor over 20. The inductor has been simulated and measured separately and the Q factor is between 30-40, which shows that the MIM capacitors have comparable Q . An electric shield surrounds the structure to minimize the coupling to Si substrate. This “open” shield structure is used to minimize eddy current flow.

3.1.3 Capacitors

In mm-wave circuits, capacitors in matching networks and resonators are usually realized as transmission lines whereas lumped capacitors are employed for AC coupling and DC bypass. Because lumped capacitors reside in the signal path, optimizing their design and layout as well as modeling their behavior accurately over a broad frequency range is of a crucial importance. For both AC coupling and DC bypass, an ideal structure would have infinite impedance at DC and zero impedance at the frequency of interest. Thus, large high quality factor capacitors with self-resonance frequency situated above the frequency of interest are desirable. Even if the capacitor self-resonates below the frequency of interest, as long as its impedance at the frequency of interest is sufficiently low (even if inductive), it can still be used effectively for both AC coupling and DC bypass. If the impedance is not sufficiently low, then it becomes important to accurately model the behavior so that the parasitics are absorbed into the matching network.

3.1.3.1 Coupling Capacitors

“Finger” (comb) capacitors, sometimes referred to as “MOM” (metal-oxide-metal) capacitors, can be easily fabricated in any modern CMOS process due to the abundance and high packing density of the metal layers. Finger capacitors are composed of a large number of parallel fingers connected to either port of the device as shown in Fig. 3.11. Each finger consists of all available metal layers stacked on top of each other in order to decrease its resistance which increasing both its quality factor and the self-resonance frequency. The metal layers are connected together through the maximum number of vias possible in order to take advantage of the additional via-to-via capacitance. If a lower parasitic substrate capacitance is

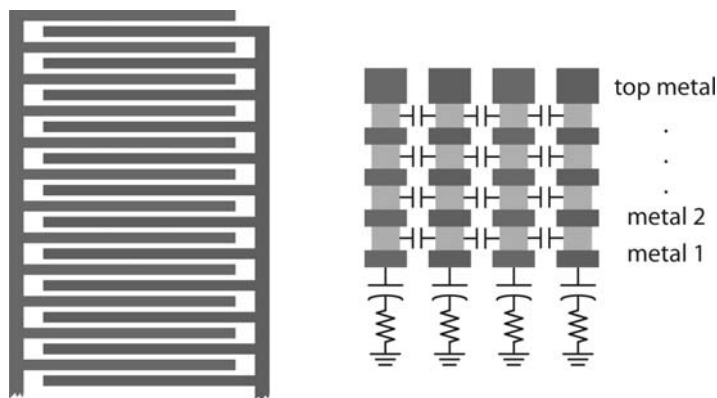


Fig. 3.11 Layout geometry of a “finger” or “comb” capacitor structure.

desired, the lowest metal layer can be omitted, but this negatively affects the quality factor of the structure. At each port, a wide slab consisting of a small 45° taper is used to connect the fingers. In order to realize a given capacitance, either a large number of short fingers in parallel or a small number of long fingers can be used. The former structure results in smaller finger resistance and inductance but larger lead parasitics due to the transition from the signal lines to the capacitor plates. The latter structure exhibits the opposite tradeoff. This tradeoff and the optimum geometry are dependant on the overall capacitor size.

Given that no new layers are needed in the fabrication of MOM capacitors, they are less expensive than Metal-Insulator-Metal (MIM) capacitors, which employ a sandwich structure with a thin high-K dielectric. If the density of MIM capacitors are sufficiently larger than “finger” structures, then the area savings may outweigh the additional fabrication costs.

3.1.3.2 Modeling

The symmetric capacitor equivalent circuit shown in Fig. 3.12 is used to model AC coupling capacitors over a broad frequency range (from DC to 100 GHz). R_1 , C_1 , and L_1 model the finger capacitor while C_2 models the oxide, and C_3 and R_2 model the substrate.

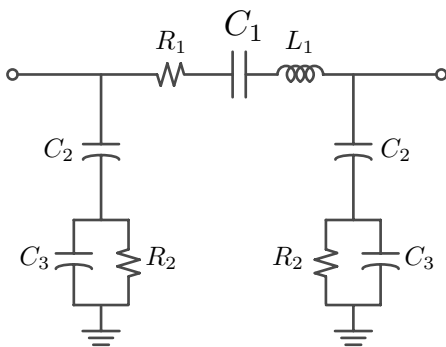


Fig. 3.12 Symmetric layout AC coupling capacitor model.

The circuit components values are optimized in order to match the model simulated network parameters to the measured network parameters. If only the S -parameters are used in the optimization, it is very difficult to extract the value of R_1 . In an S -parameter measurements, the 50Ω resistance is connected in series with each port, overshadowing R_1 which is usually in the order of hundreds of $m\Omega$'s. For this reason, Y and Z parameters are derived from the measured S parameters and used simultaneously in the optimization. Because the branch composed of R_1 , C_1 , and L_1 has a much lower impedance than the branches composed of $C_{2,3}$ and R_2 , the

former dominates the Y parameters and the latter dominate the Z parameters. Thus, a simultaneous Y and Z parameter optimization is used.

3.1.3.3 De-Coupling Capacitors

Very large capacitors (e.g. 2-5 pF) are desirable for DC bypass applications. However, scaling the AC coupling capacitors described above would result in considerably lower self-resonance frequencies that render the structure unusable as a DC decoupling capacitor. In a bypass capacitor, one port is always connected to ground (or supply) and therefore the entire bottom plate of the capacitor can be “shorted” and used as a low inductance path to ground. Having a low impedance ground plane considerably lowers the series inductance and thus increases the self-resonance frequency of the capacitor enabling large capacitors to be functional at mm-wave frequencies.

Other techniques in the layout of the MOM capacitor have appeared in the literature. The layout incorporating a mesh structure is shown in Fig. 3.13 uses fingers at right angles, resulting in both a fringe and overlap capacitance. The layers are interconnected multi-layer meshes that are interlocked together.

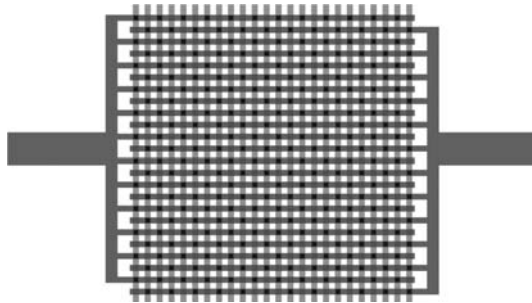


Fig. 3.13 The layout of a “mesh” capacitor structure.

3.1.3.4 Varactors

Varactors are key building blocks in many circuits, particularly in voltage-controlled oscillators. The variation in capacitance versus tuning voltage and the quality factor are important metrics. The quality factor of capacitors drops inversely with frequency due to any series resistance, and this is a severe limitation for on-chip structures. In Fig. 3.14 we plot the quality factor of a MOS varactor. Evidently the Q factor is very low in the mm-wave frequencies. To first order the MOS varactor Q factor is independent of the device size since the channel resistance and capacitance scale with the W of the structure. In inversion, the maximum capacitance is given by the effective oxide thickness

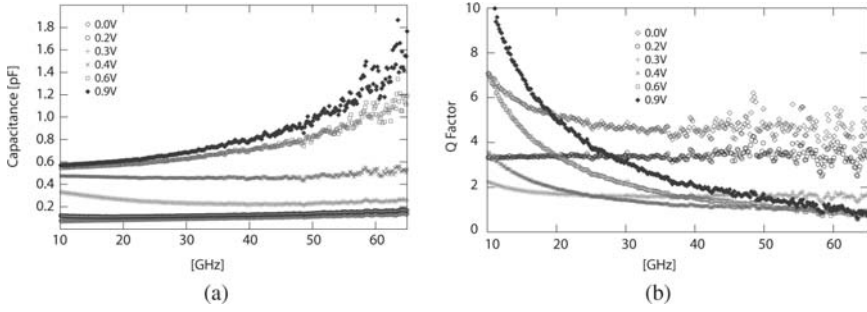


Fig. 3.14 Measured (a) capacitance and (b) quality factor of a MOS varactor ($N = 5 \times W = 40$) as a function of bias voltage.

$$C_v = W \cdot L \cdot C_{ox} \quad (3.32)$$

whereas the resistance is proportional to the gate and channel resistance. The channel conductance contribution is given by

$$G_{ch} \propto \frac{W}{L} \mu C_{ox} (V_{GS} - V_T) \quad (3.33)$$

so that the Q is independent of W and C_{ox}

$$Q = \frac{G_{ch}}{\omega C_v} \propto \frac{\mu(V_{GS} - V_T)}{L^2} \quad (3.34)$$

necessitating one to employ the shortest channel length and highest mobility possible. For this reason electrons are the preferred carriers in the channel of a MOS varactor. The low Q factor of mm-wave varactors is a major limitation since a poor quality VCO tank directly translates into poor phase noise performance unless tuning range is sacrificed. The difficulty in producing a wide tuning VCO in the mm-wave band favors transceiver architectures with lower frequency VCOs and frequency multipliers.

3.1.4 Transformers

Transformers have found many unique applications in RF CMOS and SiGe circuits. For mm-wave circuits, they can be used for driving balanced circuits, such as a Gilbert cell mixer, for AC coupling, and for impedance matching. Their compact size, the ability to AC couple and provide DC voltages, makes these devices very convenient in the design of mm-wave circuits.

Two coupled inductors, for instance two ring inductors shown in Fig. 3.15a, form a simple transformer. The measured insertion loss of the structure is shown in Fig. 3.15b. The structure exhibits about 1 dB of loss when the device is loaded

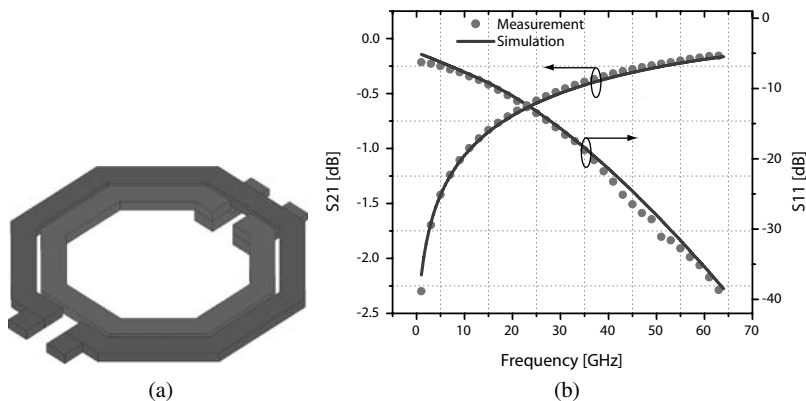


Fig. 3.15 (a) The layout of a simple 1:1 transformer. (b) The measured insertion loss and matching of the structure.

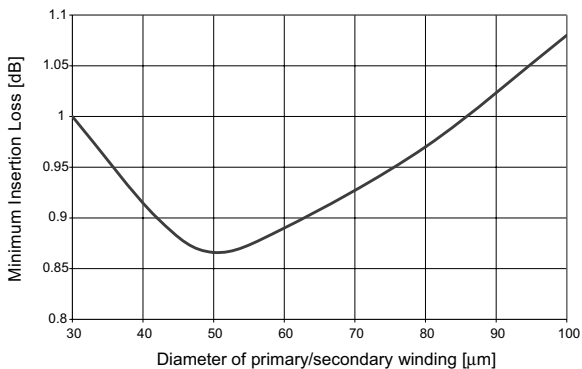


Fig. 3.16 The simulated insertion loss of a transformer as a function of outer diameter.

with the optimal source/load impedance. In many applications, the load and source capacitance can be absorbed into the transformer. More complicated geometries are possible for the realization of higher turn ratios [29]. Due to the high frequency of operation, the area of these structures must be scaled down to realize high self-resonant frequencies. For a given frequency of operation, one can determine the optimal size through EM simulation, as shown in Fig. 3.16, where the outer radius is swept to find the optimal radius. Clearly a larger inductor is beneficial since the winding resistance can be thought of as a shunt resistor to ground of value ωLQ , which should be as large as possible. A larger inductor has lower series resistance but higher substrate losses. There is thus an optimal radius that produces the best $Q \times L$ product for the windings. An example design will be highlighted in the next chapter, where the LO port of a Gilbert cell mixer is driven with an on-chip balun.

3.1.5 Resonators

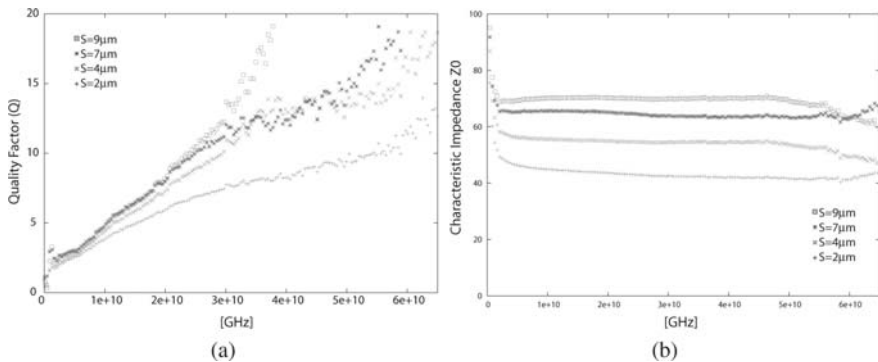


Fig. 3.17 The measured (a) quality factor $Q = \beta/2\alpha$ and (b) characteristic impedance of a CPW transmission line resonators with varying gap spacing: $S = 2\mu\text{m}$, $4\mu\text{m}$, $7\mu\text{m}$, $9\mu\text{m}$.

Resonators are useful building blocks in filters and in oscillators. We have already demonstrated that simple LC tanks can be used as resonators in the mm-wave regime, and the behavior is closely captured by a lumped equivalent circuit. As long as the structure is physically small as to keep the self-resonant frequency larger than the mm-wave band of interest, there is no problem in thinking of a small ring or spiral as a lumped inductor.

An alternative to the ring is a resonant transmission line. For instance a shorted section of transmission line will self-resonate at every odd multiple of the quarter wave frequency. As already noted, the quality factor of the line is given by $Q = \beta/2\alpha$. The Q of several CPW transmission line resonators is plotted in Fig. 3.17a based on the measured S -parameters of the transmission lines. The various electrical and magnetic loss mechanisms have already been discussed but it is noteworthy that at resonance the electric and magnetic field energy are equal, and so both contributions of loss are important. It is clear that using a wide gap spacing increases the Q , but the improvement in Q is limited since the electric field leaks into the substrate and introduces significant loss at higher frequencies. The characteristic impedance Z_0 is shown in Fig. 3.17b, where the Z_0 increases in step with the gap spacing as more magnetic energy is stored while the electrical energy storage drops. It is noteworthy again that the improvement in Z_0 is limited by the substrate, especially above 40 GHz. We often desire to alter the self-resonant frequency of a resonator, in particular in the design of voltage-controlled oscillators (VCOs). We may use a varactor loaded LC tank or even a slow wave varactor loaded transmission line. The tank inductor can be realized by a transmission line shorter than $\lambda/4$.

It is interesting to note that the optimal quarter wave line is not a uniform line [15], since the voltage/current profile and the corresponding electrical/magnetic energy

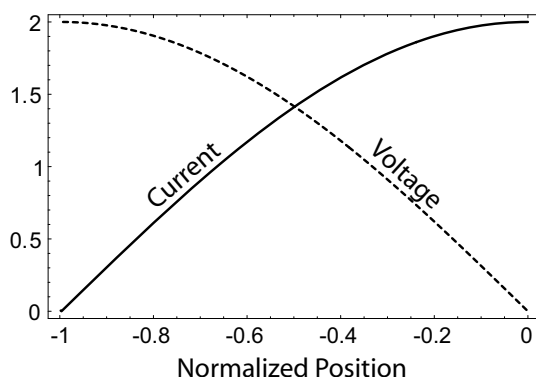


Fig. 3.18 The standing wave pattern for voltage and current along a quarter wave short-circuited line in resonance.

storage varies along the line. A differential stripline shorted at one end exhibits the standing wave mode shown in Fig. 3.18 at resonance. At the shorted end of the line, the voltage is at a minimum and the current at a maximum, so the losses at this point come only from the series resistance of the metal line. Conversely, at the driven end, the voltage is at a maximum and the current at a minimum, so the losses at this point come only from the shunt conductance between the differential lines. This phenomenon can be exploited to lower the losses of the resonator and thus raise the quality factor [15]. At the shorted end of the line, the metal conductors can be made wider to reduce the series resistance without having to worry about the increase in shunt conductance. The gap spacing can also be made wide since the voltage difference between the lines is small. Similarly, at the driven end of the line, the conductors can be made very narrow, lowering the conductance, without having to worry about the increase in series resistance. The gap spacing should be small in order to minimize field leakage into the substrate, since the fields are strongest at this end of the line. Along the rest of the line, shunt conductance and series resistance can be traded off to achieve a tapered line with much lower losses than the optimum uniform line. The optimum taper shape would thus end up looking similar to Fig. 3.19a. From this analysis we draw an important conclusion that the optimal resonator is in fact very similar to an LC circuit. To see this notice that circuit can be partitioned into a low Z_0 transmission line in series with a “loop” inductor, as shown in Fig. 3.19b. It’s clear that the low Z_0 section is mainly contributing capacitance to the tank and can be replaced by a more energy efficient MIM or MOM structure described earlier. With this substitution, we clearly identify this circuit as a an optimal LC tank.

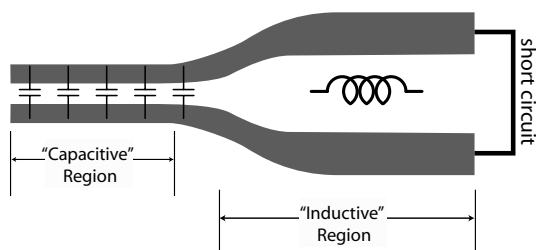


Fig. 3.19 The optimal profile for a transmission line tapers from a low Z_0 capacitive section to a high Z_0 inductive region. This taper profile resembles a lumped LC tank.

3.2 Active Devices

3.2.1 Modeling

Circuit designers typically use “compact” models for design and verification of Si integrated circuits. Compact models are the interface between the technology and the design. A circuit designer learns about a process by experimenting with the compact model, rather than running expensive and time-consuming experiments. Therefore compact models should be scalable with geometry and accurate across a wide temperature and bias voltage range.

Several good compact models have been developed for digital, analog, and RF applications [16] [17] [18] [19]. These models use a combination of physical and empirical methods to develop general equations, usually a large number of them, to describe the behavior of the device. Several parameters are embedded in each equation in order to customize it for the desired technology or device. These parameters are necessarily determined through complicated curve fitting procedures (parameter extraction) and shape the familiar model card for circuit designers. Most compact models have the advantage of describing the behavior of the device in all regions of operation at the same time. Furthermore, they provide small and large signal analysis as well as noise analysis. They also operate over a fair range for geometry, width and length of the device, over which the extracted parameters are valid. This generality however comes with an accuracy penalty if the model is used over a bias or geometry range outside of the extraction process. Moreover, the core equations in most compact models have been derived under quasi-static assumptions. This, together with the fact that most of available extracted parameters are done so for low-GHz frequency purposes, make these compact models less desirable and inaccurate for millimeter wave applications.

There are two main reasons for this inaccuracy: First of all, as mentioned before, the fact that the parameter extraction has been done in lower frequencies makes the extrapolation to mm-wave frequencies problematic. Some of device mechanisms that are not well captured at low frequencies, and naturally not modeled properly, have considerable effect on the performance of the device in higher frequencies, resulting

in some inaccuracy. The substrate network including capacitances and resistances is an example of such an effect. The inaccuracy due to this effect could be addressed by increasing the frequency range for parameter extraction process.

The second reason for the error in modeling is due to the important role of the device layout and this is more difficult to address. The device interconnections to the outside world introduce small inductors, resistors and capacitors to the model. These small components are generally negligible at lower frequencies making the device model more or less independent of layout. These components however change and in fact dominate the performance of the device as the frequency increases. This makes it crucial to include these parasitics into the model. An accurate prediction of these parasitic requires full-wave electromagnetic simulation, which is difficult and expensive to perform in a compact model, and therefore it is best to capture these parasitic through experimental means. Existing compact models can be used as the core for a hybrid customized mm-wave model. In essence, each small finger of the transistor is modeled with the “intrinsic” transistor model based on the quasi-static equations whereas the interconnection between the fingers and the interconnect are captured by electromagnetic simulation and experimental techniques.

Given the difficulties in modeling the device, one may be tempted to work directly with measured data. In traditional microwave design the common approach is to use measured S -parameter data for a specific device and treat to the transistor as a black box. This approach is very accurate in nature and accounts for all parasitics and distributed effects associated with the device and the layout. While this method is sufficient for small-signal circuit design applications, the accuracy of the S -parameter data hinges on reliable measurements of the device and de-embedding structures. As a result, the accuracy of the method may deteriorate for very high frequencies, both due to limited accuracy of test equipment and due de-embedding errors. Besides, this method is not suitable for simulation of any non-linear circuit such as mixers or oscillators or the assessment of the dynamic range of amplifiers. Moreover, because the transistor is treated as a black-box, there is no physical insight for improving the device performance or layout. Due to these issues, for mm-wave application, a combination of RF methodology and traditional microwave method is preferred even for small-signal applications.

3.2.2 Active Device Design

As was mentioned in the previous section, device performance in mm-wave frequencies is impacted by the device layout. Unlike low frequency circuit design in which the device design is absolutely in the realm of process engineers, here the circuit designer could – and should – alter the device performance drastically by changing the device layout. This enables the designer to layout the device based on the performance metric which is more important in a specific application. f_{max} , for instance, which is an indicator of the speed of the transistor, has been reported to vary from

80 GHz to up to 280 GHz for an identical CMOS 90nm process, mainly due to differences in the layout [20][21]².

The most common and important performance metrics for mm-wave devices include the f_T , f_{max} , Maximum Stable Gain (MSG), Maximum Unilateral Gain (U), output power (P_{out}), drain efficiency (η_d) and minimum noise figure (F_{min}). The selection of the appropriate metric as an optimization target depends on the specific application of the device.

If a transistor is modeled as a linear two-port network and represented with y -parameters, a common figure-of-merit used to characterize the active device is Mason's unilateral gain [6],

$$U = \frac{1}{4} \frac{|y_{21} - y_{12}|^2}{g_{11}g_{22} - g_{12}g_{21}} \quad (3.35)$$

where $g_{ij} = \Re(y_{ij})$. One property of the unilateral gain, as proved by Mason, is that the addition of lossless capacitors or inductors around the device will not change U . The conditions for a generic two-port to be active are complex [22], but for nearly all CMOS transistors, a simpler test for activity can be used, viz.

$$U > 1 \quad (3.36)$$

For most practical cases, U is a monotonically decreasing function of frequency, and the device becomes passive where $U = 1$. Since f_{max} is often beyond the frequency capabilities of the measurement system, it is common practice to use low-frequency measurements of U and report f_{max} as the extrapolated frequency where $U = 1$ (assuming a 20 dB/decade slope). At frequencies approaching f_{max} , U often drops at a rate much faster than 20 dB/decade (Fig. 3.20). Therefore, it is imperative that U is measured and modeled as closely to the targeted operating frequency of the circuit as possible to minimize errors associated with extrapolation.

Transistor Layout

Consider a multi-finger transistor composed of N_F identical devices in parallel. If ideal connections between all of the devices are assumed, the port admittance matrix of the composite device is $Y_{TOT} = N_F Y_{Finger}$. Since all of the admittances are scaled by N_F , it is easy to see from Eq. 3.35 that

$$U_{TOT} = U_{Finger} \quad (3.37)$$

Therefore, f_{max} of the multi-finger transistor is identical to f_{max} of the individual fingers. Even if non-ideal interconnect is modeled, the dominant effect for transistors much smaller than a wavelength is additional low-loss parasitic capacitance and inductance, which have a negligible effect on U , and thus f_{max} is also not greatly

² Portions of this text are taken from [3], (© IEEE 2005)

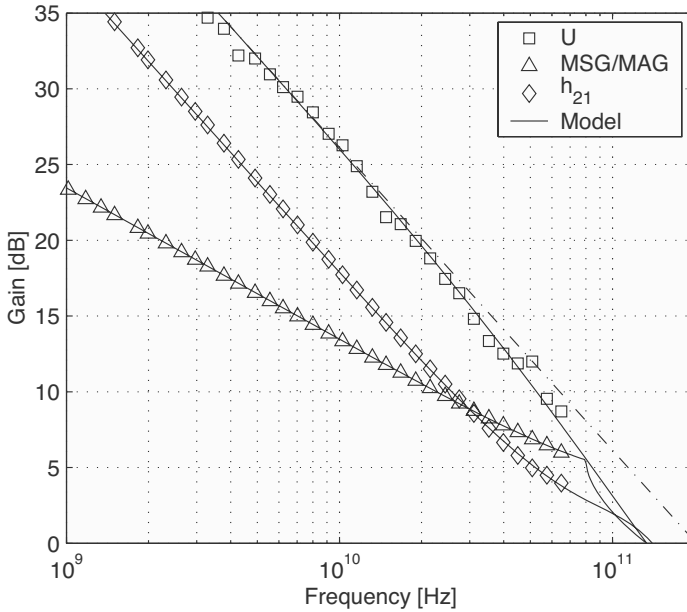


Fig. 3.20 Measured (markers) and modeled (solid lines) unilateral gain U , maximum stable gain MSG, maximum available gain MAG , and current gain h_{21} , for a typical NMOS device ($W/L = 100 \times 1 \mu\text{m}/0.13 \mu\text{m}$, $I_{DS}/W = 300 \mu\text{A}/\mu\text{m}$, $V_{DS} = 1.2 \text{ V}$). The maximum frequency of oscillation, based upon the device circuit model, is $f_{max} = 135 \text{ GHz}$. This is much lower than the value of 200 GHz attained if a 20 dB/decade slope is assumed [3] (© IEEE 2005).

affected. In the next section we will revisit this issue when we consider transistor layouts that become appreciably large to invalidate this assumption.

Therefore, for now it is sufficient to only consider the optimal layout for a single finger. The physical layout of a single finger is shown in Fig. 3.21, along with a physical model depicting the dominant high-frequency loss mechanisms. As mentioned, f_{max} is limited by resistive losses, the most significant being the gate resistance (R_G), series source/drain resistances (R_S, R_D), non-quasi-static channel resistance (r_{nqs}), and resistive substrate network (R_{sb} , R_{db} and R_{bb}) [23]. The series source and drain resistances are dominated by the intrinsic spreading resistance and are less sensitive to layout changes (such as by increasing the length of the source/drain regions). The non-quasi-static channel resistance models the effective increase in gate resistance due to the finite channel charging time, and can be shown to be inversely proportional to g_m [24]. The gate resistance, R_G , accounts for the distributed RC nature of the polysilicon gate, and can be approximated using a lumped resistor with value [8] [9]

$$R_G = \frac{R_{poly} W_F}{3n^2 L} \quad (3.38)$$

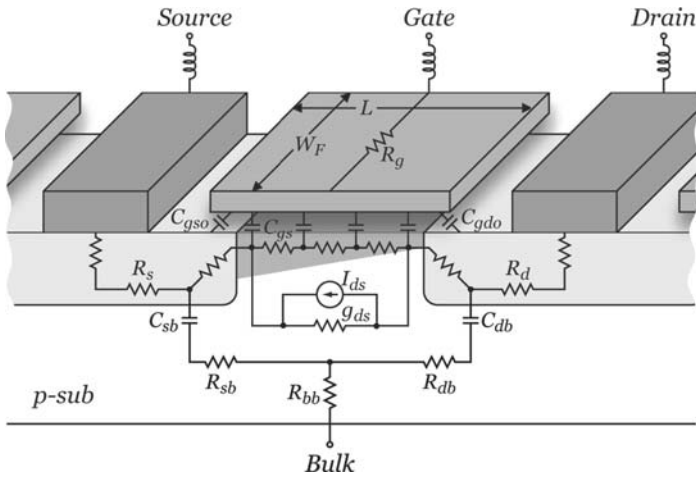


Fig. 3.21 Simplified physical model for one finger of an NMOS device [3] (© IEEE 2005).

where R_{poly} is the polysilicon gate sheet resistance, W_F is the finger width, L is the channel length, and $n = 1, 2$ is determined by the number of gate contacts.

Although using two gate contacts reduces the gate resistance by a factor of 4, one can alternatively use fingers with width $W_F/2$ and double the number of fingers to achieve the same effect. Using two gate contacts requires more complicated routing of the gate line, and often comes at the expense of increasing other parasitics, such as the gate inductance, the source resistance and inductance, and substrate resistance. Therefore, for an in-line transistor layout the narrowest possible gate fingers should be used.

To gain insight into the effect of layout on f_{max} , consider the first-order approximation [11] [12] (neglecting R_D and substrate losses),

$$f_{max} \approx \frac{f_t}{2\sqrt{R_g(g_m C_{gd}/C_{gg}) + (R_g + r_{ch} + R_S)g_{ds}}} \quad (3.39)$$

where $C_{gg} = C_{gs} + C_{gd}$ is the total gate capacitance. It is well-known that in order to maximize f_{max} , the fingers should be narrow. For a fixed I_{DS}/W (i.e., constant V_{GS}), reducing the finger width results in R_G being decreased, while all other resistances (per finger) are increased. For very narrow finger widths, the second term in the denominator of Eq. 3.39 will dominate, R_G will become negligible, and f_{max} will approach a value that is independent of W_F . The polysilicon gate sheet resistance only affects how narrow the fingers must be made. Although the dependence of f_{max}

on the complete model with all parasitics is complicated and cannot be expressed in closed-form, Eq. 3.39 possesses the correct qualitative dependence of f_{max} on W_F . Furthermore, minimizing W_F to optimize f_{max} allows the substrate contacts to be placed more closely to the device, thus minimizing the losses due to the network. Therefore, with optimal layout, f_{max} is not limited by the gate resistance, but is primarily determined by the series source/drain resistances and substrate losses.

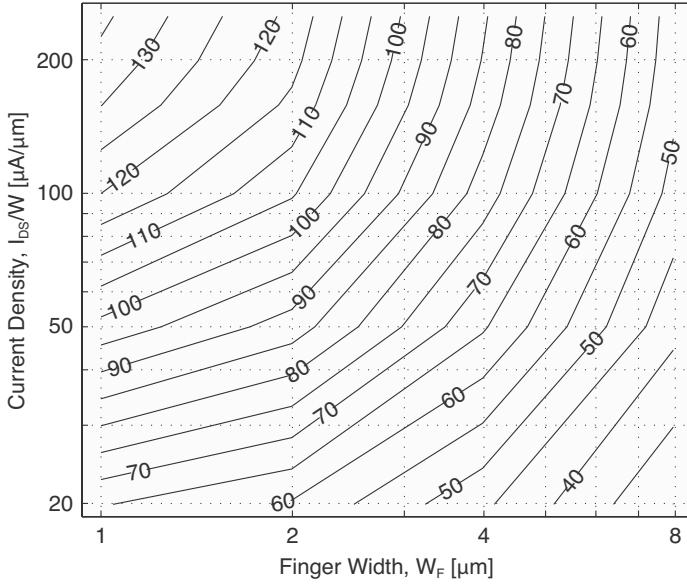


Fig. 3.22 Measured f_{max} [GHz] for minimum channel length ($L = 0.13 \mu\text{m}$) NMOS transistors. The constant f_{max} contour lines are linearly interpolated between the measurement data. The peak measured f_{max} for a $100 \times 1 \mu\text{m}/0.13 \mu\text{m}$ device biased at $I_{DS}/W = 250 \mu\text{A}/\mu\text{m}$ is 135 GHz [3] (© IEEE 2005).

The optimal transistor finger width for a commercial 130-nm digital CMOS process has been determined empirically. The measured f_{max} for NMOS transistors with minimum channel length as a function of finger width and bias current density is displayed in Fig. 3.22. Nine devices with $W_F = 1\text{--}8 \mu\text{m}$ and $N_F = 40\text{--}100$ in common-source configuration with the bulk and source grounded and the gate contacted on one side were fabricated. For six bias points ($20\text{--}300 \mu\text{A}/\mu\text{m}$) per device, a transistor model was extracted from the measured data in order to find f_{max} . The constant f_{max} contours shown in Fig. 3.22 were linearly interpolated between the measured data points.

For a constant current density, the device f_T remains fixed (e.g., $100 \mu\text{A}/\mu\text{m}$, f_T is 70 GHz). It is clear from Fig. 3.22 that, depending on the finger width, f_{max} can be much larger or smaller than f_T . Thus, the optimal layout for mm-wave applications requires CMOS transistors to be designed using many extremely narrow fingers in

parallel (less than $1\text{ }\mu\text{m}$ each). This is in stark contrast to GaAs FETs with metal gates, where relatively few fingers of wide devices ($30 - 75\text{ }\mu\text{m}$) are typically used [25]. Furthermore, the device must be biased well into strong inversion (around $100 - 300\text{ }\mu\text{A}/\mu\text{m}$) for mm-wave operation. By proper layout and biasing, though, the f_{max} of an NMOS transistor in a standard 130-nm CMOS technology can easily surpass 100 GHz, opening the possibility for mm-wave circuits.

Transistor Wiring

To determine the effect of the wiring layout on device parameters, a physical small signal model is used in order to ascertain the effect of each parasitic on the desired performance metrics of the device. The small-signal model of the device is not necessarily unique and different combination of lumped element values could satisfy the required matching between the measurement and the simulation result. As a result, in order to make the model physical, the values of these parasitics were partly determined through 3D EM simulation (Ansoft HFSS) and were set as the initial value for optimization³.

The developed physical small-signal model helps determine the effect of each parasitic element on the performance of the transistor. Ultimately this insight can be used to determine the optimal transistor wiring and layout. For example, the layout and wiring of a common source device has been optimized for f_{max} in a commercial 90nm technology process. Using the previously discussed approach, the starting point for this procedure was a $N_F = 80$ finger $80\text{ }\mu\text{m}/90\text{nm}$ sized device with an MSG of 7.5 dB at 60 GHz and the extrapolated f_{max} of 143 GHz. Compared to 130 nm technology, this device consumes less current for the same gain, but it only has a slightly higher gain at 60 GHz and only a marginally larger f_{max} . A sensitivity analysis was performed using a small-signal model and the variation of maximum unilateral gain U and maximum stable gain together with maximum frequency of operation, f_{max} , were determined.

As expected, the gate to drain capacitance and the gate resistance have the largest impact on f_{max} , and thus layout wiring should minimize them as much as possible. It is worth noting that MSG does not change with a reduction in gate and drain series resistances when the transistor is conditionally stable ($K < 1$). This makes sense since to render the device stable in this region, we have to add resistances to the input and output ports. The source resistance, however, changes the MSG since it changes the effective transconductance through its local series feedback effect.

The NMOS structure was modified based on these findings. Mainly the shape of gate and drain tapers, number of gate vias, and width of connections and gate/drain overlap regions were changed. Fig. 3.23 shows a layout comparison of the structure before and after modification. The measured performance of the initial device and that of the modified device is shown in Fig. 3.24. The f_{max} for the improved structure is up to 178 GHz as was predicted by the analysis. The maximum stable gain of

³ Portions of this text are taken from [27], (© IEEE 2007)



Fig. 3.23 Initial and improved layout for an $80\text{ }\mu\text{m}/90\text{ nm}$ NMOS device. The improved layout also includes more substrate contacts and a higher density of gate vias.

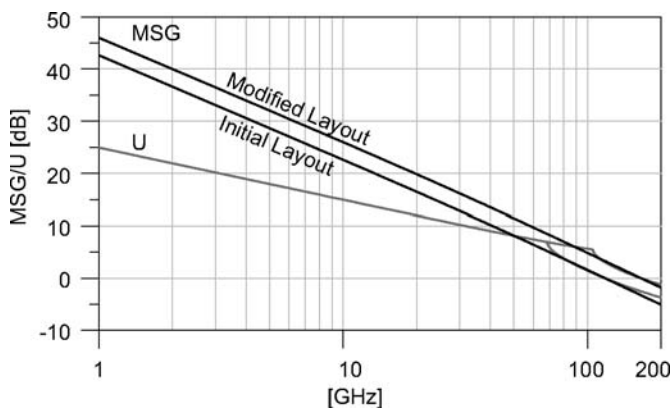


Fig. 3.24 The effect of layout improvement over Mason's Unilateral Gain U and MSG [27] (© IEEE 2007).

the device is intact however since the device is in the conditionally stable region and changing series gate and drain resistances does not change the maximum stable gain.

Round-Table Structure

Working with a simple in-line transistor layout, our previous research showed that the optimal multi-finger layout and wiring of an NMOS device could increase the f_{max} up to 20%, but increasing the performance required further innovation. This is particularly true of the available gain in conditionally stable frequencies. In order to improve the performance of the device even further, a new structure for the device

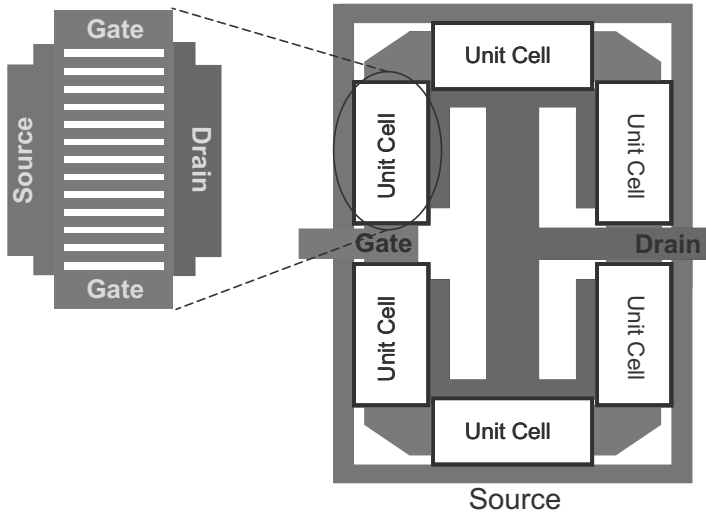


Fig. 3.25 Layout of the "Round Table" device [27] (© IEEE 2007).

was proposed [26]. The idea is to reduce the parasitic losses by using a modular approach in device design.

The building block is a standard $W = 10 \mu\text{m}$ device with $N_F = 10$ fingers, using double-gate contacts in order to decrease the finger resistance of the device as much as possible. Since each finger of the device forms a distributed RC network, double contact reduces the resistance of each finger by a factor of four [29]. These cells are then connected in a matrix or circular fashion depending on the desired size of the final transistor. Fig. 3.25 shows a $W = 60 \mu\text{m}$ NMOS using a circular connection, hence the name "Round Table". This structure uses external double-contacts (between cells) and multi-path connections between sources and drain of the sub cells.

Several dimensions of these devices were fabricated in 90nm CMOS process. Measurements were carried out up to 65 GHz and probing pads were de-embedded from the devices. Fig. 3.26 shows MSG, Mason's gain (U) and H_{21} of a $W = 40 \mu\text{m}$ round-table NMOS. The f_{max} is calculated by extrapolation of the Mason's gain U for frequencies between 20 GHz to 50 GHz, a frequency range where the most reliable data occurs. As evident, measurements suggest significant improvement in both the speed and the desired gain of these devices as compared to in-line layout transistors with the same number of fingers. Even though f_T remains almost constant (100 GHz), f_{max} improved by almost two fold, or to about 300 GHz. This is of course the extrapolated f_{max} since the device introduces new high frequency poles after 100 GHz, rendering the linear approximation of the U curve inaccurate beyond 100 GHz. The model extrapolated f_{max} is 200 GHz. Unlike the improvement of the in-line multi-finger device presented in the previous section, the MSG of the round-table device increases even at frequencies in which the device is conditionally

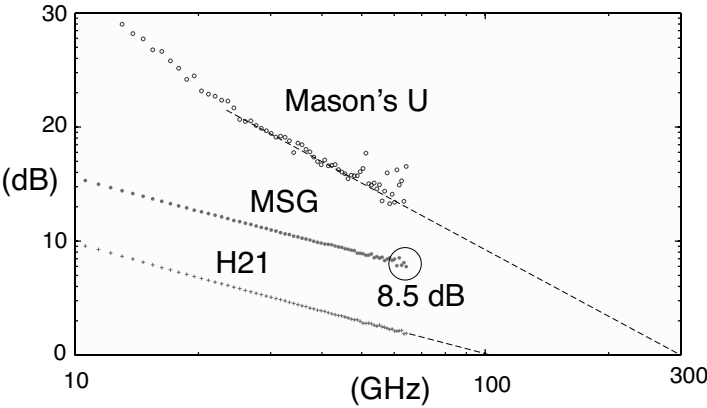


Fig. 3.26 Measured h_{21} , Mason’s Unilateral Gain (U) and Maximum Stable Gain (MSG) for a $40\,\mu\text{m}/90\,\text{nm}$ Round Table device [26] (© IEEE 2007).

stable. The MSG at 60 GHz is 8.5 dB from the value of 7.5 dB (regular NMOS) for $I_D = 28\,\mu\text{A}/\mu\text{m}$. The ratio, f_{max}/f_t , a measure of the optimality of the physical structure of the device, is greater than 2 and as large as 3 (when using the 20 dB/dec extrapolation), the highest reported for CMOS. The improvement of the maximum stable gain is the result of decreased source resistance and parasitic drain to gate capacitance that both act as internal series and shunt feedback. The improvement of f_{max} was mostly due to reduction in the gate and drain resistances.

Table 3.1 A comparison of the small-signal equivalent circuit parameters for an in-line transistor and a Round Table transistor [27] (© IEEE 2007).

	In-Line Device	Round Table Device
$R_g\,(\Omega)$	4.46	2.23
$R_d\,(\Omega)$	3.54	2.42
$R_s\,(\text{m}\Omega)$	672	438
$C_{gs}\,(\text{fF})$	35.7	57.1
$C_{gd}\,(\text{fF})$	21.3	17.2

Table. 3.1 compares the result of extracted small-signal parameters of a round-table $W = 40\,\mu\text{m}$ device to a regular optimized multi-finger $40\,\mu\text{m}$ transistor. All the resistive losses have been reduced considerably as shown in the accompanying table. The parasitic gate-source capacitance of the device is increased. This is mainly due to the increased overlap capacitance between source and gate in order to reduce the gate and source resistances. This is a good trade-off since the C_{gs} does not directly affect the f_{max} and MSG as much as the device loss.

3.2.3 Small-Signal Model

At mm-wave frequencies, series resistive and inductive parasitics become more significant. Consequently, it is critical to properly model these parasitics, in addition to the capacitive effects that are traditionally captured by digital CMOS models [23].

Considering the small margins for modeling errors, the following modeling methodology for active devices is suggested to yield a model with the highest possible accuracy:

- Since the precise layout details—connections to the gate, drain, source, and bulk, location of the substrate contacts, number of fingers, etc.—have a major impact on the parasitic elements, models should be extracted only for fixed layouts.
- The transistors in the circuit have the identical layout as the devices used for the model extraction.
- For the highest accuracy, a bias-dependent small-signal model is extracted. For increased flexibility, a large-signal transistor model based on BSIM3 has also been demonstrated to provide good results up to 65 GHz [1].

The physical model depicting the significant high-frequency parasitics was shown in Fig. 3.21, and Fig. 3.27 shows the corresponding extended circuit model. The core device is modeled using either a lumped small-signal model (Fig. 3.27) or using a standard BSIM3 model card. In addition to the parasitic resistors, series inductors must be added to all terminals— L_G , L_D , L_S —to properly model the delay effects associated with interconnect wiring. Notice that all of the capacitors (e.g., C_{gd} , C_{gs} , C_{ds}) account for both the traditionally “intrinsic” channel and overlap capacitances as well as the traditionally “extrinsic” wiring capacitances.

For each model, the extrinsic component values and device parameters were extracted from measured data using a hybrid optimization algorithm in Agilent IC-CAP [30]. S -parameters for the simulated small-signal model and measured data up to 65 GHz are shown in Fig. 3.28 for a $100 \times 1 \mu\text{m}/0.13 \mu\text{m}$ NMOS transistor biased at $V_{GS} = 0.65 \text{ V}$ and $V_{DS} = 1.2 \text{ V}$. The excellent broadband accuracy of the simulation compared to the measured data verifies that the topology of our model is correct and complete. Furthermore, it also demonstrates that distributed effects and frequency-dependent losses caused by the skin effect can be adequately accounted for using only lumped extrinsic components with frequency-independent values.

The transistor gains—Mason’s unilateral gain, maximum stable gain (MSG), maximum available gain (MAG), and current gain—for this device are plotted in Fig. 3.20.

The accurate modeling of the unilateral gain is particularly important. Unlike the MSG and current gain, Mason’s unilateral gain is a very strong function of all resistive losses. Therefore, accurately fitting the unilateral gain validates that the important loss mechanisms have been properly modeled. As mentioned earlier, these resistive losses are critical because they ultimately limit the high-frequency capabilities of the transistor.

For a common-source 90nm CMOS device, a small-signal model as shown in Fig. 3.29 is used. This lumped equivalent circuit also adds parasitic resistances and inductances as well as a substrate network to the well-known hybrid- π model of the

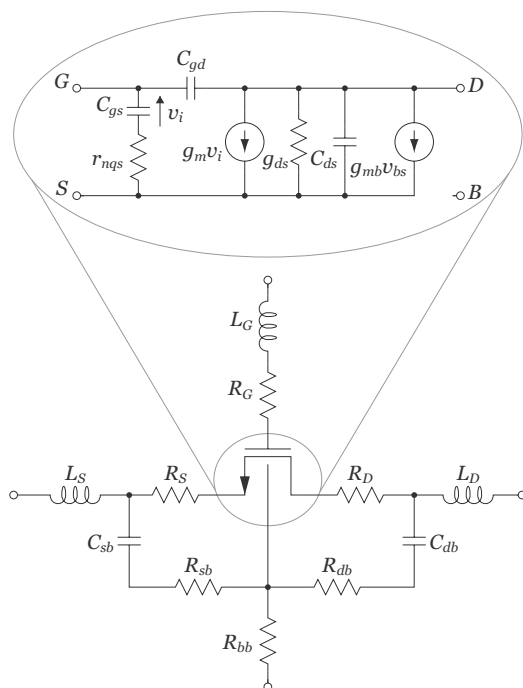


Fig. 3.27 Small-signal transistor model for an NMOS device showing the important parasitic elements [3] (© IEEE 2005).

core transistor. Figure 3.30 shows the result of this modeling technique by comparing S -parameters measurement of a sample common-source transistor to its extracted model. The model shows a good match with the measurement up to 100 GHz as evident in this figure.

3.2.4 Large-Signal Model

Although ac models are often sufficient for the design of linear circuit blocks, the optimization of nonlinear circuits such as mixers, power amplifiers, oscillators, and frequency multipliers, requires precise knowledge of the nonlinear characteristics of the active devices over a wide operation range⁴.

Large-signal device models have evolved into two basic categories: table-based [32] and physical [33][34]. At mm-wave frequencies, GaAs FET models commonly employ table-based models derived from bias-dependent linear measurements. The large-signal accuracy of table-based models is limited by the existence of discontinu-

⁴ Portions of this text are taken from [1], (© IEEE 2004)

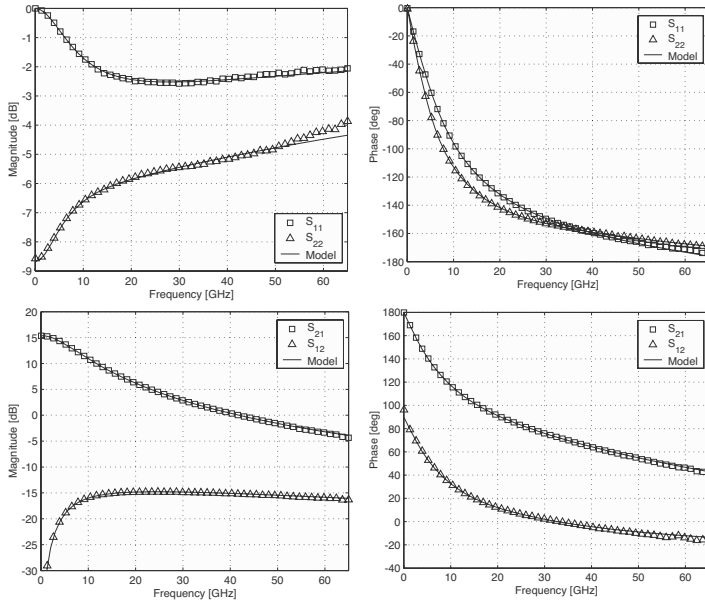


Fig. 3.28 Measured (markers) and simulated (solid lines) S -parameters for a typical NMOS device ($W/L = 100 \times 1 \mu\text{m}/0.13 \mu\text{m}$, $I_{DS}/W = 300 \mu\text{A}/\mu\text{m}$, $V_{DS} = 1.2 \text{ V}$) [3] (© IEEE 2005).

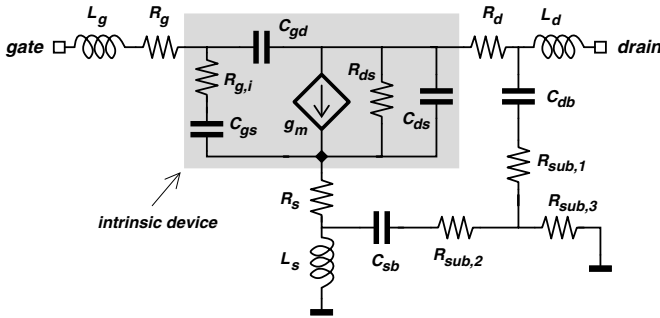


Fig. 3.29 Small-signal equivalent circuit of a common source CMOS device modeled in 90 nm technology [31] (© IEEE 2007).

ities in the model elements and nonlinearities in their interpolation due to imperfect measurement data [35]. Recent efforts to model the large-signal distortion performance of CMOS transistors with compact models have provided accurate results, but have been verified only at frequencies below 2 GHz [34][36].

In this section, we introduce an extension to the standard BSIM3 modeling procedure, and provide experimental validation of small-signal and large-signal accuracy up to 65 GHz.

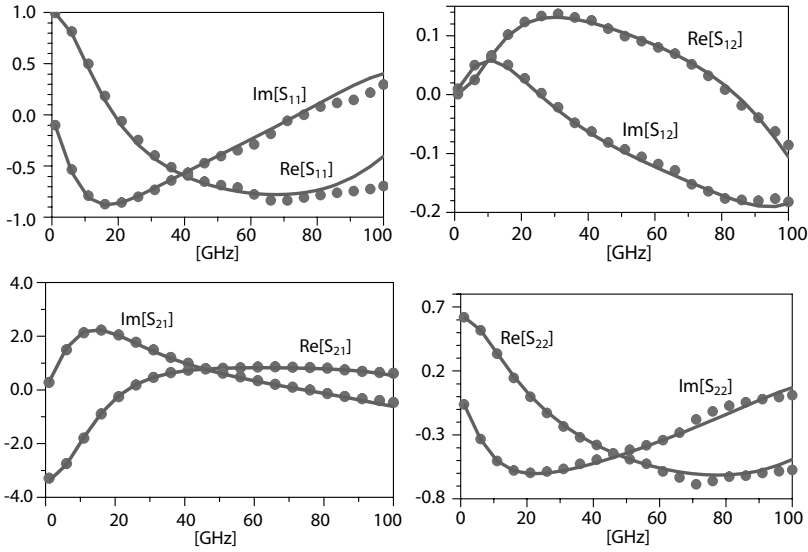


Fig. 3.30 A comparison of the measurement (marker) and modeled (line) transistor S -parameters up to 100 GHz for an NMOS device in 90 nm technology [27] (© IEEE 2007).

Modeling Methodology

The proposed modeling methodology is based on the quasi-static assumption that the mm-wave large-signal performance of a transistor is primarily governed by its dc nonlinearities, while the dynamic performance can be modeled with the addition of extrinsic parasitics to capture loss and inductive effects, which become particularly important at mm-wave frequencies.

The proposed modeling methodology uses a core BSIM model for the intrinsic transistor, augmented with extrinsic parasitics as depicted in Fig. 3.31. The series source and drain resistances are dominated by the intrinsic spreading resistance in the source-drain extension (SDE) region near the channel. The gate resistance, R_G , accounts for the distributed RC nature of the polysilicon gate, and can be approximated using a single lumped resistor⁵. Additionally, series inductors must be added to all terminals, L_G , L_D , L_S , to properly model the delay effects associated with interconnect wiring. Layout-dependent interconnect capacitances are also added around the intrinsic device, and junction diodes account for the voltage-dependence of C_{db} and C_{sb} .

⁵ The bias dependence due to the channel inversion level is ignored.

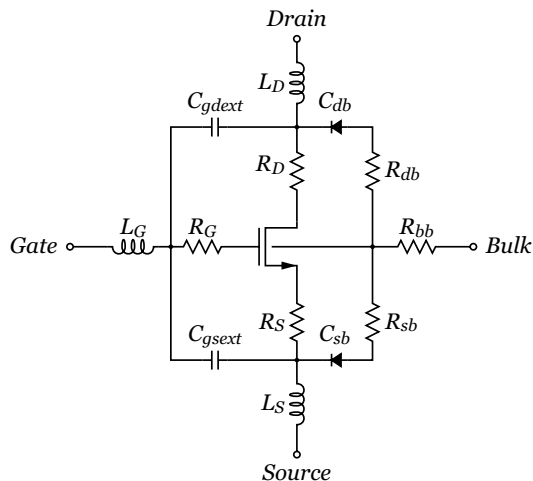


Fig. 3.31 BSIM3 core model with extrinsic parasitics [1] (© IEEE 2004).

Model Extraction

A $80 \times 1 \mu\text{m}/0.13 \mu\text{m}$ NMOS transistor, optimized for operation at mm-wave frequencies and fabricated on a standard $0.13\text{-}\mu\text{m}$ CMOS process, is used as a demonstration vehicle in the large signal modeling. De-embedding structures were used to remove the effects of the pads from the small-signal device characterization.

For dc large-signal characterization, I - V measurements were used to extract the core BSIM parameters of the fabricated common-source NMOS. As shown in Figs. 3.32a and Fig. 3.32b, a good agreement between measured and modeled dc curves can be achieved for this device. Next, over a wide bias range, extensive on-wafer S -parameter measurements to 65 GHz were performed on a Cascade Microtech probe station using an Anritsu 37397C VNA and Cascade Microtech GSG coplanar probes. The external parasitic component values for the model were extracted using a hybrid optimization algorithm in Agilent IC-CAP. S -parameters for the simulated model and measured data up to 65 GHz are shown in Fig. 3.33 for a typical bias sweep over V_{GS} for the de-embedded transistor. The broadband accuracy of the model verifies that using lumped parasitics (Fig. 3.31) is suitable well into the mm-wave region.

Large-Signal Measurement Setup

The nonlinear model described in Section 3.2.4 needs experimental verification. Two common approaches to characterize transistor mm-wave nonlinearity are with mm-

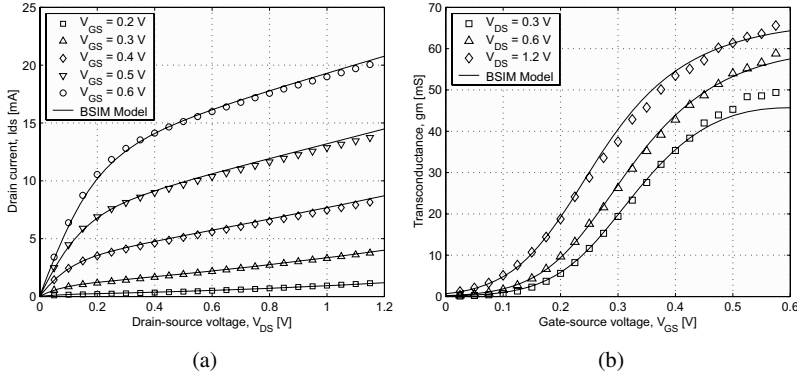


Fig. 3.32 (a) Measured and modeled I_{DS} vs. V_{DS} . (b) Measured and modeled g_m vs. V_{GS} [1] (© IEEE 2004).

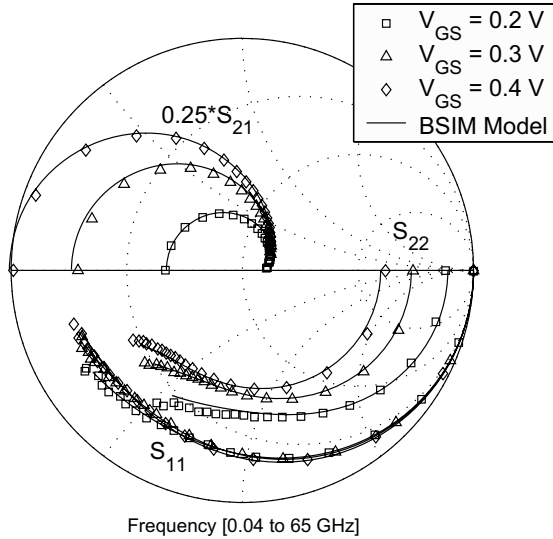


Fig. 3.33 Measured and modeled S-parameters for $V_{DS} = 1.2$ V, and $V_{GS} = 0.2, 0.3, 0.4$ V [1] (© IEEE 2004).

wave load-pull measurements and power spectrum analysis [37]. While the former requires automated tuners, the latter, which was chosen in this work, can be performed with only a synthesizer, VNA, and power meter (Fig. 3.34). The fabricated device is driven over a wide range of input power and bias conditions, while the output powers at the fundamental and harmonic frequencies are measured. In the harmonics power measurement setup of Fig. 3.34, the VNA is configured as a receiver in the

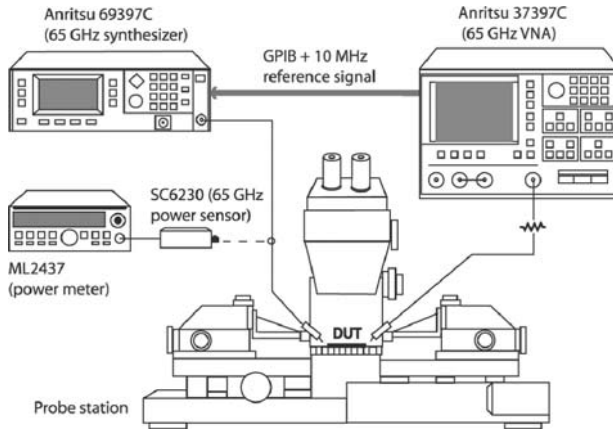


Fig. 3.34 Test setup for mm-wave harmonics measurements [1] (© IEEE 2004).

Set-On mode. In this mode of operation, the source lock circuitry of the 37397C is completely by-passed which allows all of the 37397C samplers to operate over their full dynamic range. The 65 GHz synthesizer and 37397C are locked to the same 10 MHz reference, enabling coherent reception at the harmonic frequencies. A 65 GHz high dynamic-range power sensor is used to de-embed the insertion loss of the cables, probes, adaptors, etc. from the measurements. Port2 of the VNA requires a 10-dB attenuator to avoid compression when the fundamental power is strong, thus limiting the sensitivity of the power measurement at 60 GHz to approximately -35 dBm.

Measured And Simulated Results

Fig. 3.35a shows the measured and modeled fundamental, second, and third harmonics at the output when the device is driven by a 20 GHz 0-dBm signal with a bias of $V_{DS} = 1V$ and variable V_{GS} . The output powers at the fundamental and second harmonics agree very well, whereas the agreement at the third harmonic is reasonable, but not as good. The measurement of the third harmonic for $V_{GS} > 0.4V$ is limited due to the dynamic range of the VNA samplers at 60 GHz.

Similar experiments were performed at a different input power (-2 dBm) and signal frequency (30 GHz). Fig. 3.35b shows the measured and modeled fundamental and second harmonics of the output signal for the same bias conditions. Good agreement is observed at both the fundamental and second harmonic. Since the second harmonic at 60 GHz is strong, it is within the dynamic range of the measurement setup.

Fig. 3.36a displays the measured and modeled 60 GHz large-signal amplification characteristics of the device. The model closely predicts the fundamental output

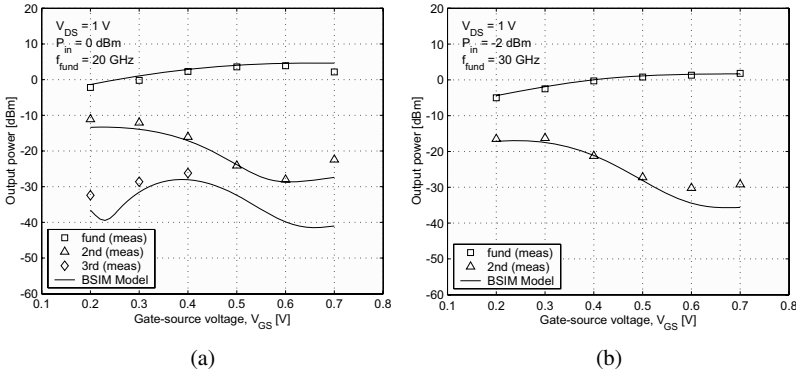


Fig. 3.35 (a) Fundamental, second, and third harmonics vs. the gate bias (V_{GS}) for $V_{DS} = 1$ V, $P_{in} = 0$ dBm, $f_{fund} = 20$ GHz. (b) Fundamental and second harmonic vs. the gate bias (V_{GS}) for $V_{DS} = 1$ V, $P_{in} = -2$ dBm, $f_{fund} = 30$ GHz [1] (© IEEE 2004).

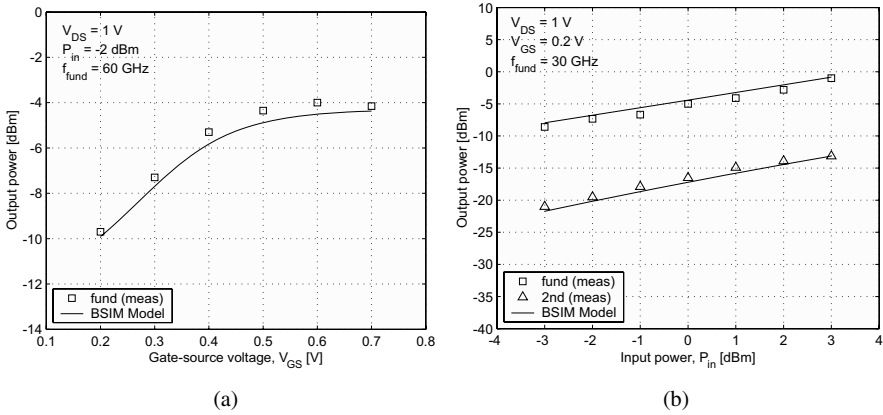


Fig. 3.36 (a) Large-signal gain vs. gate bias (V_{GS}) for $V_{DS} = 1$ V, $P_{in} = -2$ dBm, $f_{fund} = 60$ GHz. (b) Class AB power sweep curves [1] (© IEEE 2004).

power for the device over a large bias range from weak inversion to very strong inversion.

Many mm-wave circuits, such as mixers and frequency multipliers, operate by exploiting the strong nonlinearity of transistors biased near the threshold voltage, effectively using the device as a rectifier. To determine the accuracy of the model for these applications, the transistor was biased at constant $V_{GS} = 0.2$ V, and the input power was swept from $P_{in} = -3$ dBm to $+3$ dBm at a fundamental of 30 GHz. The results shown in Fig. 3.36b show that the extended BSIM model provides good accuracy for class AB operation over a wide range of input signal power.

The results validate that with accurate dc nonlinearity fitting, a simple extended compact model can predict the harmonic distortion behavior of a device up to 60 GHz. The ability to model the strongly nonlinear case of class AB operation is also verified, which suggests that the model is viable for use in the design of nonlinear mm-wave blocks such as mixers, power amplifiers, and frequency multipliers.

3.2.5 FET Noise Model

An accurate noise model is crucial in designing low noise amplifiers. Typically foundry compact noise models have been optimized to fit measurements up to 20 GHz, and typically the mismatch between measurement and the model is significant at higher frequencies due to inadequacies of compact models. This inadequacy arises from the distributed nature of the device near the f_T and uncertainty to the cause of the excess noise in short channel devices. Traditional microwave design, on the other hand, is based directly on two-port noise parameters. Although this method is the most accurate, its applicability is limited due to the difficulty in noise parameter measurement and the lack of robust de-embedding for Si devices.

The noise figure of a two-port system based on the four noise parameters is given by [2]

$$F = F_{min} + \frac{R_n}{G_s} |Y_s - Y_{opt}|^2 \quad (3.40)$$

in which F_{min} is the minimum achievable noise figure, R_n is the noise sensitivity resistance, and Y_{opt} is the optimal source noise admittance, and $Y_s = G_s + jB_s$ is the source admittance.

As technology has advanced, CMOS technology minimum noise figure F_{min} has dropped significantly, approximately like (f/f_t)

$$F_{min} = 1 + 2 \left(\frac{f}{f_t} \right) \sqrt{g_m R_g \frac{\gamma}{\alpha}}$$

The above is calculated based on Pospiezalski's noise model [38] (see next section), which has resulted in a good match with CMOS devices in mm-wave frequencies. In the limit that we use short transistor fingers to minimize the gate resistance R_g , the noise is bounded by the NQS resistance seen at the gate $R_g = 1/5g_m$ [24], which makes the fundamental noise just dependent on the technology parameters

$$F_{min} > 1 + 2 \left(\frac{f}{f_t} \right) \sqrt{\frac{1}{5} \frac{\gamma}{\alpha}}$$

Based on this simplistic prediction, a plot of F_{min} is shown in Fig. 3.37, which shows that F_{min} is as low as 2.5 dB at 60 GHz in the 90nm node. Actual measurements show that this lower bound is reached to within 1 dB in practice, which is very encouraging. For example, measurement results at mm-wave frequencies (60 GHz) on the 130 nm transistor also reveal that the minimum achievable noise figure is between 3-4 dB. This level of performance is sufficient for many applications. The remaining challenge is to build an amplifier that can actually match for low noise.

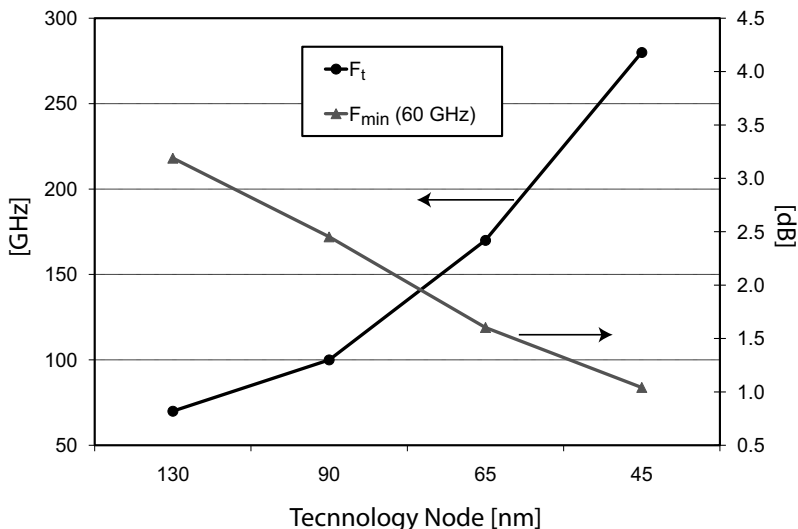


Fig. 3.37 Lower bound for minimum achievable noise figure F_{min} as a function of technology node.

3.2.5.1 Review of Noise Models

The classical Van Der Ziel RF noise model [39][40] is well known. As shown in Fig. 3.38, the high frequency noise of the device is described by two correlated

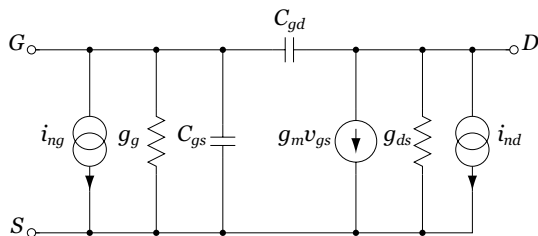


Fig. 3.38 The van der Ziel RF noise model.

current noise sources, namely the induced gate noise (i_{ng}) and the short-channel drain current noise (i_{nd}). The current noise sources are given by

$$\frac{\overline{i_{nd}^2}}{\Delta f} = 4kT \left(\frac{\gamma}{\alpha} \right) g_m \quad (3.41)$$

$$\frac{\overline{i_{ng}^2}}{\Delta f} = 4kT \delta g_g \quad (3.42)$$

$$g_g = \frac{\alpha \omega^2 C_{gs}^2}{5g_m} \quad (3.43)$$

$$c \equiv \frac{\overline{i_{ng} \cdot i_{nd}^*}}{\sqrt{\overline{i_{ng}^2} \overline{i_{nd}^2}}} \approx j0.395 \quad (3.44)$$

where c is the correlation factor between the two current noise sources. The induced gate noise increases with the square of frequency, playing an important role at microwave and mm-wave frequency. The short-channel drain current noise is proportional to γ , which increases with shorter channel length processes [41]. Flicker noise, usually considered a low frequency noise source, can play an important role at mm-wave frequencies due to the noise translation in non-linear and time-varying circuits.

Unfortunately correlated noise sources are not easily incorporated into standard based compact models. Even though it is mathematically trivial to add correlated noise sources to the simulators, due to the historical absence of such elements, other approaches have been taken to model the noise. A procedure to implement correlated noise into Verilog-A is reported in [42]. In the holistic thermal noise model of BSIM [16], shown in Fig. 3.39, the noise is partitioned into a gate noise and a drain noise. The induced gate noise is captured due to the source noise voltage, but unfortunately the phase of the correlation is not imaginary as predicted by the Van Der Ziel model. The Philips MOS11 model, shown in Fig. 3.39b, captures the gate noise automatically since the transistor is broken into a pseudo-distributed circuit [43]. Measurements show that in practice five sections is sufficient to accurately capture the induced noise.

On the opposite end of the spectrum, one can characterize noise based on measurement data, as shown in Fig. 3.40. In this approach, the transistor is modeled as a two-port noiseless black box with two external correlated input noise generators, namely the equivalent input noise voltage and the equivalent input noise current.

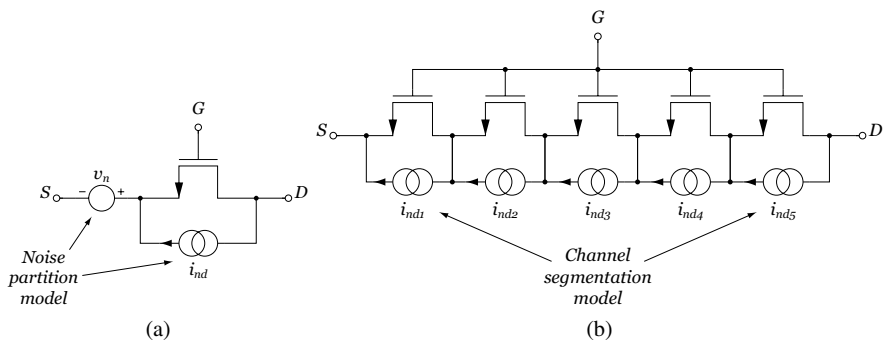


Fig. 3.39 (a) BSIM4 noise model. (b) Philips MOS 11 noise model.

The drawback is that this model is valid only for a fixed bias point and at a single frequency.

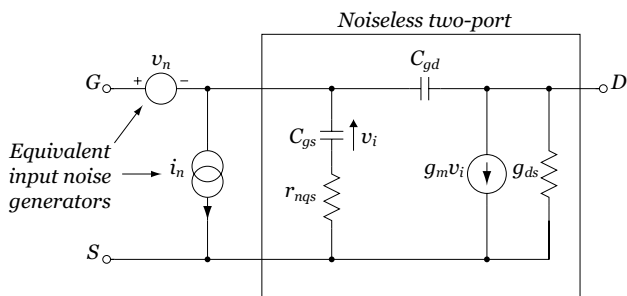


Fig. 3.40 Measurement-based noise model.

Equivalent-circuit noise models are essentially based on the Van Der Ziel model. The PRC model shown in Fig. 3.41a, models the induced gate noise and the drain current noise through the following equations

$$\frac{\overline{i_{nd}^2}}{\Delta f} = 4kTg_m \cdot P \quad (3.45)$$

$$\frac{\overline{i_{ng}^2}}{\Delta f} = 4kT \frac{\omega^2 C_{gs}^2}{g_m} \cdot R \quad (3.46)$$

$$\frac{\overline{i_{ng} \cdot i_{nd}^*}}{\sqrt{\overline{i_{ng}^2} \overline{i_{nd}^2}}} \approx jC \quad (3.47)$$

and the parameters P , R , and C are extracted from measurement data. On the other hand, the Pospieszalski model [38], shown in Fig. 3.41b, assumes the gate noise voltage and the drain noise current are uncorrelated, and it models the induced gate noise source as thermal noise contributed by the NQS resistance. It thus requires the extraction of only one parameter. Due to the simplicity of this noise model, and its ability to predict the device noise (shown in the next section), we have adopted this model in our mm-wave research.

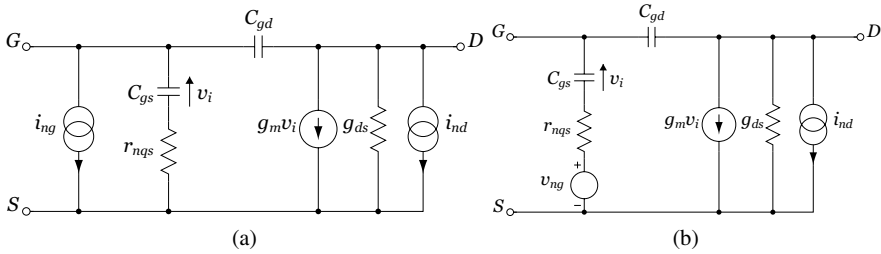


Fig. 3.41 (a) PRC noise model. (b) Pospieszalski noise model.

The basic small-signal model can include noise if one can correctly partition the gate resistance into an intrinsic and extrinsic portion. Since the extrinsic resistance is bias independent, one can use several bias points to derive the intrinsic gate resistance. Extraction of the short-channel drain noise current yields a value for γ of 1.4. The Pospieszalski model was chosen because it yields reasonably good results while being simple and requiring the extraction of only one variable, which is beneficial because high frequency measurements are usually quite difficult.

In essence the gate r_{nqs} and channel resistance (R_{ds}) are noisy. The former is kept at the gate temperature which is close to the room temperature while the equivalent temperature for the latter could reach several thousands degrees. Based on the value of noise parameter, $\gamma = 1.4$, the drain temperature is 3640°C.

Noise Measurement and Simulation Results

Noise measurement in the mm-wave regime requires a semi-custom setup due to the unavailability of commercial solutions. A typical setup is shown in Fig. 3.42, which incorporates a calibrated noise source, a source tuner to vary the source impedance, a low noise amplifier and low noise mixer to down-convert the signal to an IF for noise measurement [44]. Most noise figure meters allow an external down-converter to be used for exactly this purpose. By varying the source impedance and measuring

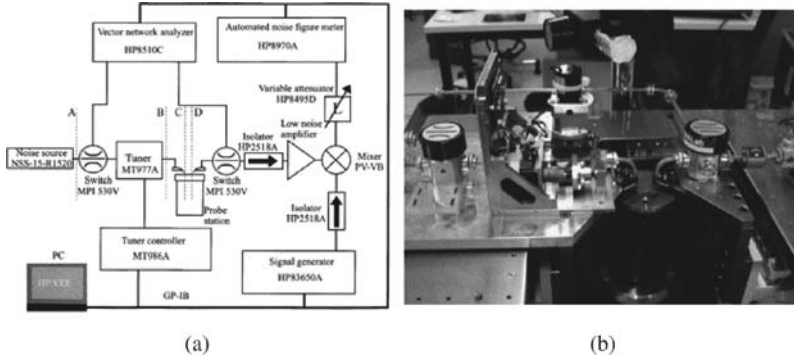


Fig. 3.42 (a) Noise measurement setup at mm-wave frequencies. (b) Photo of noise measurement setup. (Courtesy of VTT)

the noise figure of the device, we can find the noise parameters of the transistor or circuit under measurement.

In Fig. 3.43-3.44 we compare the measured and simulated minimum noise figure (F_{min}), noise resistance (R_n), and optimal impedance (Y_{opt}) using the extended transistor modeling approach for a commercial CMOS 130nm process. All simulations results use $\gamma = 1.4$. A good match is observed in many cases, especially in the trends of the noise parameters versus bias and frequency. It's interesting to note that the noise of a cascade device is substantially higher than a single transistor, a result that differs at low RF frequencies. This is because the cascade device contributes negligible noise at low frequency due to the large degeneration impedance at the cascade node. At high frequency, the capacitance at this node presents less degeneration to the casocde device, increasing the noise. Equivalently, the drain noise of the cascode circulates within the device at low frequencies but flows through the node capacitance at high frequency, which flows through the output.

The minimum noise figure and noise sensitivity of a $W = 40 \mu\text{m}$ round-table device in a 90 nm commercial CMOS is simulated based on the proposed model and shown in Fig. 3.45. The F_{min} increases nearly linearly with frequency. However, the noise sensitivity R_n decreases in the millimeter wave region. This translates into a smaller penalty when deviating from the optimal source impedance. Another key point in designing low noise amplifiers is the trade-off between gain and noise optimization. In this case the optimal gain and noise circles for the same device are close to each other on the Smith Chart. This together with the smaller noise sensitivity means that a respectable noise figure could be achieved even by designing the amplifier only for gain performance.

In Fig. 3.46 we compare the measurement and simulated results for the 60 GHz 3-stage amplifier discussed in [3]. Measurement and simulation results match reasonably well. Most importantly, the trend with frequency and bias is well predicted, allowing the circuit designer to optimize the performance of an amplifier.

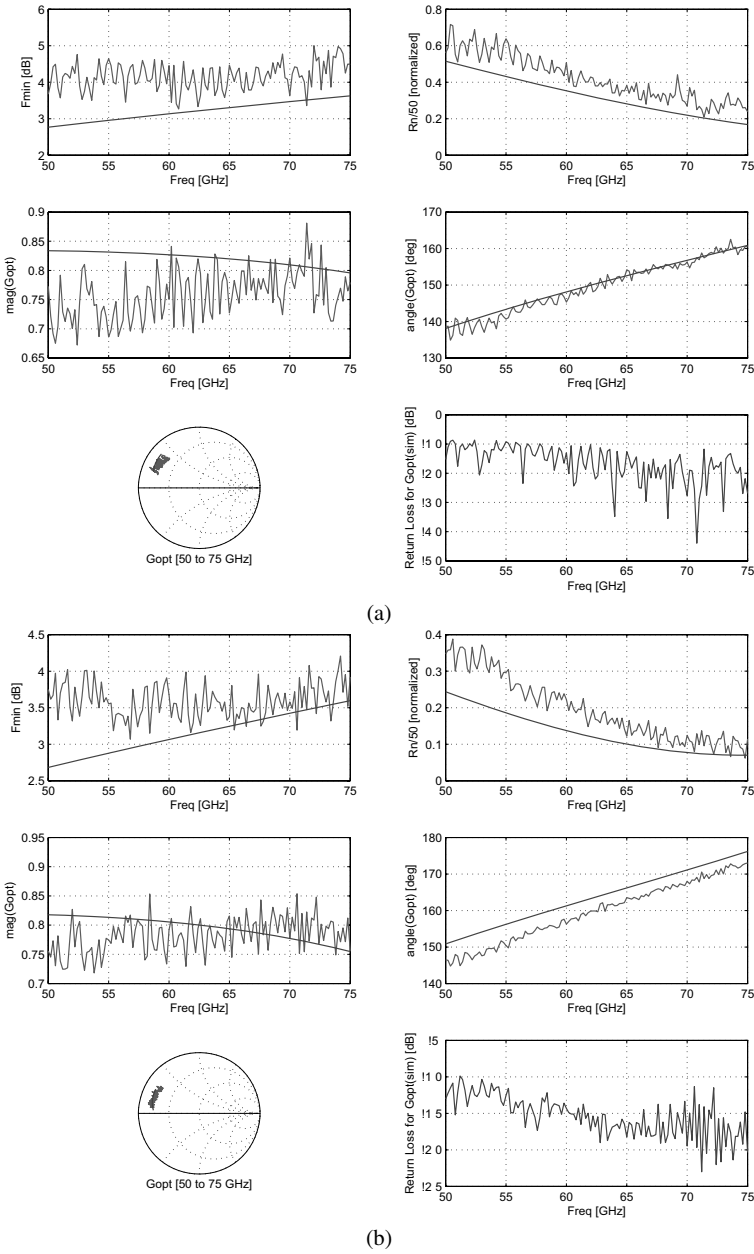


Fig. 3.43 Measurement and simulation results for the minimum noise figure (F_{min}), the noise resistance (R_n), and the optimal impedance (Y_{opt}) using the extended transistor modeling approach for the devices biased at $I_{DS}/W = 150 \mu\text{A}/\mu\text{m}$. NMOS devices sized at: (a) $40 \times 1 \mu$ (b) $60 \times 1 \mu$.

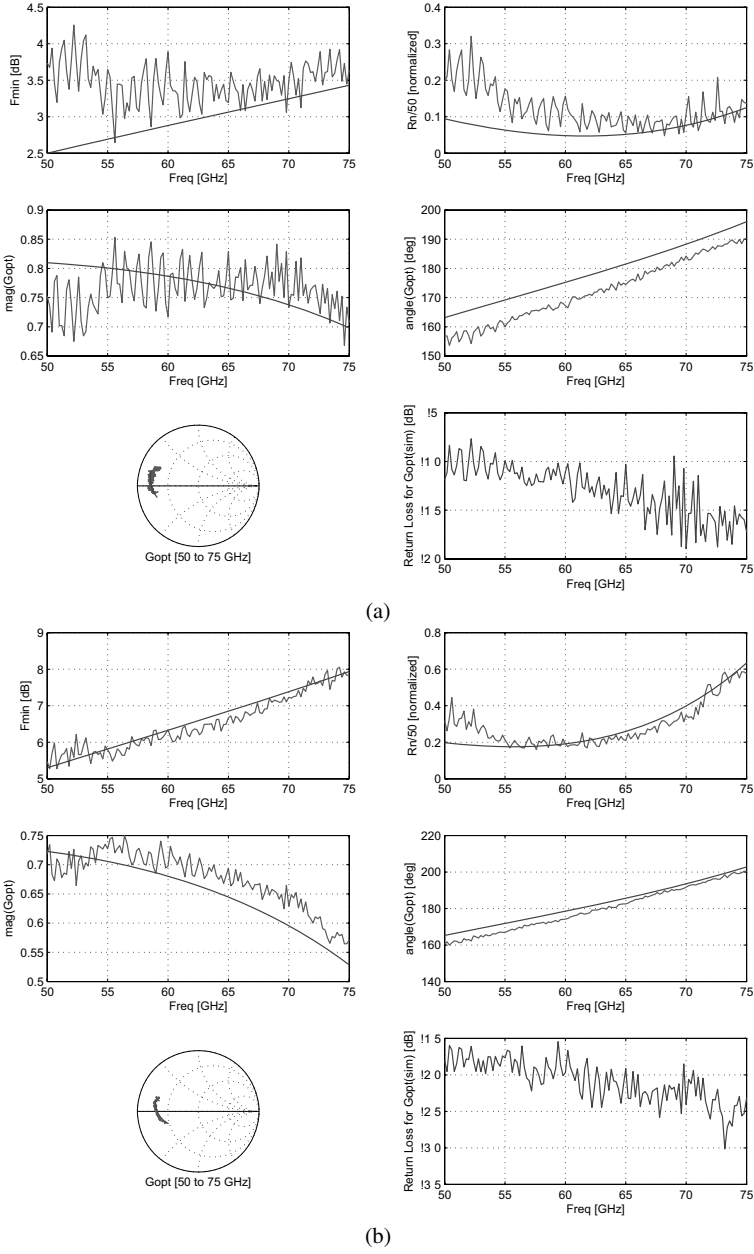


Fig. 3.44 Measurement and simulation results for the minimum noise figure (F_{min}), the noise resistance (R_n), and the optimal impedance (Y_{opt}) using the extended transistor modeling approach for the devices biased at $I_{DS}/W = 150 \mu\text{A}/\mu\text{m}$. NMOS devices sized at: (a) $80 \times 1 \mu\text{m}$ (b) Cascode $40 \times 2 \mu\text{m}$.

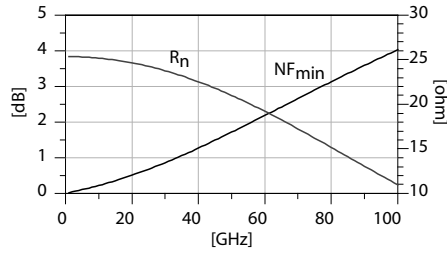


Fig. 3.45 Simulated NF_{min} and noise sensitivity parameter R_n vs. frequency for a Round Table device.

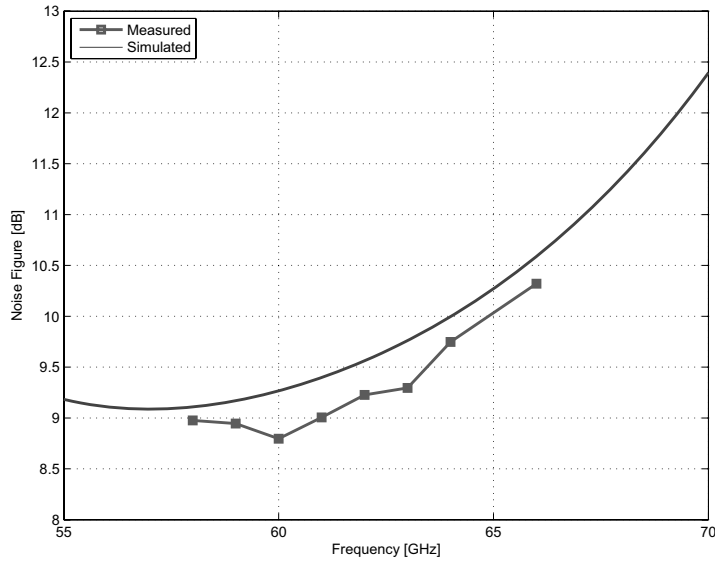


Fig. 3.46 Measurement and simulation results of noise figure for a 60 GHz amplifier [3] (© IEEE 2005).

3.3 Conclusion

In this chapter we have highlighted the design and modeling of passive and active components for mm-wave applications. The design and modeling approach is a result of several years of experiments and characterizations performed on various technology nodes (180nm, 130nm, and 90nm). Custom layout is integral for achieving performance near the limits of activity of the process. In a modern CMOS or

SiGe process, there is a rich variety of devices available to the designer, and yet the designer should not shy away from customizing these structures for mm-wave application requirements. We have demonstrated that custom tailored transmission lines, transistors, and capacitors can be used for higher quality factors, higher gain, and higher self-resonant frequency. The accuracy of the design approach and the models is borne out by measurements and by circuit building blocks described in Ch. 4. By building a library of active and passive components, and by carefully modeling the components into the mm-wave regime, one forms the foundation for a predictable and robust mm-wave design methodology. We have focused our attention on CMOS technology, but certainly with passive devices almost all observations and comments can be applied directly to SiGe or any other technology incorporating a semi-conductive substrate. The design methodology for active devices is highly dependent on the underlying device physics. We have shown, however, that the intrinsic transistor model remains valid well into the mm-wave regime, only requiring careful measurements and modeling of the extrinsic device. This simplifies Si based design tremendously, since active devices are carefully characterized and modeled over process and temperature.

References

1. S. Emami, C. H. Doan, A. M. Niknejad, R. W. Brodersen, "Large-signal millimeter-wave CMOS modeling with BSIM3," *RFIC Digest of Papers*, pp. 163-166, June 2004.
2. D.M. Pozar *Microwave engineering*, Wiley, New York, 1998.
3. C. H. Doan, S. Emami, A. M. Niknejad, and R. W. Brodersen, "Millimeter-wave CMOS design," *IEEE Journal of Solid-State Circuits*, vol. 40, pp. 144-155, Jan. 2005.
4. C. P. Yue and S. S. Wong, "On-chip spiral inductors with patterned ground shields for Si-based RF IC's," *IEEE J. Solid-State Circuits*, vol. 33, pp. 743-752, May 1998.
5. S. Ramo, J. R. Whinnery, and T. Van Duzer, *Fields and Waves in Communication Electronics*, 3rd ed. New York: Wiley, 1994.
6. S. J. Mason, "Power gain in feedback amplifiers," *IRE Trans. Circuit Theory*, vol. CT-1, pp. 20-25, June 1954.
7. C. Doan, Ph.D. Dissertation in Preparation, U.C. Berkeley.
8. C. Enz, "An MOS transistor model for RF IC design valid in all regions of operation," *IEEE Trans. Microwave Theory Tech.*, vol. 50, pp. 342-359, Jan. 2002.
9. B. Razavi, R.-H. Yan, and K. F. Lee, "Impact of distributed gate resistance on the performance of MOS devices," *IEEE Trans. Circuits Syst. I*, vol. 41, pp. 750-754, Nov. 1994.
10. T. C. Edwards and M. B. Steer, *Foundations of Interconnect and Microstrip Design*, 3rd ed. New York: Wiley, 2000.
11. S. P. Voinigescu, S. W. Tarasewicz, T. MacElwee, and J. Iowski, "An assessment of the state-of-the-art 0.5 μ m bulk CMOS technology for RF applications," in *IEDM Tech. Dig.*, Dec. 1995, pp. 721-724.
12. J. N. Burghartz, M. Hargrove, C. Webster, R. Groves, M. Keene, K. Jenkins, D. Edelstein, R. Logan, and E. Nowak, "RF potential of a 0.18- μ m CMOS logic technology," in *IEDM Tech. Dig.*, pp. 853-856, Dec. 1999.
13. B. Kleveland, C. H. Diaz, D. Vook, L. Madden, T. H. Lee, and S. S. Wong, "Exploiting CMOS reverse interconnect scaling in multigigahertz amplifier and oscillator design," *IEEE J. Solid-State Circuits*, vol. 36, pp. 1480-1488, Oct. 2001.
14. G. Carchon, W. De Raedt, and B. Nauwelaers, "Novel approach for a design-oriented measurement-based fully scalable coplanar waveguide transmission line model," *IEE Proc. Microwave, Antennas Propagat.*, vol. 148, pp. 227-232, Aug. 2001.
15. W. F. Andress and D. Ham, "Standing wave oscillators utilizing wave-adaptive tapered transmission lines," *IEEE J. Solid-State Circuits*, vol. 40, no. 3, pp. 638-651, Mar. 2005.
16. The BSIM 4.4 Manual: <http://www-device.eecs.berkeley.edu/bsim3>.
17. MOS11, http://www.semiconductors.philips.com/Philips_Models/mos_models/model11/.
18. EKV model website, <http://legwww.epfl.ch/ekv/>.
19. PSP model website, http://www.nxp.com/Philips_Models/mos_models/psp/.
20. J. C. Guo, W.Y. Lien, M.C. Hung, C.C. Liu, C.W. Chen, C.M. Wu, Y.C. Sun, and Ping Yang, "Low-K/Cu CMOS Logic Based SoC Technology for 10Gb Transceiver with 115GHz f_T, 80GHz f_{MAX} RF CMOS, High-Q MiM Capacitor and Spiral Cu Inductor," *VLSI Digest of Technical Papers*, pp. 39-40, June 2003.
21. L. F. Tiemeijer et al., "Record RF performance of standard 90 nm CMOS technology," *IEDM Technical Digest*, pp. 441-444, Dec. 2004.
22. R. Spence, *Linear Active Networks*, London: Wiley, 1970.
23. C. Enz, "An MOS transistor model for RF IC design valid in all regions of operation," *IEEE Trans. Microwave Theory Tech.*, vol. 50, pp. 342-359, Jan. 2002.
24. Tsividis, Y., *Operation and Modeling of the MOS Transistor*, 2nd ed. Boston : WCB/McGraw-Hill, c1999.
25. I. Bahl and P. Bhartia, *Microwave Solid State Circuit Design*, 2nd ed. Hoboken, NJ: Wiley, 2003.
26. B. Heydari, M.Bohsali, E.Adabi, A. M. Niknejad, "Low-Power mm-Wave Components up to 104GHz in 90nm CMOS," *ISSCC Dig. Tech. Papers*, pp. 200-201, Feb. 2007.

27. B. Heydari, M. Bohsali, E. Adabi, A.M. Niknejad, "mm-Wave devices and circuit blocks up to 104 GHz in 90nm CMOS," to appear in *IEEE J. Solid-State Circuits*.
28. D. Huang, W. Hant, N.-Yi Wang, T.W. Ku, Q. Gu, R. Wong, M.-C.F. Chang, "A 60GHz CMOS VCO using on-chip resonator with embedded artificial dielectric for size, loss and noise reduction," *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, pp. 1218-1227, Feb. 2006.
29. A. M. Niknejad, *Electromagnetics for High-Speed Analog and Digital Communication Circuits*, 1st Edition. Cambridge University Press, 2007.
30. Agilent IC-CAP 2002 User's Guide, <http://eesof.tm.agilent.com>.
31. B. Heydari, P. Reynaert, E. Adabi, M. Bohsali, B. Afshar, M. A. Arbabian and A. M. Niknejad, "A 60-GHz 90-nm CMOS cascode amplifier with interstage matching," to be presnted at *EU Microwave Conference (EuMic)*, 2007.
32. J. Wood and D. E. Root, "Bias-dependent linear scalable millimeter-wave FET model," *IEEE Trans. Microwave Theory Tech.*, vol. 50, pp. 2352-2360, Dec. 2000.
33. BSIM3, <http://www-device.eecs.berkeley.edu/~bsim3>
34. R. Van Langevelde, L. F. Tiemeijer, R. J. Havens, M. J. Knitel, R. F. M. Ores, P. H. Woerlee, and D.B.M. Klaassen, "RF-distortion in deep-submicron CMOS technologies," in *Electron Devices Meeting*, pp. 807-810, Dec. 2000.
35. K. Koh, H.-M. Park, and S. Hong, "A spline large-signal FET model based on bias-dependent pulsed I-V measurement," *IEEE Trans. Microwave Theory Tech.*, vol. 50, pp. 2589-2603, Nov. 2002.
36. T. Y. Lee and Y. Cheng, "MOSFET HF distortion behavior and modeling for RF IC design," in *CICC Conf. Dig. Tech. Papers*, Sep. 2003, pp. 87-91.
37. I. Angelov, H. Zirath, and N. Rorsman, "Validation of a nonlinear transistor model by power spectrum characteristics of HEMT.s and MESFET.s," *IEEE Trans. Microwave Theory Tech.*, vol. 43, pp. 1046-1052, May 1995.
38. M.W. Pospieszalski, "Modeling of Noise parameters of MESFETs and MODFETs and Their Frequency and Temperature Dependence," *IEEE Trans. on Microwave Theory and Techniques*, vol. 37, pp.1340-1350, Sept. 1989.
39. A. Van Der Ziel, *Noise in Solid State Devices and Circuits*, Wiley, New York, 1986.
40. A. Van Der Ziel, "Thermal Noise in Field Effect Transistors," *Proc. IEEE*, pp.1801-12, Aug. 1962.
41. A. Abidi, "High-frequency noise measurements on FETs with small dimensions," *IEEE Trans. Electron Devices*, vol. 33, pp. 1801-1805, Nov. 1986.
42. C. McAndrew, G. Coram, A. Blaum and O. Pilloud, "Correlated Noise Modeling and Simulation," *Workshop on Compact Modeling*, pp. 40-45, 2005.
43. A. J. Scholten, L. F. Tiemeijer, R. van Langevelde, R. J. Havens, A. T. A. Zegers-van Duijnhoven, V. C. Venezia, "Noise Modeling for RF CMOS Circuit Simulation," *IEEE Trans. on Electron Devices*, vol. 50, pp.618-632, March 2003.
44. <http://www.vtt.fi/>.

Chapter 4

Amplifiers and Mixers

Ali M. Niknejad, Brian A. Floyd, Sohrab Emami, Babak Heydari, Ehsan Adabi, Bagher Afshar

4.1 60 GHz Low-Noise Amplifiers: What's Different?

The key performance requirements of the 60 GHz low-noise amplifier (LNA) are power gain, noise figure, linearity, stability, impedance matching, power dissipation, bandwidth, and design robustness to process/voltage/temperature variation. These basic requirements are universal for LNAs, and as will be shown, the basic design methodologies at 60 GHz are not all that different than those at much lower frequencies. The circuit topologies, however, will be different to account for the three fundamental differences of 60 GHz design compared to lower frequency design, which are (1) designing using transistors operating much closer to their cutoff frequencies¹, (2) operating with signals with small wavelengths resulting in distributed effects within actual components of the circuit, and (3) designing with parasitic elements which represent a much larger portion of the total impedance or admittance on a given node. The implications of these three differences are now briefly discussed, and then illustrated through circuit examples later on in the chapter.

4.1.1 Transistors Closer to Cutoff

Transistors operating closer to their cutoff frequencies have less gain and higher noise figure. A useful way to estimate the maximum available power gain, G_{MAG} , in the transistor near cutoff is to compute how many decades below cutoff the operating

Niknejad, Emami, Heydari, Adabi, and Afshar:
Berkeley Wireless Research Center, University of California, Berkeley, e-mail: niknejad@berkeley.edu

Brian A. Floyd:
IBM T. J. Watson Research Center, Yorktown Heights, NY, e-mail: brianfl@us.ibm.com

¹ True at least at the time of this writing.

frequency lies and then multiplying by 20 dB, as follows:

$$G_{MAG} \approx 20 \log_{10} \left(\frac{f_{MAX}}{f_o} \right), \quad (4.1)$$

where f_o is the operating frequency and f_{max} is the unity power-gain frequency. Note that this is valid only near cut-off where the available power gain in the transistor is equal to the maximum available gain. At lower frequencies, the transistor has enough gain to be potentially unstable; thus, **the suitable power-gain metric is the maximum stable gain (MSG) which no longer has the 20-dB/decade slope.**² As an example, a transistor with an f_{max} of 240 GHz should have a G_{MAG} of approximately 12 dB at 60 GHz.

As a result of the lower available gain per transistor, multi-stage topologies are likely required to provide enough gain to suitably deemphasize the noise contribution from the mixer and subsequent stages. Multi-stage amplifiers consume more power and area, have poorer linearity, and have more internal nodes which require stability checks. The noise performance of a multi-stage amplifier may or may not deviate from the minimum achievable noise performance in the technology, depending on whether or not each gain stage is designed for minimum noise or maximum gain. Reduced transistor gain means less margin for process/temperature/voltage variations, less margin for de-Q'ing output loads (for broader bandwidth response), and less margin for series or shunt feedback. A multi-stage amplifier will have more cascaded tuned amplifier responses, each with their own temperature degradation. As a result, a larger temperature variation can be expected. In summary, operating closer to device cutoff means poorer performance across the board for the LNA.

4.1.2 Small Wavelengths

At 60 GHz the wavelength of light in free-space is approximately 5 mm. In SiO₂, the wavelength is roughly halved or 2.5 mm. When signals traverse components which are an appreciable size of a wavelength, then they actually betray their true nature of being waves rather than simple stationary voltages and currents, resulting in a noticeable phase delay across that component. As a result, distributed effects must be considered as part of the design process. One rule-of-thumb for the distance which signifies the *lumped/distributed* boundary is 5% of the wavelength, or 125 μ m at 60 GHz in SiO₂. In practice this boundary has been reduced to about 1% of the wavelength, or 25 μ m at 60 GHz in SiO₂, to accurately model the circuit performance.

Operating in the distributed regime is nothing to fear. Anyone who has designed circuit boards in the low GHz frequency range is already familiar with the fundamentals and the basic toolbox. One implication of the distributed regime is that

² Alternately, the unilateral power gain or Mason's gain can be used with this simple equation, (4.1).

impedance matching is required on internal mm-wave nodes within the circuit to ensure maximum power transfer. Another implication is that any interconnect within the circuit which is an appreciable size of a wavelength should be treated as a transmission line and accurately modeled. Transmission lines, therefore, become very important elements in the entire millimeter-wave portion of the radio, as they are used as both interconnects and to realize passive components. A final implication (or benefit, actually) of operating in the distributed regime is that many traditional microwave structures, such as hybrid couplers, now become viable on-chip. Thus, the 60 GHz circuit designer has a wider pallet of devices at his or her disposal, each with certain trade-offs.

4.1.3 Parasitics at 60 GHz

The parasitic reactive components one cares about in a circuit are shunt capacitance on a node, creating an admittance which is proportional to frequency, and series inductance on a node, creating an impedance proportional to frequency. Both of these worsen as frequency increases. As an example, a lumped capacitance value at 60 GHz may be ~ 30 fF (or $-j88 \Omega$). A parasitic capacitance of 3 fF represents 10% of the total capacitance on that node and can therefore shift the resonant frequency by $\sqrt{1.1}$ or 5%. Clearly, 3 fF parasitic capacitances are commonplace in any design, and they must therefore be accurately extracted and modeled. Turning to an inductor example, a lumped inductor value at 60 GHz may be 150 pH (or $+j56 \Omega$); thus, 15 pH represents 10% of the total series inductance on that node. Technologies with reasonably thick back-ends-of-the-line (say, $10 \mu\text{m}$) can have large vias from the top level metal to the front-end devices, and this via can exhibit a noticeable inductance. As an example, a via inductance of ~ 4 pH was found both empirically and through electromagnetic simulation for a via of $4\text{-}\mu\text{m}$ height. This value of $1 \text{ pH}/\mu\text{m}$ is equivalent to the 1-nH/mm rule-of-thumb for bondwires.

Another parasitic component which is relevant is series resistance. At 60 GHz the skin-depth in the copper and aluminum metal alloys used in most foundries is approximately $0.5 \mu\text{m}$. Top-level metals thicker than this will exhibit frequency-dependent losses, but these layers will have the smallest loss per unit width. Also, ground planes thinner than the skin depth will not completely terminate electromagnetic fields from above; thus, additional current can flow beneath these very thin ground planes in other layers which may be more resistive (like the silicon substrate), resulting in more loss than intended.

4.2 Low-Noise Amplifier Design Methodology

Let us now review some basic relationships between gain, noise figure and linearity. As its name suggest, the low-noise amplifier should provide moderate to high power

gain while introducing a minimal amount of noise. The gain in the LNA deemphasizes noise contributions from subsequent receiver stages by amplifying the noise present at the input by a large enough value such that subsequent noise contributions are small by comparison. Therefore, the signal-to-noise ratio (SNR) of the entire receiver is effectively set to the SNR at the output of the LNA. This can be seen through Friis formula for cascaded noise factor, as follows [1]:

$$F = \frac{(S/N)_{IN}}{(S/N)_{OUT}} = F_{LNA} + \frac{F_2 - 1}{G_{LNA}} + \frac{F_3 - 1}{G_{LNA} \cdot G_2} + \dots \quad (4.2)$$

where F_i and G_i are the noise factors and available power gains, respectively, of individual stages in the receiver. Thus, to minimize the total receiver noise figure, the first stage in the receiver should amplify the input signal by as large an amount as possible while adding minimal noise.

The LNA gain cannot be arbitrarily increased, however, due to its negative impact on large-signal performance. Clearly, the gain present in the LNA makes it more difficult for subsequent receiver stages, such as the mixer, to remain linear. Linearity is often quantified in terms of a third-order intercept point (IP3) as well as a 1-dB compression point. The expression for the total IIP3 for a cascaded system can be expressed as follows [2]:

$$\frac{1}{iIP3_T} = \frac{1}{iIP3_{LNA}} + \frac{G_{LNA}}{iIP3_2} + \frac{G_{LNA} \cdot G_2}{iIP3_3} + \dots \quad (4.3)$$

where $iIP3_i$ and G_i are the input-referred IP3 (in watts) and the power gain (in watts/watt) of each individual stage of the receiver. For gains greater than one, the linearity of latter stages will dominate the total receiver linearity. Therefore, to maximize the receiver's IIP3, the latter stages' linearity should be maximized while reducing or limiting the total preceding gain.

Equations (4.2) and (4.3) imply an acceptable range of LNA gains which will meet both the system's noise figure and linearity requirements. It should be pointed out that these requirements do not always have to be met concurrently, in which case it is possible to design the LNA to have adjustable or switched gains to allow for a wider receiver dynamic range.

4.2.1 Input Match Optimization for Noise and Power

The heart of any LNA design is the input impedance match which is optimized for minimum noise factor together with maximum power transfer. A *fundamental objective of the LNA is to achieve a coincident input power match and input noise*

match. The most insightful way to examine power and noise matching trade-offs is to utilize noise parameters. Any two-port network can be represented with four noise parameters which describe how the noise factor behaves with respect to the input match [3]. A variety of noise factors can be used, with the three most common sets shown below.

$$F = F_{\min} + \frac{R_n}{G_S} |Y_S - Y_{opt}|^2 \quad (4.4)$$

$$F = F_{\min} + \frac{G_n}{R_S} |Z_S - Z_{opt}|^2 \quad (4.5)$$

$$F = F_{\min} + 4r_n \frac{|\Gamma_S - \Gamma_{opt}|^2}{|1 + \Gamma_{opt}|^2 (1 - |\Gamma_S|^2)} \quad (4.6)$$

In all three, the noise factor reaches a minimum value, F_{\min} , at the optimum input impedance match, expressed in either admittance (Y_{opt}), impedance (Z_{opt}), or reflection coefficient (Γ_{opt}). The sensitivity of noise factor with respect to impedance mismatch is captured by either the noise resistance (R_n) or noise conductance (G_n), depending on whether the admittance or impedance domain is used. Larger noise resistances (or conductances) result in greater sensitivity to noise mismatch.

To obtain a low noise factor, the LNA design should first provide for the minimum F_{\min} possible by selecting the optimum bias point for the transistor, the optimum layout for that transistor, and of course the appropriate base technology for the transistor with suitable f_T/f_{max} . Second, the LNA design should provide an input matching network which transforms the source impedance (admittance) to Z_{opt} (Y_{opt}), where Z_{opt} should be designed to coincide with the matching condition for maximum power transfer. This is to say that Z_{opt} *should be the complex conjugate of the input impedance to the amplifier, Z_{in}* . Failing to make the noise match and power match coincident results in either sub-optimal input return loss or sub-optimal noise figure. Noise figure calculations which directly compute F for a 50Ω source obscure the existence and relevance of Z_{opt} (Y_{opt}) and hide the fundamental trade-off which exists between noise and power matching.

The noise and power match requirement is illustrated in Fig. 4.1, for reflection coefficients rather than input impedances. Here, the basic input-match requirement for the LNA is to make the power and noise match coincident, meaning that Γ_{in} and Γ_{opt} should be complex conjugates of one another.

4.2.2 Transistor Noise Parameters

The noise parameters of the transistor can be derived by equating the basic “noisy” circuit network with a “noiseless” network with input voltage and current noise sources, v_n , and i_n , as shown in Fig. 4.2. By equating the short-circuit output current in both networks for open-circuited and short-circuited input conditions, i_n and v_n can be derived, respectively [4].

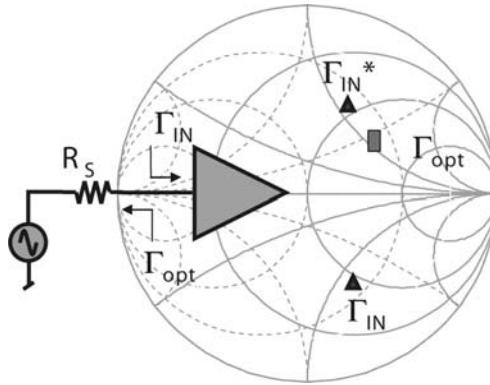


Fig. 4.1 Illustration of impedance matching requirements for LNA together with Smith chart. The input reflection coefficient Γ_{in} should be the complex conjugate of the optimum noise reflection coefficient Γ_{opt} .

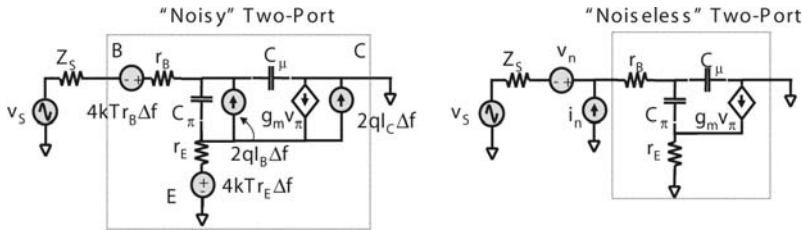


Fig. 4.2 Schematic of BJT with dominant noise sources defined (left) and equivalent input noise generators (right).

Figure 4.2 shows the dominant noise sources for a BJT³. The noise resistance and conductance which define the noise power in v_n and i_n , respectively, can be readily derived to be

$$R_n = \frac{\overline{v_n^2}}{4kT\Delta f} \approx \frac{1}{2g_m} + r_B + r_E \quad (4.7)$$

$$G_n = \frac{\overline{i_n^2}}{4kT\Delta f} \approx \frac{g_m}{2} \left(\frac{\omega}{\omega_T} \right)^2 \quad (4.8)$$

Since these two equivalent input noise sources (v_n and i_n) originate from the same noise sources in the circuit (e.g., shot noise and thermal noise), they are necessarily correlated. A noise correlation power, C_n , can be defined as

³ Note that the noise sources for a CMOS transistor are located in the same positions, where the collector and base shot noises should be replaced with thermal noise in the channel lumped to the drain or the gate (also known as gate-induced noise), respectively, and base/emitter resistance should be replaced with gate/source resistance [5].

$$C_n = \frac{\overline{i_n^* v_n}}{4kT\Delta f} \approx \frac{-j}{2} \left(\frac{\omega}{\omega_T} \right). \quad (4.9)$$

Note that these parameters, R_n , G_n , and C_n , are another valid set of noise parameters, where C_n can have both real and imaginary components. These four noise parameters can be translated into the noise parameters of interest in (4.5), through the following formulas [6]:

$$F_{\min} = 1 + 2\operatorname{Re}(C_n) + 2\sqrt{G_n R_n - [\operatorname{Im}(C_n)]^2} \quad (4.10)$$

$$Z_{\text{opt}} = \sqrt{\frac{R_n}{G_n} - [\operatorname{Im}(C_n)]^2} - j\operatorname{Im}(C_n). \quad (4.11)$$

Substitution of (4.7)-(4.9) into (4.10) and (4.11), yields

$$F_{\min} \approx 1 + \left(\frac{\omega}{\omega_T} \right) \sqrt{2g_m(r_B + r_E)} \quad (4.12)$$

$$Z_{\text{opt}} \approx \left(\frac{\omega_T}{\omega} \right) \sqrt{\frac{2(r_B + r_E)}{g_m}} + \frac{j}{\omega(C_\pi + C_\mu)} \quad (4.13)$$

Equation (4.13) is an important result from this exercise, as the optimal LNA should have a Z_{opt} equal to the complex conjugate of the input impedance, Z_{in} .

4.2.3 Common-Base vs. Common-Emitter

Using (4.12) and (4.13), we can now compare the basic input stages for an amplifier in terms of noise and power match, namely the common-base and the common-emitter stages, depicted in Fig. 4.3. The noise parameters in either configuration are identical; however, the input impedances are quite different.

A common-base stage has an input conductance equal to the device transconductance, g_m . If a shunt inductor is used to resonate out the input capacitance, then Z_{in} is simply $1/g_m$. This cannot be made to be equal to the real part of Z_{opt} of (4.13); thus, a common-base LNA input stage suffers from either a poor noise match or a poor power match. This is commonly recognized at lower frequencies and persists at millimeter-wave frequencies.

A common-emitter amplifier without any feedback suffers from a very high quality-factor (low resistance) input impedance which is difficult to impedance match. As a result, feedback is typically employed, with emitter degeneration quite

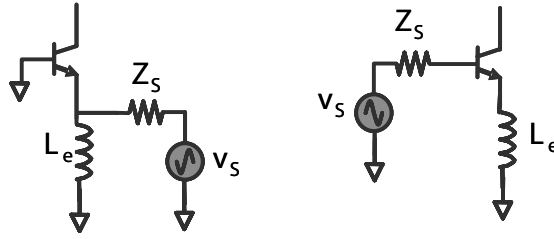


Fig. 4.3 Basic schematic of common-base input stage and common-emitter with degeneration input stage.

common. The common-emitter amplifier with inductive degeneration can achieve a coincident noise and power match, as the emitter inductor provides a degree of freedom. The input impedance of this configuration can be easily derived by reflecting the emitter impedance into the base by multiplying it by one plus the current gain ($\beta(\omega) = \omega_T / j\omega$). As a result,

$$Z_{in} = \omega_T L_e + j\omega L_e + \frac{1}{j\omega(C_\pi + C_\mu)} \quad (4.14)$$

It can be shown that the noise parameters of a network with ideal feedback are identical with respect to F_{min} and R_{opt} , where only X_{opt} has been reduced by the series feedback (i.e., the impedance in the emitter) [7]. Examining (4.13) and (4.14), it can then be seen that the noise and power match can be made coincident by selecting the emitter degeneration to form the real part of the input impedance (typically 50Ω) and then by selecting the size of the transistor to make R_{opt} 50Ω as well.

The simple design procedure for the LNA input stage can be summarized as follows:

1. Simulate noise measure (NM)⁴ versus collector current density (J_C) to choose optimal *range* of collector current density (J_C) for minimum noise measure.
2. Scale emitter length such that $R_{opt} = 50\Omega$, increasing device size (emitter length) to reduce the optimum noise impedance.
3. Add degeneration inductor to create real input impedance for power match.
4. Complete input match with series inductor in the base.

It should be pointed out that this is the same design methodology advocated at lower frequencies by Voinigescu [5], and it has been found to work at mm-wave frequencies as well. At mm-wave, the device physics of the transistor have not changed;

⁴ Noise measure is the noise figure of an arbitrarily long cascade of identical amplifiers. It is a more useful metric for noise performance of amplifiers with low gain, since noise contributions from second-, third-, etc. stage amplifiers becomes relevant [3].

thus, it should come as no surprise that the methodology to achieve optimum noise performance has not changed as well. Note that the procedure listed about is bipolar-centric; however, the same procedure also applies to MOSFETs as well, with the appropriate substitution of noise parameter quantities and device nomenclature.

Finally, one additional qualification for this design methodology is required. This procedure assumes that the amplifier is unilateral. That is, it assumes that the reverse isolation is quite good, meaning that the output impedance matching network does not change the input impedance. This is typically true of cascade amplifiers; however for single-transistor amplifiers, particularly at 60 GHz, the reverse isolation can be low. As a result, the concept of simultaneous conjugate input and output matching can be adopted [3]. This means that the input power matching goal of 50Ω should be replaced with a goal of achieving the simultaneous conjugate match condition. This then requires that the output matching network for the amplifier be designed as well for simultaneous conjugate matching.

4.3 Low-Noise Amplifier Examples

Several bipolar and CMOS LNAs are now presented. We begin with two bipolar amplifiers, one with a common-base input stage and one with a degenerated common-emitter input stage. Next we consider several CMOS amplifier design approaches. These examples show mm-wave LNA design techniques and topologies and also illustrate the performance differences between the various approaches.

4.3.1 Bipolar LNA (v1), Common-Base Input

A simplified schematic of the first version of a bipolar 60 GHz LNA is shown in Fig. 4.4 [8], [9], [10]. This amplifier was implemented in a $0.13\text{-}\mu\text{m}$ bipolar-only process with NPN HBTs with 200 GHz f_T and 280GHz f_{max} [11]. The input stage is a common-base amplifier, which exhibits higher gain compared to the degenerated common-emitter⁵. Also, its high reverse isolation decouples the input and inter-stage matching networks. The second stage of the LNA is a common-emitter cascode amplifier with emitter degeneration. Each stage provides ~ 8 dB of gain when biased at 3 mA and $V_{CC} = 1.8$ V.

Impedance matching at the input, inter-stage, and output are realized with single-stub tuners. This is an example of a distributed-based network topology. An advantage of the single-stub tuner is that it does not require prohibitively small series capacitors, since the network is realized solely with transmission lines. Operation of the tuner is best described through an example of matching a given admittance, Y_L , to 50Ω . The series line first transforms Y_L to $Y' = 0.02 + jB$, i.e., rotating along a constant

⁵ Portions of this text is taken from [10], (© IEEE 2005)

VSWR circle on the Smith chart. The shunt stub then generates an admittance of $-jB$ which adds directly to Y' , i.e., moving along the $0.02\Omega^{-1}$ conductance circle to the center of the Smith chart, resulting in a matched condition. The stub lengths are $\sim 100\ \mu\text{m}$ ($TL_{4,7}$) while the series transmission line lengths are $\sim 300\ \mu\text{m}$ ($TL_{3,6}$). Additionally, this amplifier uses metal-insulator-metal (MIM) AC coupling capacitors ($C_{2,3}$) operating beyond self-resonance are used between the two stages and at the output.

At 60 GHz, the silicon substrate exhibits low impedance; thus, care was taken in circuit design and layout to ensure adequate reverse isolation and stability. Metal-1 ground shields are used throughout the amplifier, and substrate ties are included wherever possible. Also, metal-2 V_{CC} planes are used together with many MIM bypass capacitors to realize a low-impedance supply. A die photograph of the LNA is shown in Fig. 4.5. The die size is $0.9 \times 0.6\ \text{mm}^2$. Diamond-shaped pads are used at the input and output to reduce the parasitic capacitance of the pad.

The measured S -parameters for the LNA are shown in Fig. 4.6, together with the simulated results. The LNA is unconditionally stable over 30–110 GHz. At 61.5 GHz, the gain is 14.7 dB and the reverse isolation is 40 dB, when biased at 6 mA from a 1.8-V supply. The input return loss is 6 dB, while the output return loss is 17 dB. Model-to-hardware correlation is excellent for S_{21} , S_{12} , and S_{22} . The miss in S_{11} , though, was not expected, particularly given that the output match was right on target. An alternate version of the LNA which had coplanar waveguide (CPW) tapers at the input and output [9] showed S_{11} and S_{22} better than -12 dB, as expected. Since the CPW taper absorbs pad parasitics, the miss in input match for the *pad*-LNA is attributed to the input bondpad transition.

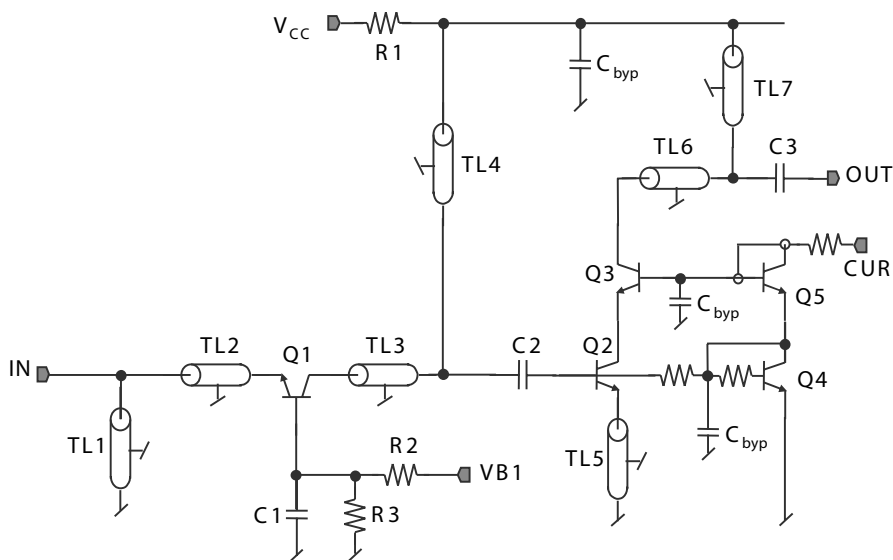


Fig. 4.4 Simplified schematic of bipolar LNA, version 1, common-base input [10] (© IEEE 2005).

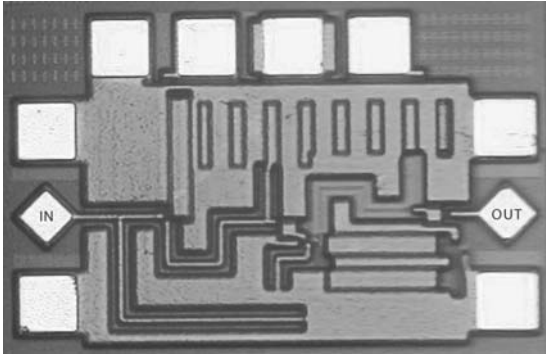


Fig. 4.5 Die micrograph of the common-base 60 GHz LNA, version 1 [10] (© IEEE 2005).

Figure 4.7 shows the measured and simulated NF of the LNA (with pads) at 6 mA and 1.8 V. No on-chip loss has been de-embedded from this result. The measured NF is 4.5 dB at 61.5 GHz, while the simulated NF is 4.6 dB. The taper-LNA also has a NF of 4.5 dB when the insertion loss of the tapers is de-embedded. The simulated minimum NF of the LNA is 4.2 dB, while the NF_{min} of a single transistor is 3.1 dB. From simulation, the input common-base device contributes 80% of the added noise (split nearly equally between collector shot noise and thermal noise from base resistance), while the second-stage cascode contributes 10%. The remaining 10% comes from other assorted sources. As expected, under this good noise matching condition, the input power match is poor, with an input return loss of only 6 dB. Table 4.1 summarizes the other performance metrics for this LNA, including linearity, power consumption, and reverse isolation.

Table 4.1 Comparison between bipolar 60 GHz LNAs, version 1 and 2

LNA	Version 1	Version 2
Topology	2-stage: Common-Base Cascode	4-stage: Common-Emitter (x2) Cascode (x2)
Spot Freq.	61.5 GHz	61.5 GHz
Gain	15 dB	20 dB
Image Rejection	> 15 dB	> 26 dB
NF	4-6 dB	4.8-6.2 dB
S11	-6 dB	-15 dB
S22	-17 dB	-15 dB
iCP1dB	-20 dBm	-29 dBm
VCC	1.8 V	2.7 V
Current	6 mA	10 mA

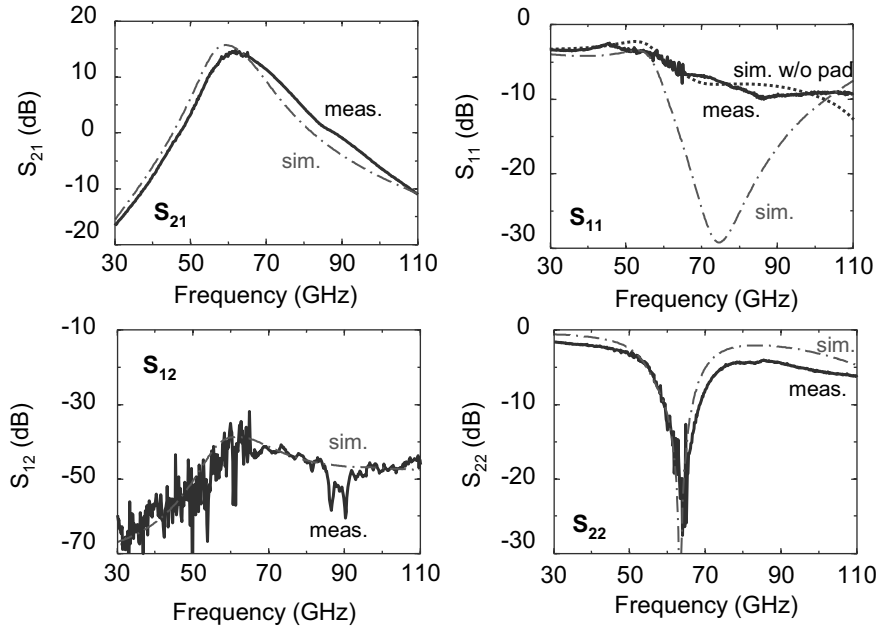


Fig. 4.6 Measured and simulated S -parameters for common-base 60 GHz LNA, version 1 [10] (© IEEE 2005).

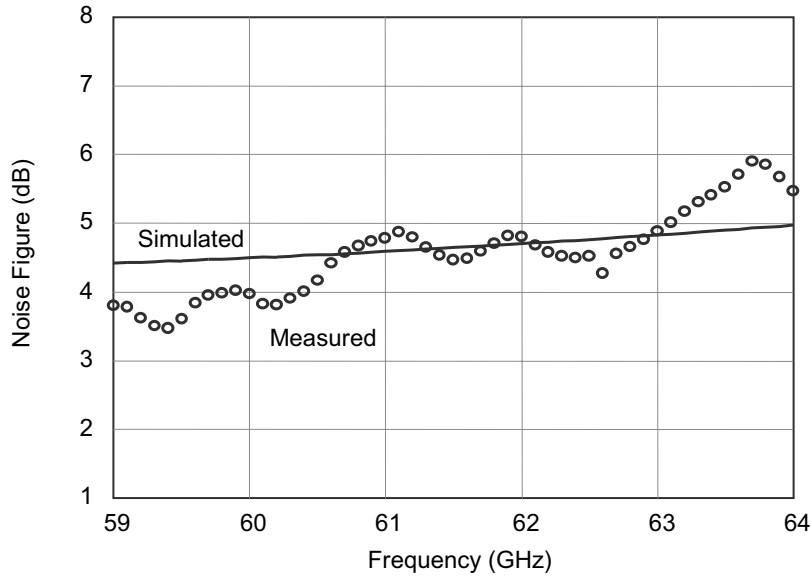


Fig. 4.7 Measured and simulated noise figure of 60 GHz common-base LNA, version 1 [10] (© IEEE 2005).

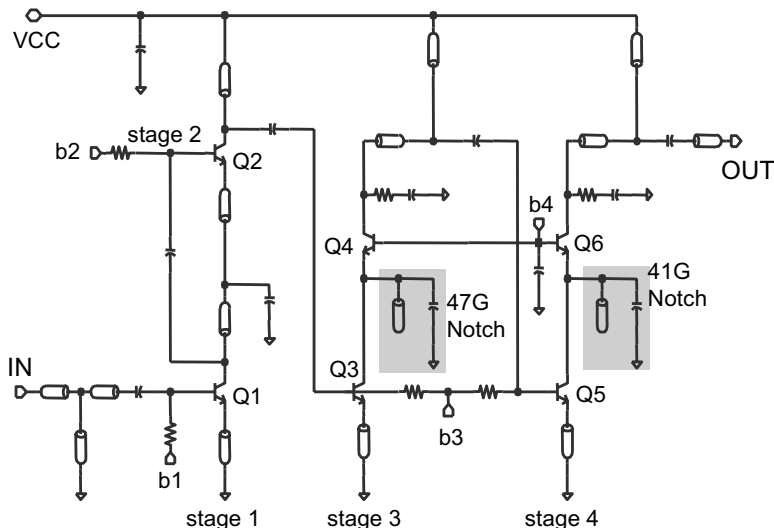


Fig. 4.8 Simplified schematic of 4-stage, 60 GHz bipolar LNA, version 2, common-emitter input [14] (© IEEE 2006).

4.3.2 Bipolar LNA (v2), Common-Emitter Input

Fig. 4.8 shows a simplified schematic of a second version of a bipolar 60 GHz LNA [12], [13], [14]. This LNA was implemented in a $0.13\text{-}\mu\text{m}$ SiGe BiCMOS technology which features NPN HBTs with $200\text{ GHz } f_T$ and $265\text{ GHz } f_{MAX}$. This LNA was designed for a complete 60 GHz receiver; thus, it has additional features and requirements. First, it was designed to have $\sim 20\text{ dB}$ of gain to achieve the lowest possible noise figure for the receiver. Second, it was designed to include embedded image-reject notch filters, since the receiver employs a superheterodyne topology. In the receiver, the image frequency is between 42 and 46 GHz; hence, notches were designed at 41 and 47 GHz to bracket this image band. Finally, the LNA was designed to work with the receiver's 2.7-V supply voltage.

The LNA is a four-stage design. The first two stages (transistors Q1 and Q2) are inductively-degenerated common-emitter amplifiers. These amplifiers are designed for minimum noise figure at $\sim 4\text{ dB}$ per stage and provide roughly 5 dB of gain in each stage. Stages 1 and 2 are stacked one on top of the other for current re-use. Stages 3 and 4 in the LNA are cascode amplifiers. Image rejection is provided in these two stages by notch filters at 47 and 41 GHz. Each notch filter, placed at the junction between the two cascode devices (Q3-Q4 and Q5-Q6), contains an open-circuited microstrip transmission-line stub, $\lambda/4$ in length at the notch frequency. The quarter-wavelength transforms the open-circuit on the far end of the line to a short-circuit at the near end of the line; therefore, all small-signal current coming out of the common-emitter stage of the cascade is shunted to ground. An additional shunt capacitor is included with the stub to resonate out the incremental inductance which

exists at the RF frequency (60 GHz). The resultant response from the filter is a series resonance to ground at the image and a parallel resonance at RF.

The noise figure (NF) of each cascode amplifier is approximately 6 dB; hence, stages 1 and 2, each with 5 to 6-dB gain and approximately 4-dB NF, were added to reduce the cascaded NF and increase gain. Fig. 4.9 shows the measured and simulated S -parameters of the LNA while Fig. 4.10 shows the measured noise figure. The measurements indicate 5 to 6.2-dB NF and 20-dB gain. Additional LNA measurements are summarized in Table 4.1, including power consumption, linearity, return loss, and image rejection.

Comparing versions 1 and 2 of the bipolar LNAs (see Table 4.1), we see that the common-emitter and common-base amplifiers have very similar noise figure performance, achieving in the realm of 5-dB noise figure. However, the common-base input stage has a poor input return loss of only 6 dB, while the degenerated common-emitter amplifier has a good input return loss of about 12 dB. This confirms our earlier assertion that the degenerated common-emitter topology provides a better input power *and* noise match than the common-base topology. Both amplifiers provide adequate RF power gain while consuming 10 to 30 mW of dc power.

4.3.3 CMOS Common Source Amplifiers

In this section we describe the design of two common source mm-wave amplifiers reported in [15]. A low power LNA was designed to operate in the 60 GHz band, while a prototype 104 GHz amplifier demonstrates the possibility of designing a linear circuit at a frequency higher than the f_T of the technology node⁶.

Many reported mm-wave CMOS amplifiers use cascode devices [17] [18] [19]. One advantage of the cascode device is that they can be made unconditionally stable at the operating frequency, making the design more robust and simplifying matching networks. However the noise figure of a cascade is relatively high compared to common source device when operating close to f_T due to the reduced degeneration on the cascade device at high frequencies. For this reason, common source devices are preferred for ultra low noise designs close to the limits of activity. The round-table common source device described in Section 3.2.2 are conditionally unstable up to 95 GHz, requiring careful design of the source and load networks. Consequently for the 60 GHz LNA, the source/load impedance for each stage should be selected to be as far as possible from the instability regions while maintaining a good gain/noise performance [3]. The maximum stable gain of common source devices with similar layout methods is more or less independent of the width, enabling low power design by using smaller devices.

CPW transmission lines are used extensively in the design for impedance matching, interconnect wiring, and the bias networks. Transmission line lengths are kept much shorter than $\lambda/4$ in order to reduce losses and minimize the noise contributions

⁶ Portions of this section are taken from [15] and [16], (© IEEE 2007)

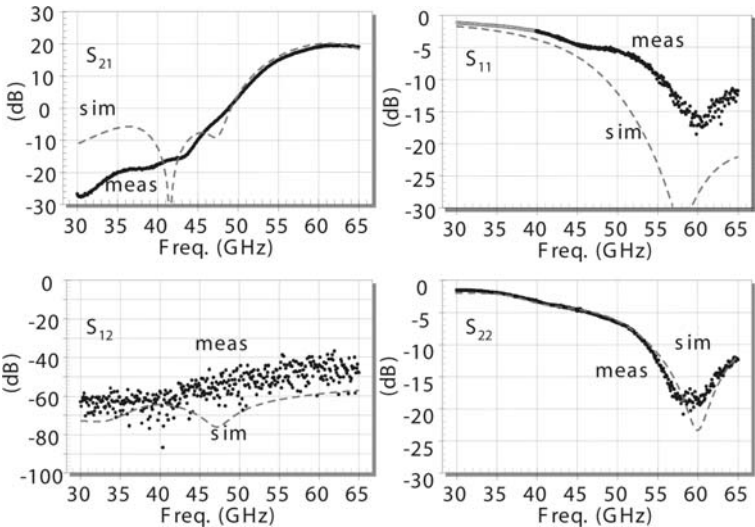


Fig. 4.9 Measured and simulated S -parameters of 4-stage 60 GHz bipolar LNA, version 2, common-emitter input.

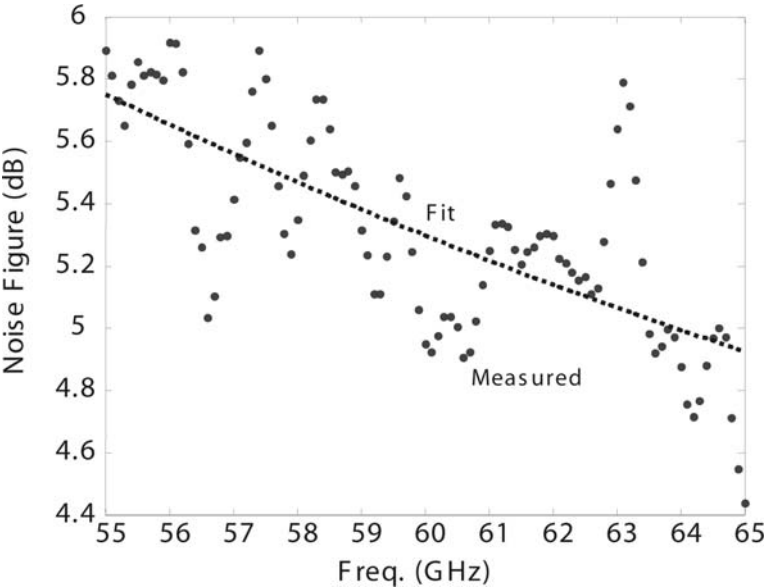


Fig. 4.10 Measured noise figure of 4-stage 60 GHz bipolar LNA, version 2, common-emitter input.

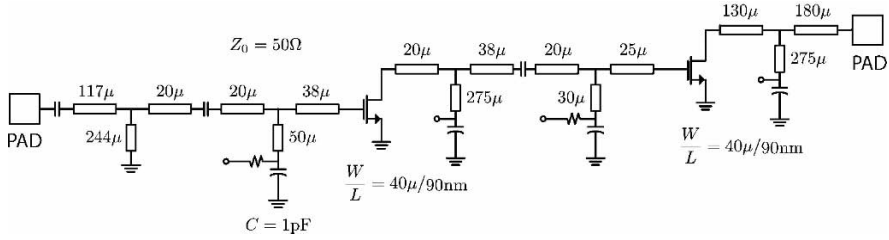


Fig. 4.11 Schematic of 60 GHz common source amplifier [15] (© IEEE 2007).

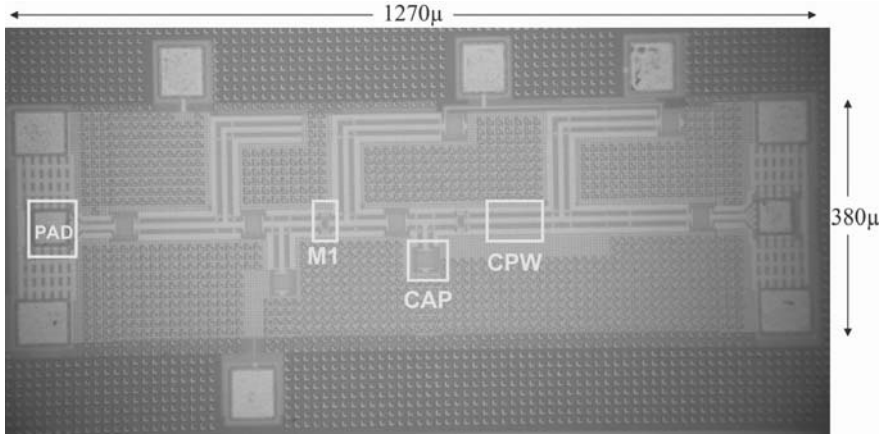


Fig. 4.12 Die photo of 60 GHz common source amplifier [15] (© IEEE 2007).

at the input of the amplifier. Finger MOM capacitors are used for ac-coupling of the input and output as well as between stages. They are also used as bypass capacitors for the DC feed lines. These capacitors are non-ideal and behave inductively at 60 GHz due to their low self-resonance frequency (47 GHz). This does not cause any issue as long as they are well modeled and used as a part of the matching network to shorten the length of transmission lines.

The schematic and the microphotograph of the 60 GHz LNA is shown in Fig. 4.11 and Fig. 4.12. Two $W = 40\mu\text{m}$ round-table devices are used running 6.5mA and 4mA from a 1V power supply. A grounded transmission line is used in before the first stage to suppress low frequency gain peaking of the circuit. Input and output pads are also modeled and used as part of matching networks. The output port is designed to match to 50Ω . The input matching network is simultaneously optimized for noise and power to get at least -10 dB of input reflection and 6 dB of noise figure at the center frequency.

The 104 GHz prototype amplifier is essentially very similar to the 60 GHz LNA. NMOS devices $W = 50\mu\text{m}$ width are unconditionally stable at this frequency and employed in three amplification stages. Compared to the 60 GHz amplifier, the length

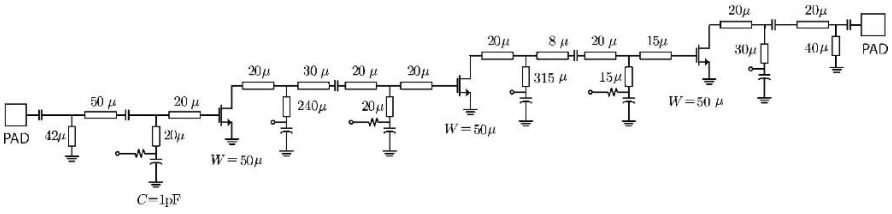


Fig. 4.13 Schematic of 100 GHz common source amplifier [16] (© IEEE 2007).

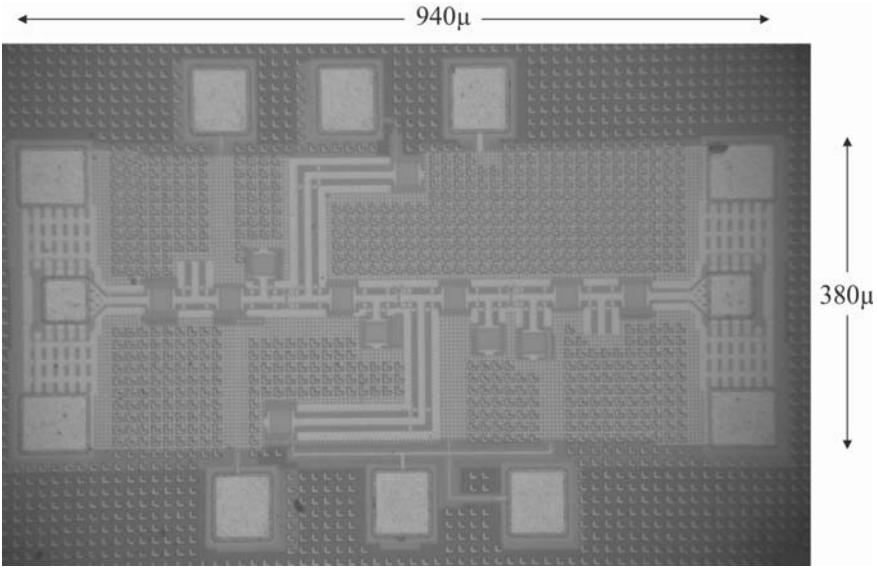


Fig. 4.14 Die photo of 100 GHz common source amplifier [15] (© IEEE 2007).

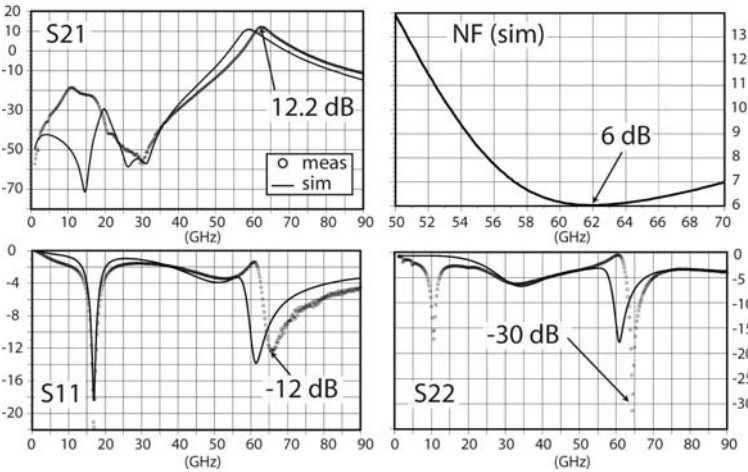


Fig. 4.15 Measured and modeled *S*-parameters for the 60 GHz common source amplifier [15] (© IEEE 2007).

of transmission lines are naturally shorter and a significant part of matching networks is also now realized through the series inductance of coupling and bypass capacitors. This design was based on de-embedded measurement data of the active devices and models of passive devices. Fig. 4.13-4.14 show the schematic and micrograph of the circuit

The measured and modeled S -parameters for the 60 GHz amplifier are shown in Fig. 4.15. The amplifier achieves a peak power gain of 12.2 dB at 63 GHz and the input and output return losses are -13 dB and -25 dB respectively. The measurement is in a good agreement with the simulation with 1 GHz of frequency mismatch. Also the output match of the circuit has discrepancy with the simulated result, pushing the circuit to the edge of instability. Process variation could explain the frequency mismatch, and based on the post measurement simulations, it was verified that a 5% reduction in T_{ox} could match the measured center frequency with the simulation. The source of mismatch in the S_{22} is not completely clear; however a variation in the drain to body capacitance could also explain this discrepancy. As the drain to body capacitance decreases, a larger share of the signal flows to the output and causes S_{22} to increase. This also causes output gain circles and load stability circles to approach each other. A model of the transistor with T_{ox} and C_{db} variation was compared to the measurements results and shown in Fig. 4.16a-b, which qualitatively resembles the observed deviations in S_{11} and S_{22} .

Fig. 4.17a shows the power sweep measurement of the LNA. The circuit has a measured +4 dBm 1-dB compression point which matches the predicted value based on the large signal model of the device. This translates into 23% of power added efficiency, making it suitable as a pre-driver for a power amplifier or the output stage of a short-range transmitter. The simulated noise figure is 6 dB (Fig. 4.17b) in the pass band of the amplifier while the measured noise figure is 6.5 dB. The measurements is done using a similar amplifier driving a down-conversion mixer and the effect of the mixer noise and cable losses were de-embedded⁷.

The measured S -parameters of the 104 GHz amplifier are shown in Fig. 4.18. The amplifier has the peak gain of 9.34 dB at 103.8 GHz. The input and output reflection coefficients are -9.8 dB and -5.5 dB respectively. The circuit draws 22mA from a 1V power supply.

4.3.4 CMOS Common Gate Amplifiers

Common gate amplifiers are seen much less frequently at mm-wave frequencies, despite the fact that they provide an active broadband input impedance match. The downside of the common gate amplifier is much lower power gain (since the current gain is unity) and much higher noise. The full drain noise flows into the input circuit without attenuation (due to unity current gain). Under an input match condition, $1/g_m = R_0$, the drain current contributes nearly equal noise to the output as the

⁷ The amplifier also featured a small amount of inductive degeneration added for stability. The simulated noise figure was the same as the tested amplifier.

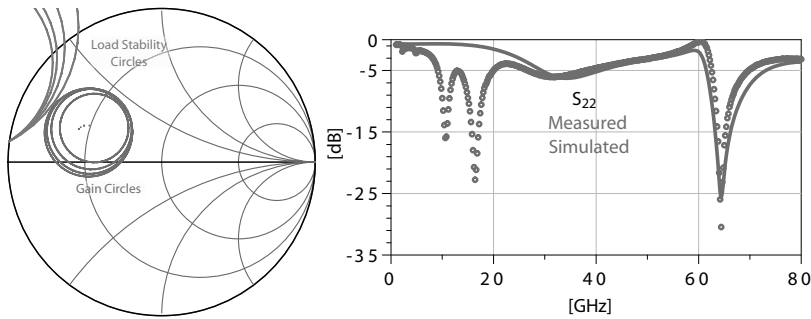


Fig. 4.16 The effect of the drain-body capacitor on gain and stability circles (left) and output scattering parameter S_{22} (right) for 20% variation in C_{db} [16] (© IEEE 2007).

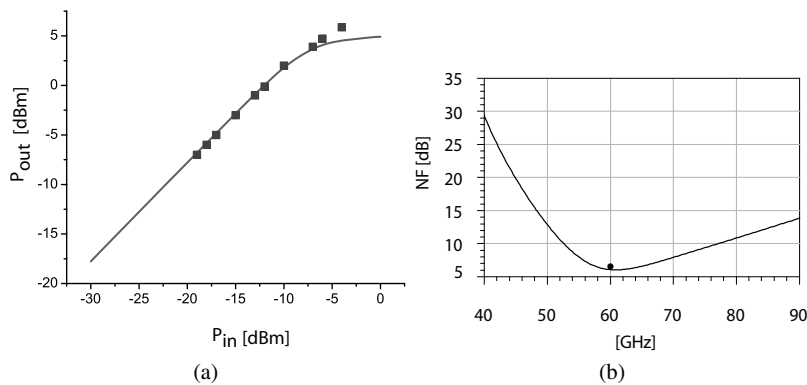


Fig. 4.17 a) Measurement of the output power of the 60 GHz common source amplifier. (b) Simulated and measured (single data point) noise figure of LNA [16] (© IEEE 2007).

source, setting a lower limit to the noise figure at 3dB. Despite these shortcomings, some researchers have found the performance to be adequate at microwave frequencies and advocate the use of the common gate amplifier [20]. It should be noted that the maximum unilateral gain of a transistor is invariant property, regardless of which terminal is taken as common. Thus the f_{max} of a common gate is equal to a common source or common drain amplifier. Thus one can always increase the gain of a common gate amplifier by using an input matching network.

4.3.5 Differential Pair Amplifiers

Differential amplifiers are used extensively in analog and RF circuit design. Differential operation has several benefits, including common mode rejection at the input and

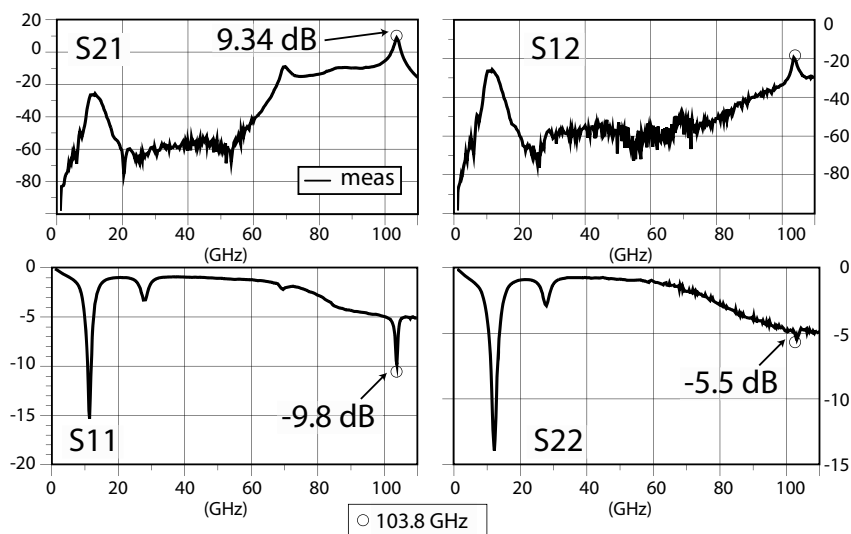


Fig. 4.18 Measurement of the S -parameters of the 100 GHz common source amplifier [15] (© IEEE 2007).

power/ground noise rejection for fully balanced circuits. Since the integrated circuit environment is noisy as a result of high levels of integration, digital switching noise can be rejected using a balanced approach. This argument is quite valid for the VCO and non-linear or time-varying circuits where lower frequency noise can couple to RF, but given the frequency isolation between the mm-wave band and the baseband operating at the low GHz range, the benefits of a differential amplifier at mm-wave are not so much for isolation but for other reasons. Since fully balanced circuits form virtual ground nodes, the physical ground connection is much less problematic for differential circuits. This is a big benefit in mm-wave design since it is difficult to realize a low impedance ground plane in a modern IC process.

A downside of differential and balanced circuits is the doubling of current and the requirement to operate with differential signals. Today most test equipment is designed to operate into a single-ended environment, making it difficult to directly characterize differential circuits. This limitation is temporary and true differential mm-wave equipment will be introduced in the near future. Even if off-chip signals are single-ended, transformers (baluns) can be used quite easily to convert signals to differential form. Another advantage of a differential amplifier is the higher output swing, which is beneficial in improving the dynamic range of amplifiers. Several groups have demonstrated differential amplifiers at mm-wave frequencies [21] [22].

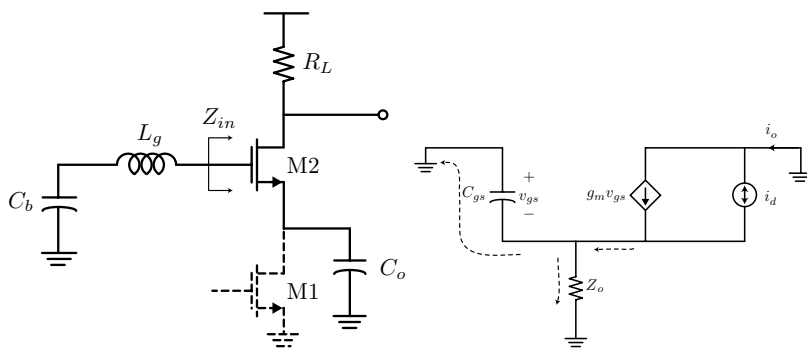


Fig. 4.19 The high-frequency degeneration of a cascode device produces noise and potential instability.

4.3.6 Multi-Stage Amplifier Design

The design of multi-stage amplifiers requires a careful choice in the topology and biasing of each stage. In receiver applications, the input stage should be designed for low noise, preferring single stage (common source or gate) rather than cascode stages. The reduced gain of a single stage amplifier, though, can result in a lower distortion intercept point, since more cascade amplifiers are required to retain the gain. Here we review a few different amplifier approaches with some examples from CMOS. We begin with cascode amplifiers, which can almost be treated as single stage amplifiers, except at high frequencies when we desire to improve the performance.

4.3.6.1 Cascode Noise and Stability

Cascode devices play a crucial role in analog circuit design. At lower frequencies, the main reason for using a cascode transistor is the increase in gain. At mm-wave frequencies, the major benefit is a reduction of the reverse gain S_{12} which in turn results in an increase of the MSG for a given frequency, and which lowers the frequency for which the stability factor k becomes larger than one (unconditional stable device). Regular cascode transistors, however, have a large parasitic capacitance on the interstage node. This capacitance will short-circuit the small signal current at higher frequencies, thus reducing the gain of the cascode structure.

With reference to Fig. 4.19, to calculate the contribution of the cascode device to the noise of the transistor, we can see that the noise has two paths to ground, one through the transistor itself and one through the load, which produces output noise. When the impedance looking into the transistor is much smaller than the current through the load, which occurs at low frequencies, the current circulates through the transistor and has virtually no impact on the output noise. However, as the impedance of the transconductance stage drops due to the capacitance at the shared junction node, the current divides and an appreciable fraction flows to the output

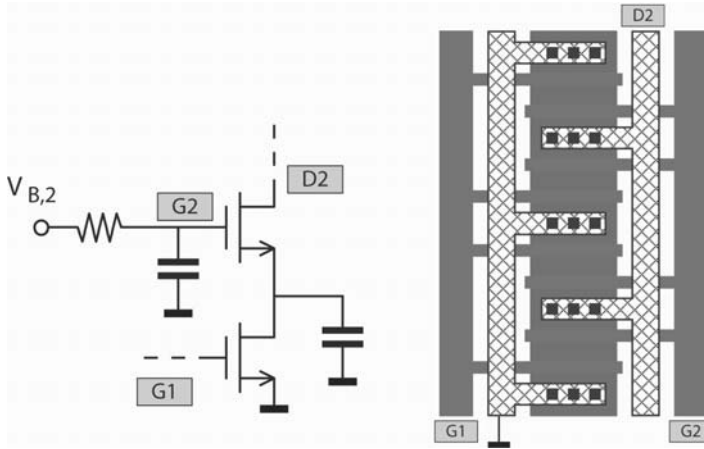


Fig. 4.20 Layout of a shared-junction cascode device [23] (© IEEE 2007).

$$i_o = i_d \frac{1/g_m}{1/g_m + Z_o} = i_d \frac{1}{1 + g_m Z_o} \quad (4.15)$$

$$\overline{i_o^2} = \overline{i_d^2} \left| \frac{1/g_m}{1/g_m + Z_o} \right|^2 = \overline{i_d^2} \left| \frac{1}{1 + g_m Z_o} \right|^2 \quad (4.16)$$

At low frequencies $Z_o \approx r_o$, which makes the noise contribution small. At higher frequencies, $Z_o \approx 1/j\omega C_{gs}$, which shows that at frequencies approaching the device f_T , the noise contribution increases appreciably.

The stability of cascode amplifiers requires special considerations, particularly due to the high frequency capacitive degeneration on the cascode device due to the output capacitance of the transconductance stage. As shown in Fig. 4.19, the impedance looking into the gate of the top transistor at high frequencies is easily shown to be

$$Z_{in} = \frac{1}{j\omega C_o} + \frac{1}{j\omega C_{gs}} - \frac{g_m}{\omega^2 C_o C_{gs}} \quad (4.17)$$

which means that in order to make the circuit stable, enough resistance must be present at the gate node at frequencies where the input impedance has a negative real part. In practice this can come from a physical resistor or it can be realized through of the losses of the bypass capacitor.

4.3.6.2 Shared-Junction Cascode

To overcome the problems of the cascode device, a shared-junction cascode should be used to improve the mm-wave performance. Shared junction devices are used very commonly in analog circuits to reduce the capacitance of the device. The layout

of such a structure is shown in figure 4.20. The gate of the cascode device needs to be properly biased and should be connected to a low-impedance AC ground. This is done by using a small bypass capacitance to ground, placed as close as possible to the gate of the cascode transistor. Additional low frequency bypass should be provided by a large MOS or MIM capacitor. Fig. 4.21 shows the measured MSG of a $20\mu\text{m}$ shared-junction nMOS cascode, compared to a common-source transistor. It can clearly be seen that below about 40GHz, the cascode structure achieves a higher power gain. However, at 60GHz, the performance of the cascode device is similar to that of the $20\mu\text{m}$ regular multi-finger device. Also note that the cascode device is unconditionally stable above 30GHz ($k > 1$)⁸.

For the cascode models, a custom large signal model has been developed. For a shared-junction cascode, the substrate is shared by the two devices, which is clearly visible in figure 4.22. Figure 4.23 show a comparison between the measurements and the model of a $40\mu\text{m}$ shared-junction cascode device. A good agreement between both is achieved, which validates the modeling approach.

4.3.6.3 A 60GHz Cascode Amplifier with Interstage Matching

The shared-junction cascode still has a non-negligible parasitic capacitance at the interstage node. This capacitance will reduce the gain at higher frequencies, making the cascode less beneficial at mm-wave. This can clearly be seen in figure 4.21. Indeed, at 60GHz, the MSG of the round-table common-source transistor is the same as the MAG of the shared-junction cascode device.

A different approach followed in [23], is to create a cascode structure with two round-table transistors, and to place an interstage matching network between the common-source and the common-gate stage. This circuit combines the excellent mm-wave performance of the round-table layout with the DC current re-use of a cascode topology. This solution also achieves a lower DC power consumption compared to a two-stage cascode amplifier with two round-table common-source transistors. It also achieves more gain for the same DC power consumption when compared to a shared-junction cascode transistor.

Fig. 4.24 shows the simulated MSG and k -factor of a cascode amplifier, composed of two $40\mu\text{m}$ round-table devices, with an interstage matching. At 60 GHz, a gain of 11.5 dB is achieved with a DC power consumption of only 6.7 mW. This figure includes the power loss of the interstage matching network. As a comparison, the two-stage cascaded amplifier described in section 4.3.3 achieves a similar gain of 12dB, but at a higher power consumption of 10.4 mW. Also, this compound device compares favorable to a regular shared-junction cascode device that achieves an MSG of 8.5dB for the same power consumption of 6.7mW. Clearly, the current re-use and interstage matching allows one to achieve more gain for a given DC power consumption. The proposed schematic, using this compound cascode device, is shown in figure 4.25.

⁸ Portions of this section are taken from [23] (© IEEE 2007)

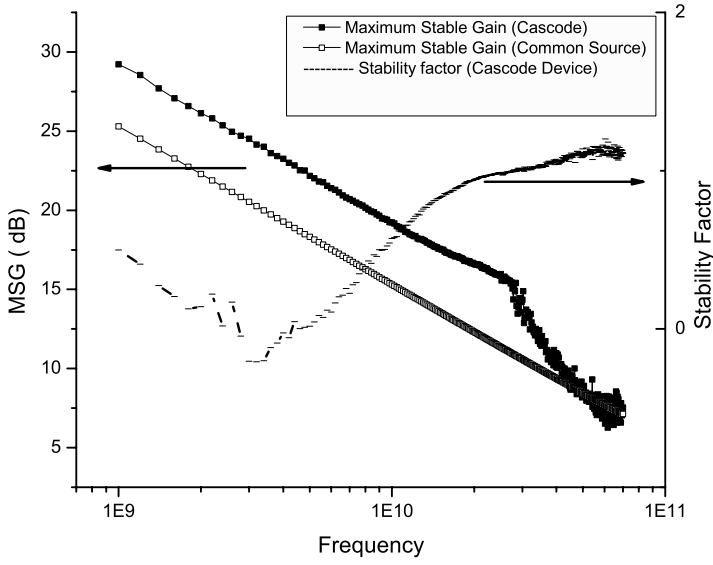


Fig. 4.21 Measured MSG of a $20\mu\text{m}$ shared-junction cascode and a $20\mu\text{m}$ round-table common-source device [23] (© IEEE 2007).

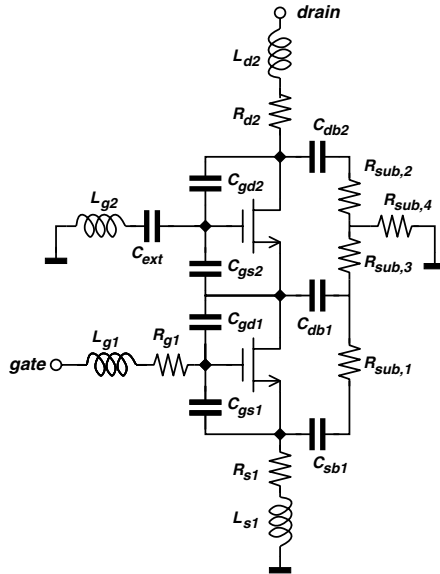


Fig. 4.22 Measurement-based small-signal model of the shared-junction cascode device [23] (© IEEE 2007).

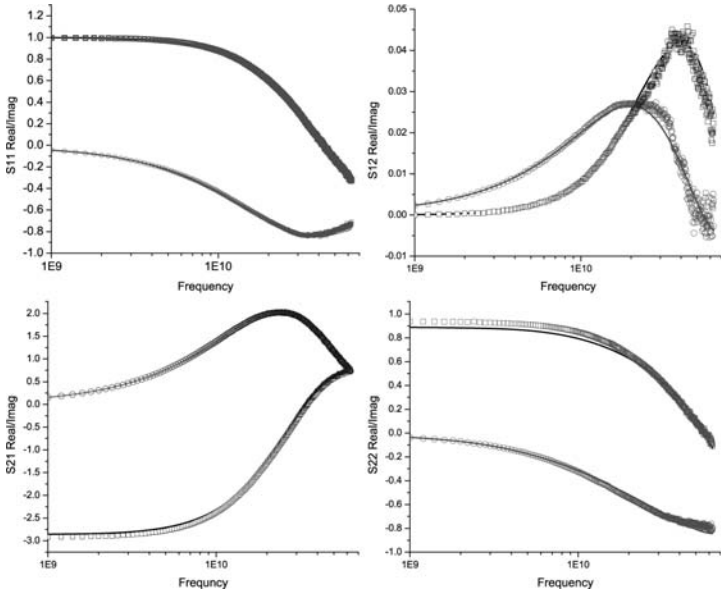


Fig. 4.23 Measured S -parameters versus model of a $40\text{ }\mu\text{m}$ shared-junction cascode device [23] (© IEEE 2007).

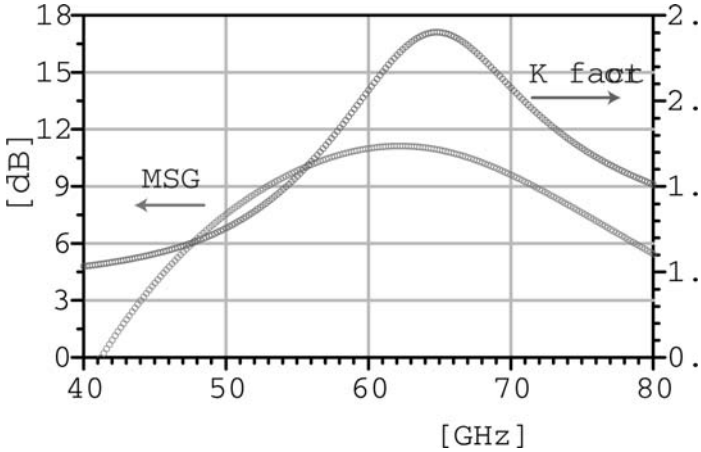


Fig. 4.24 Simulated MSG and k -factor of cascode amplifier [23] (© IEEE 2007).

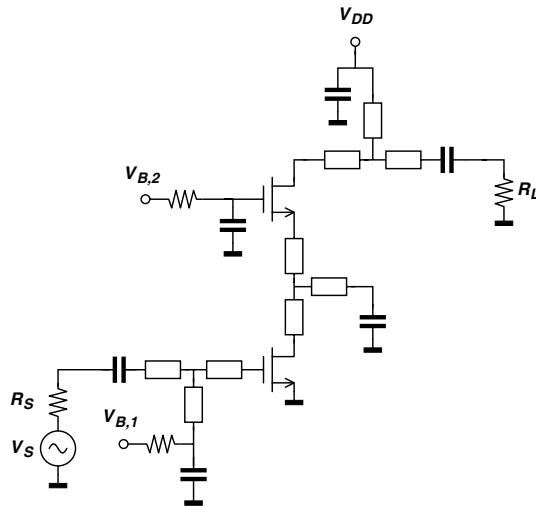


Fig. 4.25 Schematic of the cascode amplifier [23] (© IEEE 2007).

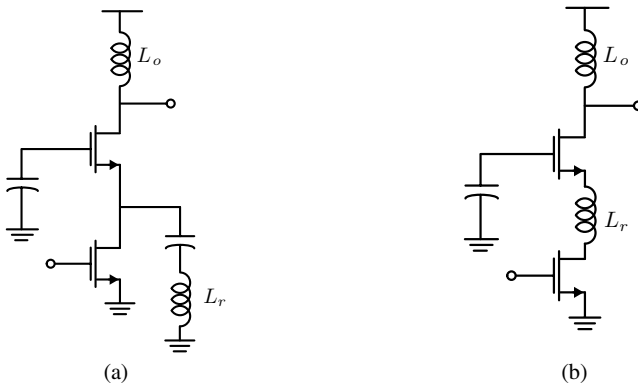


Fig. 4.26 Two techniques to reduce the impact of the cascode shared junction capacitance. Raising this impedance (a) through a resonant LC tank or through a (b) series inductor improves the noise figure of the cascode device.

Similar to previous designs, coplanar transmission lines were used for both input, output and interstage matching.

Another approach is to resonate out the capacitance at the shared junction node, as reported in [24] and shown in Fig. 4.26a. A short section of transmission line or an inductor can be used to realize a low Q resonant circuit around the frequency of interest. An alternative approach is to place a series inductor between the devices, as shown in Fig. 4.26b [25]. This is the preferred way to improve the noise as the technique rejects noise over a broad range of frequencies and is less sensitive to process variations.

4.3.7 A Two-Stage 30 GHz Amplifier

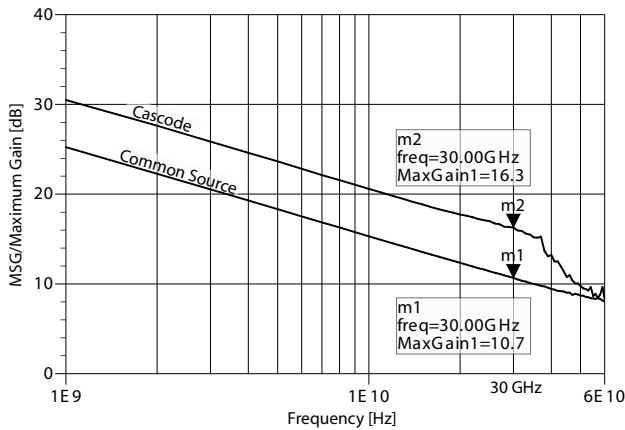


Fig. 4.27 Maximum stable gain for $40\mu\text{m}$ wide common source and cascode transistors biased at $0.2\text{mA}/\mu\text{m}$ [26] (© IEEE 2007).

A two-stage 30 GHz amplifier was designed in a 90nm CMOS process with the goal of obtaining low noise amplification and unconditional stability, allowing the amplifier to be used as a general purpose building block. A good output compression point is also useful for applications such as an LO buffer⁹.

The selection of the active devices is the first critical step. Fig. 4.27 shows maximum stable (MSG) gain plots for $40\mu\text{m}$ wide round table common source transistor and also a $40\mu\text{m}$ wide cascode device both biased at $0.2\text{mA}/\mu\text{m}$ which is a good compromise bias point for both high MSG and low NF_{min} . As depicted in the figure at 30 GHz the CS has a $\text{MSG}=10.7$ dB whereas the cascode has a substantially higher MSG of 16.2 dB. The frequency of operation is below the pole of the cascode device results from a compact shared-junction layout discussed previously.

Similar to previous designs, the amplifier uses CPW transmission lines extensively. Coplanar waveguide structure was used rather than microstrip due to higher achievable Z_0 . To reduce the conductive losses, two top metal layers were strapped together. A $Z_0 = 51\Omega$ was obtained by setting the signal line width to $10\mu\text{m}$ and gap spacing to $7\mu\text{m}$. A typical custom MOM capacitor employed in the design has an area of $25\mu\text{m} \times 25\mu\text{m}$ and 914 fF of capacitance, with a measured resonance frequency of 41 GHz.

As depicted in the Fig. 4.28, the amplifier consists of two stages. Since a single transistor has a better noise performance over a cascode, the first stage is an amplifier with a single round table transistor. Due to the large gain of the first stage, the noise of the second stage is only of marginal concern and thus a cascode amplifier was placed in the second stage to improve the gain and isolation of the overall amplifier.

⁹ Portions of this section are taken from [26] (© IEEE 2007)

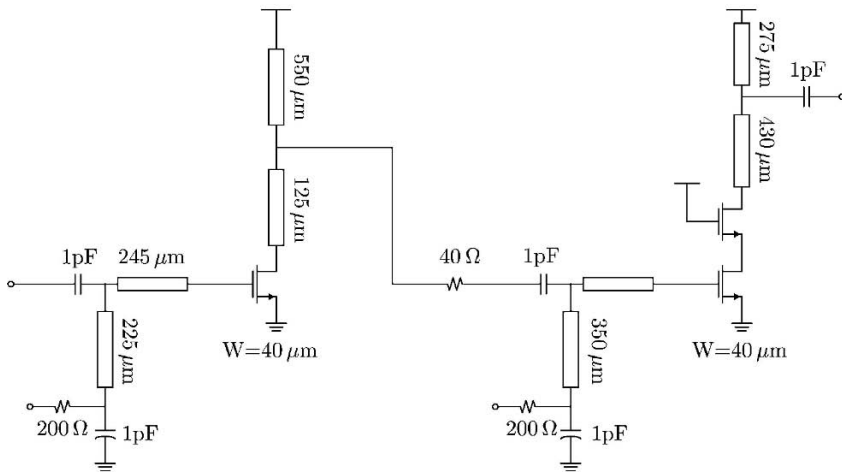


Fig. 4.28 Schematic of the two-stage 30 GHz LNA [26] (© IEEE 2007).

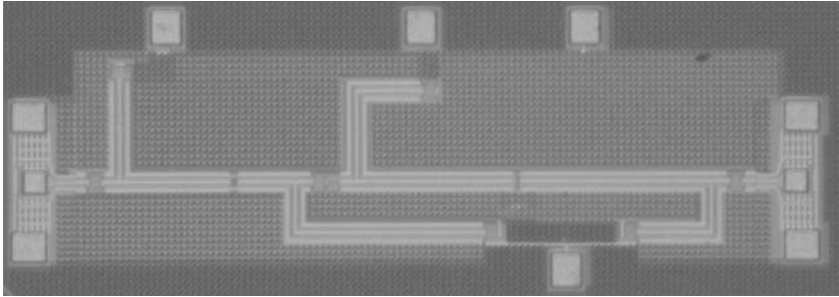


Fig. 4.29 Two-stage 30 GHz amplifier die photo implemented in a 90nm CMOS process [26] (© IEEE 2007).

At 30 GHz the cascode device is not unconditionally stable and therefore to realize a robust and stable amplifier a $40\ \Omega$ stabilizing resistor was used in the signal path between the two stages. This resistor lowers the gain but has a little effect on the noise since the first stage has a 12 dB of voltage gain. Due to non-negligible gain at an unwanted band around 5 GHz, $200\ \Omega$ resistors were used in the gate bias lines to de-Q the matching network and dampen the gain at 5 GHz.

The fabricated prototype, shown in Fig. 4.29, has been characterized. Measurements were taken directly using wafer probes. The measured S -parameters are shown in Fig. 4.30, displaying good match between measurements and simulation. The amplifier has a peak gain of 20 dB at 28.5 GHz and 3-dB bandwidth from 27.6 GHz to 30.2 GHz. The amplifier is well matched, with input and output return loss better than 15 dB and 20 dB respectively over the 3 dB bandwidth. The large signal and

distortion measurements verify +2 dBm of output power at the 1-dB compression point (input $P_{1dB} = -17$ dBm) and -7.5 dBm of IIP3.

Noise measurements were performed using the gain method. The input of the amplifier is matched and the output noise power is measured on a spectrum analyzer. The spectrum analyzer noise floor was too high (-140 dBm/Hz) for a direct measurement, so a second amplifier module with 28 dB of gain at 30 GHz was used to raise the output noise higher than the spectrum analyzer noise floor (around -127.5 dBm). The noise figure is calculated at a few points across the band and we find the minimum noise figure of 2.9 dB at 28 GHz (where the gain is close to its peak) and the NF is lower than 4.2 dB over the bandwidth of the amplifier (Fig. 4.31a). The amplifier consumes 16.25 mW of power from a 1V supply voltage. The current is nearly evenly divided between the stages. Linearity is satisfactory for a 1 V design (Fig. 4.31b).

4.4 Mixers and Frequency Translation

Mixers are key components of an RF front-end, translating the IF signal to a carrier frequency for transmission and from an RF carrier back down to IF for detection. Mixers can be categorized into two families, active and passive mixers. Active mixers have power gain whereas passive mixers have insertion loss. Each mixer essentially modulates either the transconductance of an amplifier or the resistance of a switch to produce the mixing action through time-varying mechanism. The devices are usually modulated with a large LO signal to maximize the conversion gain. At mm-wave frequencies it is very difficult to get a large LO power out of a silicon device, especially with CMOS technology. Even though the device non-linearity will produce mixing products even with a weak LO, the conversion gain is too small when mixers are operated in such a fashion. Hence, the conversion gain and noise figure requirements must be obtained at a reasonable LO power level (~ 0 dBm). Also the design goal of low-noise, together with mm-wave modeling difficulties, limits the use of complex mixer topologies.

4.4.1 Single Transistor Mixers

Single-gate GaAs FET mixers with good conversion gain and noise figure have been used successfully for mm-wave applications [28]. The use of standard CMOS for single-gate mixers at mm-wave range has been explored in [27] where we reported a 60 GHz quadrature-balanced single-gate mixer implemented in a main stream CMOS technology.¹⁰

¹⁰ Portions of this section are taken from [27] (© IEEE 2005)

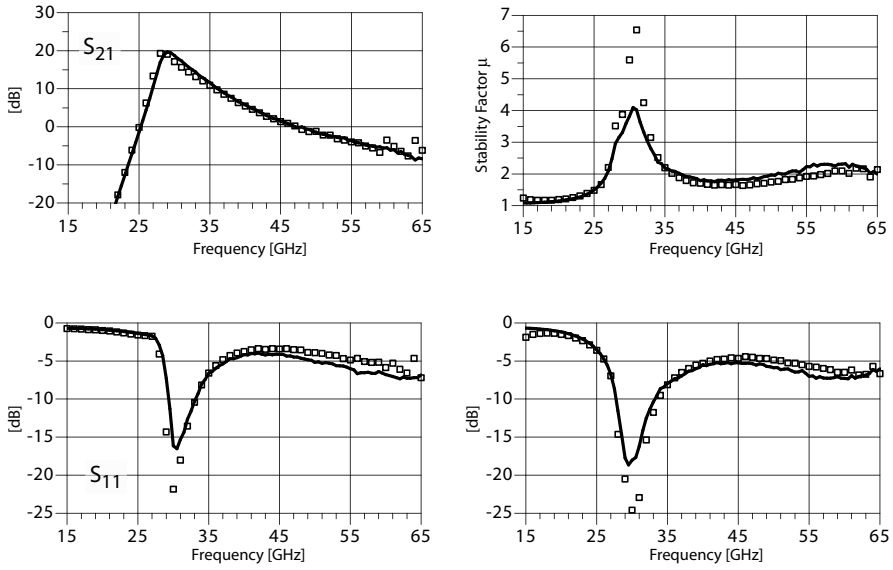


Fig. 4.30 Measured and simulated S -parameters and stability factor μ of two-stage 30 GHz amplifier [26] (© IEEE 2007).

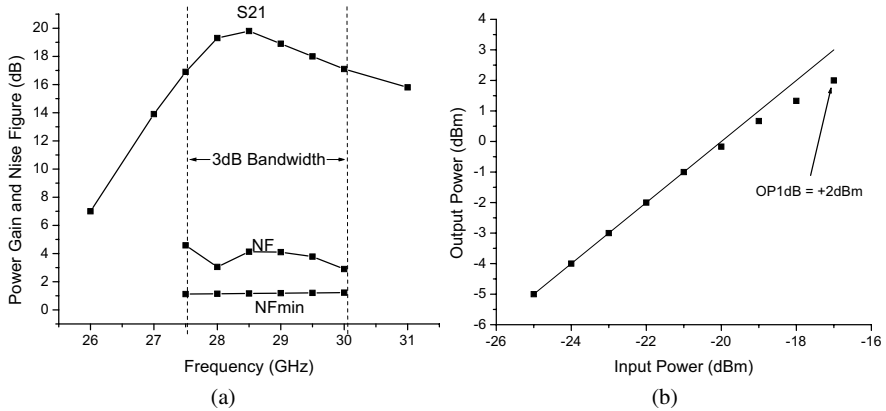


Fig. 4.31 Measurements of two stage 30 GHz amplifier (a) noise figure and (b) compression characteristics [26] (© IEEE 2007).

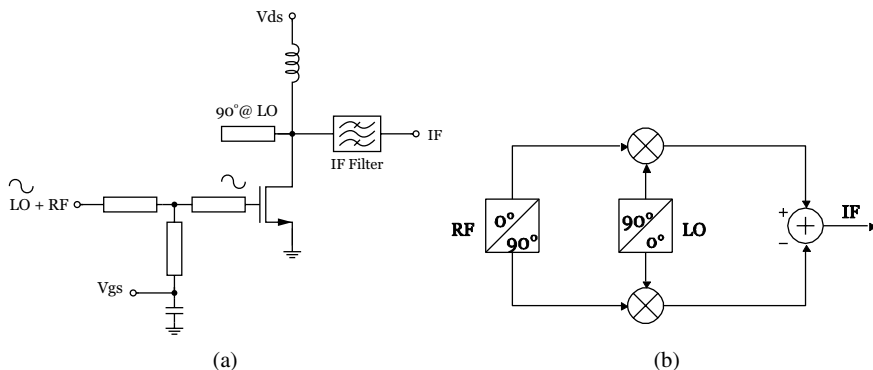


Fig. 4.32 (a) Simplified single-gate mixer. (b) Quadrature balanced architecture [27] (© IEEE 2005).

A simple single-gate mixer architecture is shown in Fig. 4.32a. It is a transconductance mixer, as the time-varying $g_m(t) = g_m(t + T_{LO})$ of the common-source stage is the main source of frequency conversion. The LO signal, applied at the gate of the transistor together with the RF signal, varies the g_m around the dc bias point, and therefore modulates the RF gain of the common-source stage and consequently provides frequency conversion.

For a standard CMOS process, as shown in Fig. 4.33, the region close to the threshold voltage exhibits a steep change of transconductance vs. the gate-source voltage. The large LO signal can efficiently modulate the g_m of a transistor biased in this region. Moreover, by choosing the bias point close to V_T , minimal dc power consumption is achievable. Using only one common-source NMOS transistor promises a good noise figure at mm-wave frequencies. Also, because the mixer is effectively a common-source device, it is very easy to model with a simple transistor structure.

Single-gate mixers have one major practical implementation problem: they require a hybrid or elaborate power combining circuit to combine the LO and RF signals. Typically the hybrids are bulky, and their insertion-loss adds directly to the mixer's noise figure. Fortunately, due to the high frequency of operation at 60 GHz, the hybrid can be easily integrated on-chip. Also, by using a balanced architecture, as shown in Fig. 4.32b, better spurious response and LO noise rejection can be obtained [29].

The quadrature balanced mixer that consists of two unit single-gate mixers, and a 90° branch-line hybrid is shown in Fig. 4.34. The mixer was designed to down-convert from the nominal RF frequency of 60 GHz to a low-gigahertz IF of 2 GHz using a 58 GHz, 0-dBm LO. In this particular CMOS process, choosing a gate source voltage, V_{GS} , slightly above the threshold voltage, V_T , of the $80 \times 1 \mu\text{m}/0.13 \mu\text{m}$ device maximizes the fundamental frequency component of g_m . The nonlinear device model was then used in ADS harmonic balance simulations to estimate the optimum large-signal matching impedances for the gate at 60 GHz for this bias point. By taking advantage of the intrinsic device capacitances, it is possible to provide the

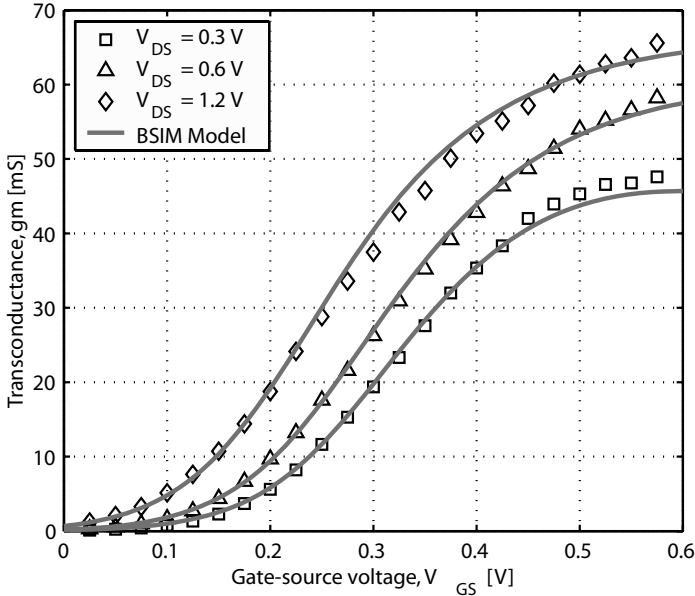


Fig. 4.33 Measured and modeled g_m versus V_{GS} [27] (© IEEE 2005).

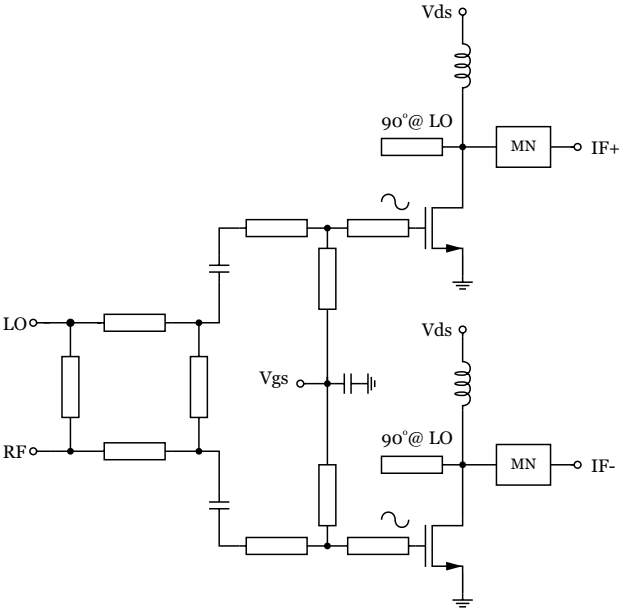


Fig. 4.34 Simplified circuit diagram of the single-gate quadrature balanced mixer [27] (© IEEE 2005).

required 90° phase shift by using transmission lines shorter than $\lambda/8$. This significantly reduces the insertion-loss and the physical size of the hybrid. The IF matching network consists of on-chip lumped LC components. At 2 GHz, a quality factor of 7 was measured for the main spiral inductor.

Simulations accounting for parasitics indicate that the core of the NMOS single-gate mixer is only conditionally stable up to 60 GHz due to the large Miller capacitance. Short-circuiting the drain node at mm-wave frequencies breaks the feedback loop (through C_{gd}) and improves the stability. Using an open quarter-wave stub is a popular and practical technique. Moreover, suppressing the LO at the drain minimizes the variation of drain-source voltage V_{DS} , and consequently minimizes the performance degradation due to other nonlinear elements such as C_{db} and g_{ds} .

The nonlinear device model described in Section 3.2.4 along with the transmission line electrical models were implemented in ADS environment. The harmonic balance and small-signal simulators along with an optimizer were used to fine tune the transmission line lengths and optimize the performance of the mixer. From the simulations, a SSB noise figure of 11.5 dB is expected for the mixer (RF=60 GHz, IF=2 GHz, PLO=0 dBm).

The mixer was fabricated using a 6-metal layer, 130-nm digital CMOS process, and the chip area is $1.6 \times 1.7 \text{ mm}^2$, including pads. The die photo of the fabricated mixer is shown in Fig. 4.35. All the measurements were taken on wafer by using GSG probes on a Cascade Summit probe station. Fig. 4.36a shows the conversion gain of the mixer for different LO power levels. The RF frequency is 60 GHz, the IF frequency is 2 GHz, and the RF input power is -25 dBm. Conversion loss of better than 2 dB was achievable for a low LO power of 0 dBm.

The RF and LO frequencies were also varied to obtain the frequency characteristics of the mixer for fixed IF frequency of 2 GHz and LO power of 0 dBm. As shown in Fig. 4.36b, the 3-dB RF bandwidth of the mixer is more than 6 GHz. Fig. 4.37a shows the small-signal input return loss of the mixer. The center frequency is slightly below the targeted frequency, but there is still better than 15-dB return loss at 60 GHz. The LO-RF leakage performance, shown in Fig. 4.37b, is adequate for typical mm-wave receiver systems as additional isolation comes from the LO buffer and multi-stage LNA. The measured input-referred 1-dB compression point is -3.5 dBm. Power consumption is only 2 mA from a 1.2-V supply.

4.4.2 Dual Gate Mixers

The “dual-gate” mixer, shown in Fig. 4.38, consists of a cascode transistor driven by the RF signal at the input stage and the LO signal at the cascode stage. Since the LO and RF ports do not need to be combined, this mixer is more compact than a FET mixer discussed above. The transconductance of the input device is modulated by the LO signal, which alters the drain to source voltage. The LO bias effectively modulates the S_{21} of the cascode device, producing a time-varying transconductance stage. The input stage is biased in saturation but near the triode region, where the sensitivity

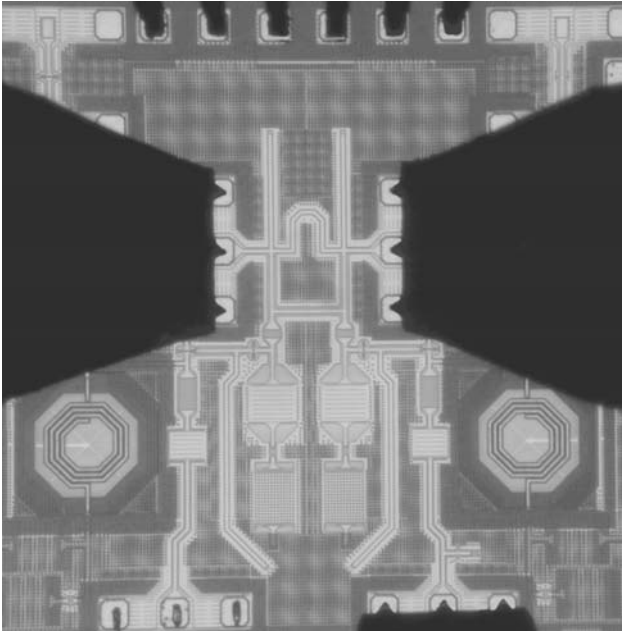


Fig. 4.35 Chip photo of single-gate quadrature balanced mixer [27] (© IEEE 2005).

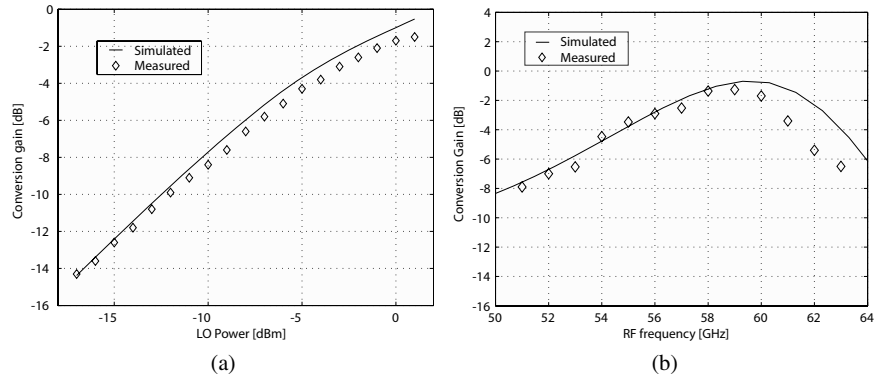


Fig. 4.36 Single gate quadrature balance mixer results. (a) Mixer conversion gain vs. LO power (RF = 60 GHz, IF = 2 GHz, VGS = 200 mV, PRF = -25 dBm). (b) Conversion gain vs. RF frequency (IF = 2 GHz, VGS = 200 mV, PRF = -25 dBm, PLO = 0 dBm) [27] (© IEEE 2005).

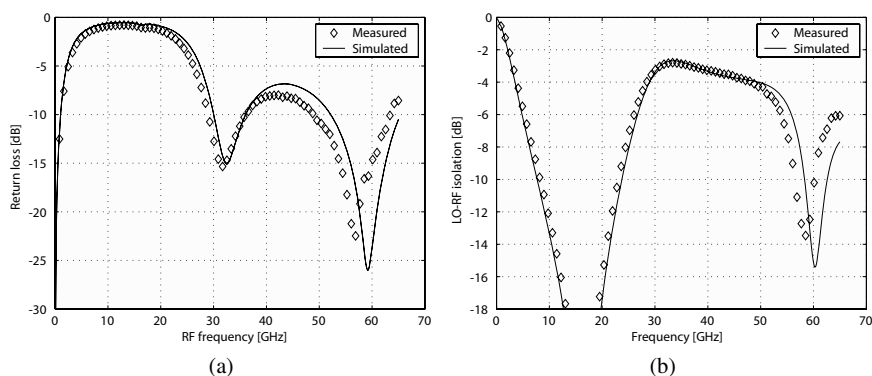


Fig. 4.37 Single-gate quadrature balanced mixer measurement results. (a) Input return-loss frequency characteristics. (b) LO-RF isolation frequency characteristics [27] (© IEEE 2005).

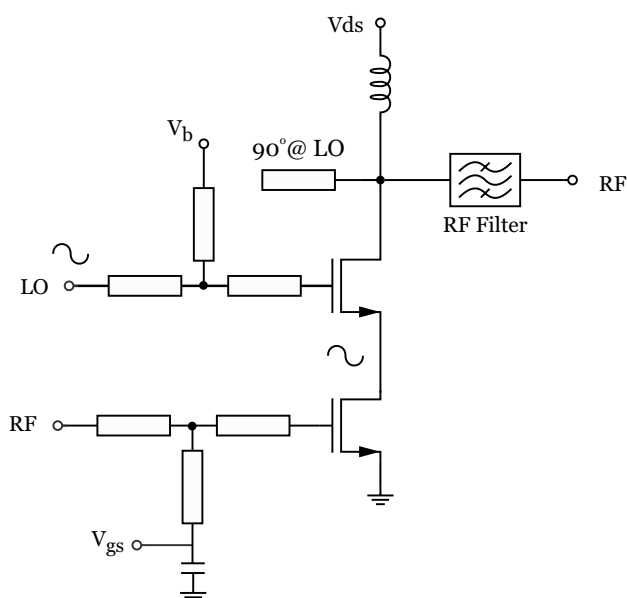


Fig. 4.38 A "dual gate" mixer is implemented in standard CMOS technology with a cascode configuration.

of the transconductance is maximized. This also has a side benefit of reducing the power consumption of the mixer. The RF and LO ports are matched to 50Ω through co-planar transmission lines. The LO signal is matched at 40 GHz and the RF is matched at 60 GHz. The DC bias stub shorts the LO leaking into the RF, and an LO stub likewise shorts the RF signals at the LO port. This results in good isolation and improved stability. An IF filter is designed at 20 GHz for further processing of the signal. Since the LO and RF signal are widely separated, the image signal at 80 GHz is of little concern, and it is naturally filtered by the LNA and mixer passband response.

A layout of the mixer in 130nm technology appears in Fig. 4.39. The chip area is $0.8\text{ mm} \times 1.5\text{ mm}$. The measured performance has better than 3 dB conversion loss for an LO power of 0 dBm. The input referred 1-dB compression point is -4 dBm, with a return loss better than 15 dB at 60 GHz. Simulations indicate a noise figure of 12.5 dB. In summary, the dual gate mixer provides high LO-RF isolation, conversion gain, and obviates the need for a combiner. Even though the common source transistor is operated in the linear region to maximize the sensitivity of the transconductance to drain voltage, the cascode transistor acts as an IF amplifier. The stability of the design is carefully controlled by providing a reactive termination at the gate of the common gate transistor and through careful device modeling.

4.4.3 Gilbert Cell Mixers

Differential Gilbert cell type mixers are commonly used at RF frequencies but have several problems at mm-wave frequencies. The use of differential circuits is more risky because the core transistors are used as full three-port devices, including the influence of the body-effect due to the non-grounded source. Since most devices are characterized as two-ports with a common ground, one must rely more on the model rather than the measurements to employ devices in this configuration.

Consider a simple Gilbert-cell type single-balanced structure using a single-ended RF input to drive the G_m -stage and a differential switching stage for frequency conversion, shown schematically in Fig. 4.40. For example, the RF drive is at 60 GHz and LO modulation is at 58 GHz. The down converted 2 GHz IF signal is coupled through an output balun and finally a source follower is used to drive the 50Ω output. The differential side of the output balun also serves as a high impedance load for the mixer. In a fully integrated mixer, the output balun can be omitted if a fully-balanced IF stage VGA is employed.

The mixer uses a resonant tuning network at the drain of the transconductance stage to resonate out all the parasitic at that node. At high frequencies the parasitics at that node will shunt the signal to ground resulting in signal attenuation. By adding the tank, one can recover a significant amount of conversion gain.

A single-turn coupled inductor structure is used for realizing the high frequency balun. Conventionally, at microwave frequencies “rat-race” couplers or distributed structures are used for single-ended to differential conversion. While transmission

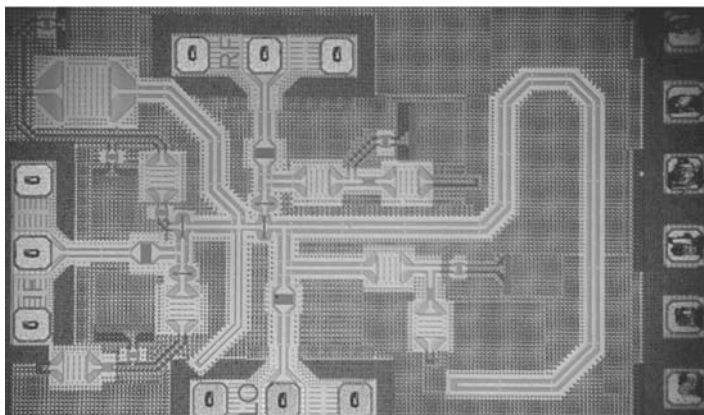


Fig. 4.39 Layout of 130nm CMOS “dual-gate” mixer.

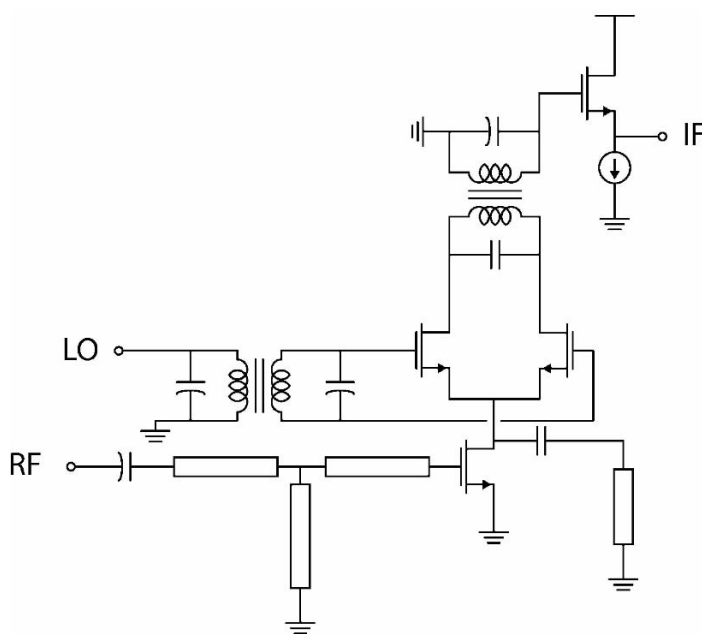


Fig. 4.40 Schematic of a single-balanced Gilbert cell mixer.

line couplers are very effective and offer good phase and amplitude balance, they are bulky (see Fig. 4.35). For example a “rat-race” balun needs arms spanning a quarter wave length. At 60 GHz in the Si substrate/oxide, this length is about $600\mu m$, resulting in very large balun and relatively high insertion loss. A loop inductor based

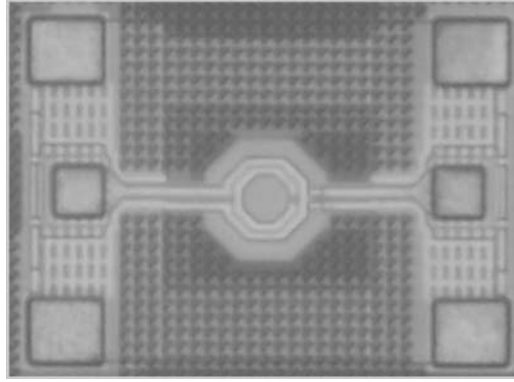


Fig. 4.41 Layout of a coupled ring transformer (die photo).

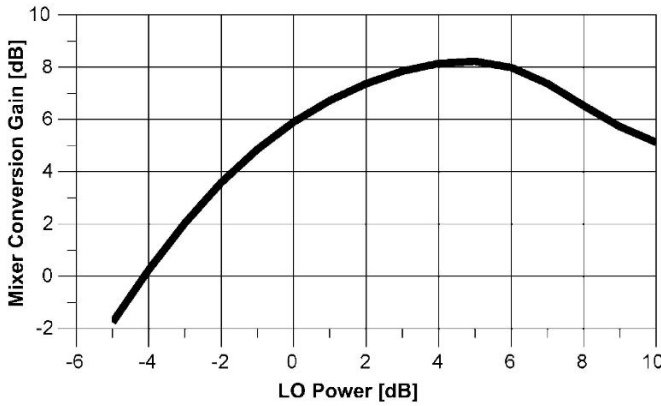


Fig. 4.42 Conversion gain of Gilbert cell mixer versus LO power (dBm).

balun uses arms of length $24\mu\text{m}$ and is realized in the area of the order of $80 \times 80\mu\text{m}^2$ and shown in Fig. 4.41.

In order to optimize and design the balun, the minimum insertion loss, related to the quality factor and coupling factor of the windings. For a bi-conjugate match, the transformer minimum insertion loss is given by [30]

$$IL_{min}(Q, k) = 1 + \frac{2}{Q^2 k^2} - 2\sqrt{\frac{1}{Q^4 k^4} + \frac{1}{Q^2 k^2}} \quad (4.18)$$

Using this metric an optimal structure with a single turn and appropriate metal width and spacing has been designed. The tranformer balun has an insertion loss of 1 dB while providing a match at the LO port, maximizing the LO swing.

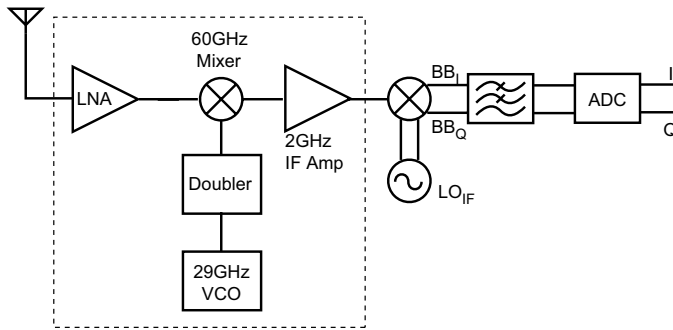


Fig. 4.43 Block diagram of a 60 GHz front-end implemented (red) in 130nm CMOS [31] (© IEEE 2007).

The simulated performance of such a mixer has 7 dB conversion gain, a linearity of $iIP3 = -5$ dBm, and a power consumption of 7 mW. The LO port needs to be driven with an LO power of 2 dBm to saturate the conversion gain. By increasing the LO power, the conversion gain increases up to a point and then drops, as shown in Fig. 4.42. In practice an LO power beyond 4 dBm is impractical and most likely the LO power will be around 0 dBm. The drop in gain is due to the change in operating point of the transconductance device due to operation in the triode region.

4.5 Examples of Integrated Front-Ends

4.5.1 CMOS 130nm 60 GHz Front-End

A highly-integrated 60 GHz CMOS front-end receiver consisting of an LNA, a quadrature balanced down-converting mixer, a 30 GHz VCO, and a frequency doubler, shown in Fig. 4.43. Individual circuit blocks were also fabricated to characterize their performance separately. The front-end is fabricated in a 130-nm standard digital CMOS technology, which has a substrate resistivity of $10 \Omega\text{-cm}$ and 6 layers of metalization. Transit frequencies (f_T) and maximum frequencies of oscillation (f_{max}) of 85 GHz and 135 GHz, respectively, have been reported for NMOS devices in this process [17].¹¹

The LNA topology consists of three stages of cascode devices with input, output, and interstage reactive matching, as shown in Fig. 4.44. Cascode transistors, used in order to reduce the Miller capacitance and improve stability, are biased at $150 \mu\text{A}/\mu\text{m}$, and the MAG is 6.0 dB at 60 GHz per stage. CPW transmission lines are used extensively in the design for impedance matching, interconnect wiring, and the bias networks. All lines are kept as short as possible to minimize losses. Since the

¹¹ Portions of this section are taken from [31] (© IEEE 2007)

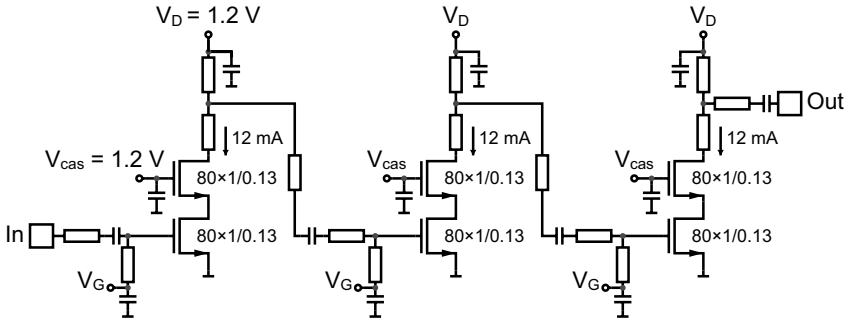


Fig. 4.44 Simplified schematic of the 60 GHz 3-stage amplifier using CPW transmission lines [31] (© IEEE 2007).

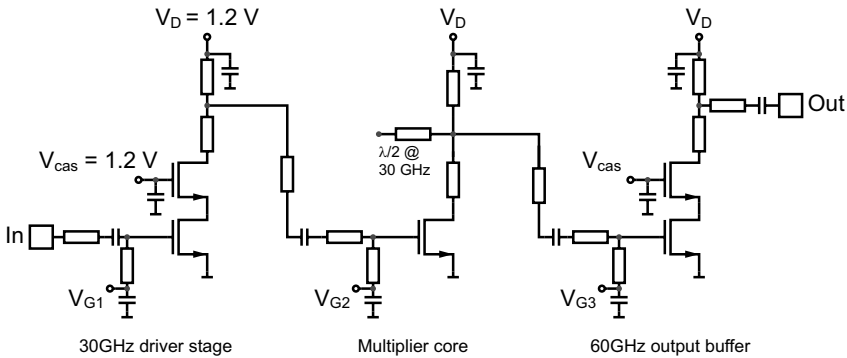


Fig. 4.45 Simplified circuit diagram of the frequency doubler [31] (© IEEE 2007).

integrated LNA output does not need to drive a pad or be matched to an off-chip 50Ω load, it is matched directly to the input of the quadrature mixer, providing higher performance than the standalone amplifier reported in [17].

The quadrature balanced down-converting mixer described in Sec. 4.4.1 consists of two unit single-gate mixers, and a 90° branch-line hybrid. The mixer was designed to down-convert from the nominal RF frequency of 60 GHz to a low-gigahertz IF of 2 GHz using a 58 GHz, 0-dBm LO. The simplified schematic of the 3-stage frequency doubler is shown in Fig. 4.45. It consists of a 30 GHz input driver, the core multiplier, and the 60 GHz LO buffer. Cascode devices with input, output, and interstage reactive matching were used to implement the input driver and output buffer stages. The second stage is the frequency doubling stage. A common-source NMOS transistor is biased close to the threshold voltage to efficiently generate a second harmonic component. Therefore, the doubler can be considered as a half-wave rectifier. A combined matching network and fundamental frequency rejection filter at the drain of this transistor supports the generation of the 60 GHz component and rejects the 30 GHz input component. A standalone version of the doubler, matched to the off-

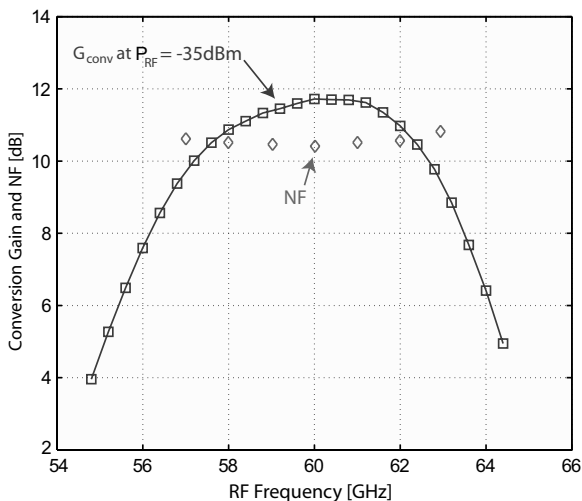


Fig. 4.46 Measured conversion-gain and NF of the integrated front-end [31] (© IEEE 2007).

chip 50Ω input (at 30 GHz) and output (at 60 GHz) with pads, is implemented and characterized separately. A -6 -dBm, 29 GHz input signal is applied to the doubler and a conversion gain of the multiplier of 7.2 dB is measured at 58 GHz. The 3-dB bandwidth at this input power level is from 53 GHz to 62.5 GHz. The fundamental frequency at the output of the doubler is >35 dB down. The doubler alone consumes 2 mA while the input and output stages consume 8 mA and 12 mA respectively.

A 28.4-29.4 GHz Pierce VCO is designed to generate the input signal to the frequency doubler. The stand-alone implementation of the VCO provides -3 dBm output power to a 50Ω load at 29 GHz. The measured phase noise is around -93 dBc/Hz at 1 MHz offset. The integrated version of the VCO with the doubler is fabricated separately. The integrated VCO/doubler block provides 2 dBm output power at 58 GHz and the measured phase noise is -86 dBc/Hz at 1MHz offset.

Fig. 1.6 depicts the die photo of the integrated front end. The chip area is about 7 mm^2 including pads. On-wafer measurements were performed using a Cascade Microtech probe station. An Anritsu 37397C VNA was used for S -parameter measurements. The input return loss at the RF port is better than 15 dB. Fig. 4.46 shows the conversion characteristics of the front-end with conversion gain of 11.8 dB at 60 GHz. The RF and LO frequencies were varied to obtain the frequency characteristics of the mixer for fixed IF frequency of 2 GHz. To obtain a broadband frequency characteristics, another fabricated version of the front-end with an external 30 GHz LO was used. The 3 dB RF bandwidth of the mixer is about 6 GHz. The measured input-referred 1-dB compression point of the front-end is -15.8 dBm at 60 GHz. Noise figure (NF) measurement of the integrated front-end was performed using a Millitech WR-15 noise source, WR-15 waveguide probes, and an Agilent N8973A NF measurement system. The measured NF of the downconverter is shown in Fig. 4.46. The

NF is 10.4 dB at 60 GHz. The total power dissipation of the integrated front-end is 64 mA from a 1.2 V supply.

4.5.2 *SiGe Transceiver Chipset*

A 60 GHz transceiver [14] has been realized in SiGe BiCMOS-8HP (0.13- μm) technology. A block diagram of the two-chip transceiver chipset is shown in Fig. 4.47. A dual-conversion superheterodyne radio architecture is selected over a homodyne approach due to its lower carrier feed-through in the transmitter and better IQ quadrature accuracy. The variable intermediate frequency (IF) frequency plan for the radio is based on a single 16 to 18 GHz local oscillator (LO) which is tripled to generate a 48 to 54 GHz LO for the mm-wave mixers or divided by two to provide a 8 to 9 GHz LO for the IF mixers. With an 8 to 9 GHz IF, the image frequency for the superheterodyne is 16 to 18 GHz below the RF frequency, which is sufficiently low enough to allow for low-Q on-chip image filters. Finally, the 16-18 GHz LO frequency results in a simpler frequency synthesizer design, since no high-frequency dynamic dividers are required. Also, the tuning-range and phase-noise requirements can be more easily met in a voltage controlled oscillator design at 18 GHz than at 54 GHz.

The receiver (Rx) chip begins with the image-reject low-noise amplifier from section 4.3.2, followed by a single-balanced Gilbert-cell mixer. This mixer provides about 9 dB of conversion gain with a cascaded upper single-sideband NF of 13 dB at 25°C (mixer + following stages). Following the mm-wave mixer is a variable-gain IF amplifier (IFVGA), which provides between -10 and +10 dB of gain. IF filtering is provided by tuned R-L-C loads at the mixer output and the IFVGA output. The IF signal is downconverted to baseband frequency through a pair of quadrature double-balanced Gilbert-cell mixers with unity gain. Following the IF mixers are unity-gain baseband output buffers to drive external 100 Ω differential loads. A harmonic frequency tripler generates the mm-wave LO signal at 48-54 GHz, while a standard emitter-coupled logic divide-by-two circuit provides the 8-9 GHz LO signal. Finally, the PLL is a type-II fourth-order PLL operating from a 520 to 610 MHz reference clock, and generating a 16.6 to 19.5 GHz output signal.

A die photograph of the Rx is shown in Fig. 4.48. The die size is $3.4 \times 1.7 \text{ mm}^2$ to the outside of the pad frame. The LNA is at the lower left, and the spiral inductors in the Rx mixer and IFVGA are visible to the right of the LNA. The frequency tripler is in the center, and the PLL occupies the right third of the chip. The chip contains > 300 NPNs, > 1000 FETs, and > 90 transmission lines and inductors.

On-wafer measurements were made on the full Rx with PLL. As shown in Fig. 4.49, at 25°C, the Rx power conversion gain is 38 to 40 dB and the NF is 5 to 6.7 dB. The image rejection is 30 to 40 dB, IIP3 is -30 dBm, and input P_{1dB} is -36 dBm, all shown in Fig. 4.50. The Rx consumes 195 mA from 2.7 V, 50 mA of which is in the baseband output buffers. At 85°C, NF remains below 8 dB while gain drops 5 to 6 dB relative to 25°C. The Rx and Tx PLL measurements show a VCO

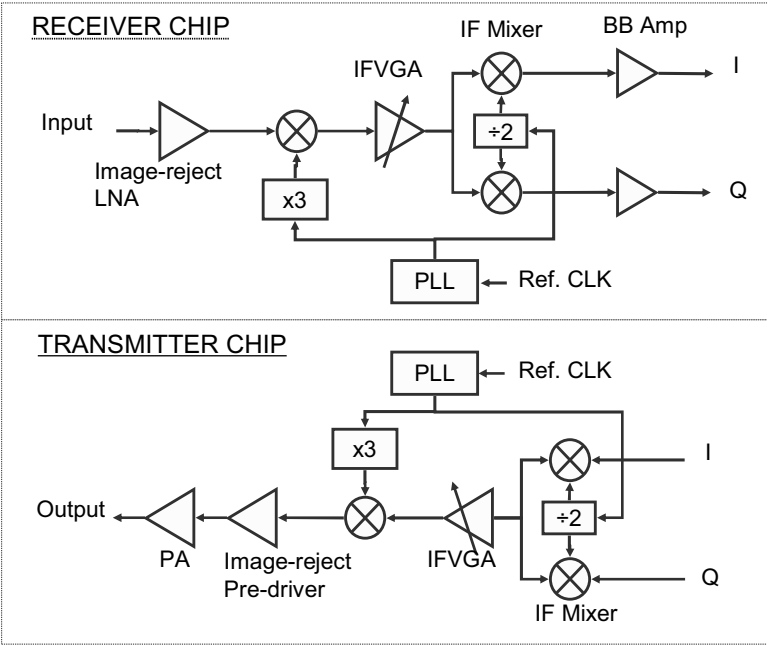


Fig. 4.47 Block diagram of 60 GHz transceiver chipset implemented in 0.13- μm SiGe BiCMOS technology [14] (© IEEE 2006). Dual-conversion superheterodyne architectures are used in both the receiver and transmitter.

phase noise of -115 to -120 dBc/Hz at 10-MHz offset with an RMS jitter less than 1.5° integrated over 0.1-1GHz. Additional receiver measurements are summarized in Table 4.2.

Table 4.2 Summary of measured receiver performance

Metric	Measured Value
Frequency Range	59-64 GHz
Gain	38-40 dB
Noise Figure	5-6.7 dB
S11, RF in	-15 dB
Image Rejection	> 30 dB
P_{1dB} (in)	-36 dBm
IIP3	-30 dBm
Phase Noise (1MHz), tripled	-85 to -90 dBc/Hz < -130 dBc/Hz floor
I/Q Balance	0-4 degrees, <1 dB
Power Dissipation	195 mA, 2.7 V

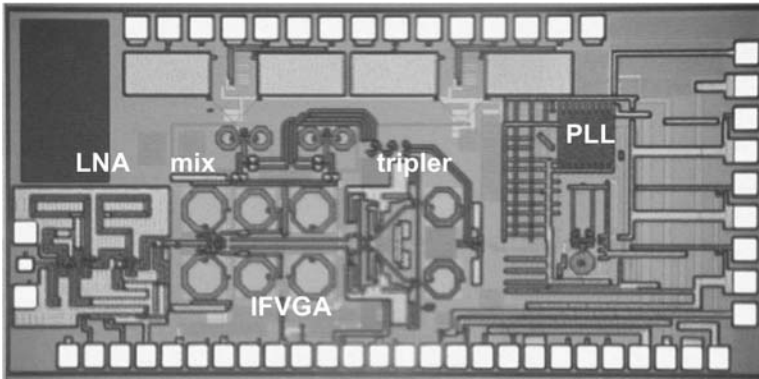


Fig. 4.48 Die photograph of 60 GHz receiver. The die size is $3.4 \times 1.7 \text{ mm}^2$ [14] (© IEEE 2006).

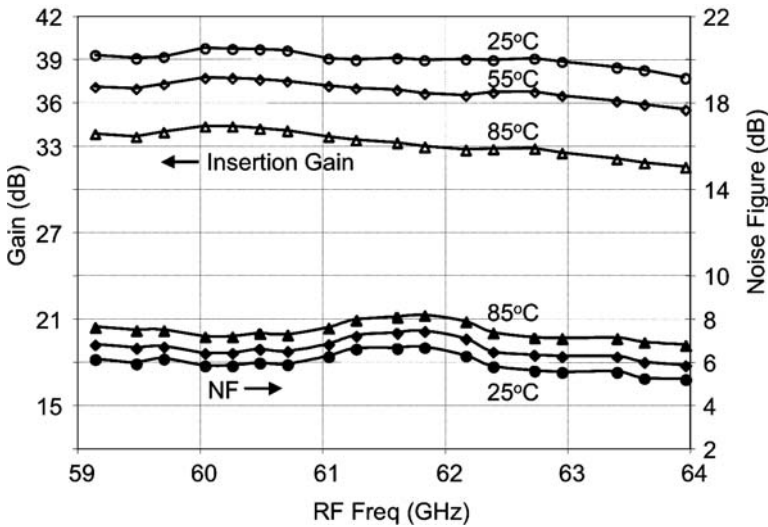


Fig. 4.49 Measured conversion gain and noise figure over temperature of 60 GHz receiver [14] (© IEEE 2006).

Referring to the block diagram in Fig. 4.47, the transmitter (Tx) chip includes a PA, an image-reject driver, an IF-to-RF upmixer, an IF amplifier strip, a quadrature baseband-to-IF mixer, a PLL, and a frequency tripler. The PA provides 15 dB of gain with 10.5-dBm P_{1dB} and 17-dBm P_{sat} . The image-reject driver provides 8-12 dB of gain, 6-8 dBm output P_{1dB} , and >25 dB image rejection. Finally, the IF and LO paths in the transmitter are the reverse of that in the receiver.

A die photograph of the Tx is shown in Fig. 4.51. The die size is $4.0 \times 1.6 \text{ mm}^2$. The PA and the differential output pads are on the left, adjoined to the right by the

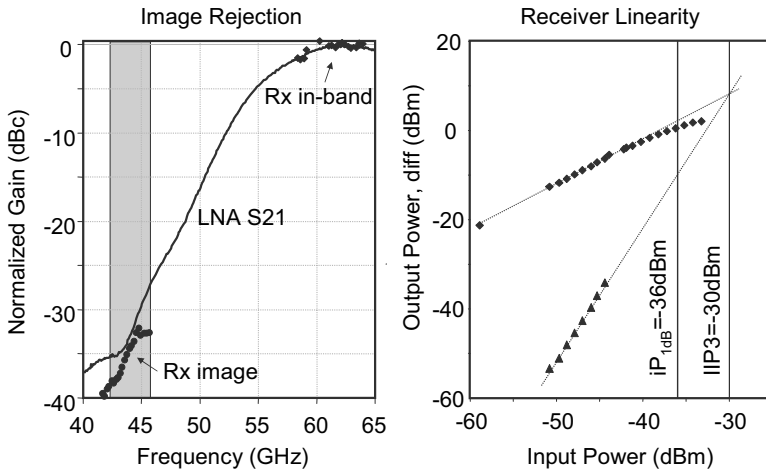


Fig. 4.50 Measured image rejection and linearity of 60 GHz receiver [14] (© IEEE 2006).

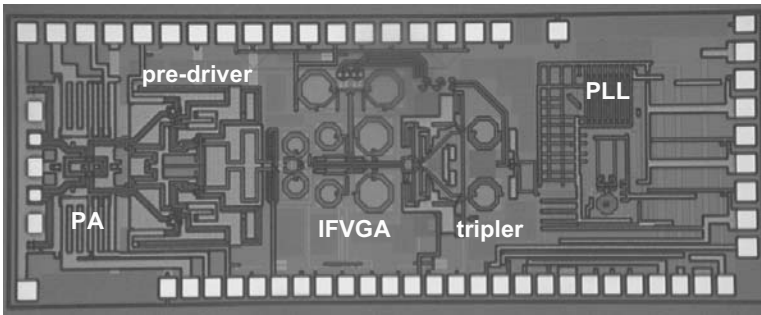


Fig. 4.51 Die photograph of 60 GHz transmitter. The die size is 4 x 1.6 mm² [14] (© IEEE 2006).

pre-driver, IF-to-RF mixer, IFVGA, and frequency tripler. The PLL occupies the right third of the chip, and the baseband-to-IF mixer contains the two spiral inductors at the top center. The Tx chip contains > 300 NPNs, > 1000 FETs, and > 170 transmission lines and inductors.

On-wafer measurements were performed on the full transmitter with PLL. The Tx conversion gain is 42-36 dB. Fig. 4.52 shows the measured output power and conversion gain versus I/Q input power of a 100-MHz CW tone for three different bands (59, 61.5, and 64 GHz). $P_{1dB(out)}$ is 10 to 12 dBm, while P_{sat} is 16 to 17 dBm, and the conversion gain is 34 to 37 dB. Across the temperature range from 5 to 85°C, P_{1dB} remains constant while gain drops 7 dB, as shown in Fig. 4.53. The spurious response of the Tx has been measured in 500-MHz steps across the band. At -25 dBm input power, which is near the P_{1dB} , 20 to 30 dB of image suppression, 20 to 25 dB of carrier suppression, and 20 to 25 dB of sideband suppression are observed. The

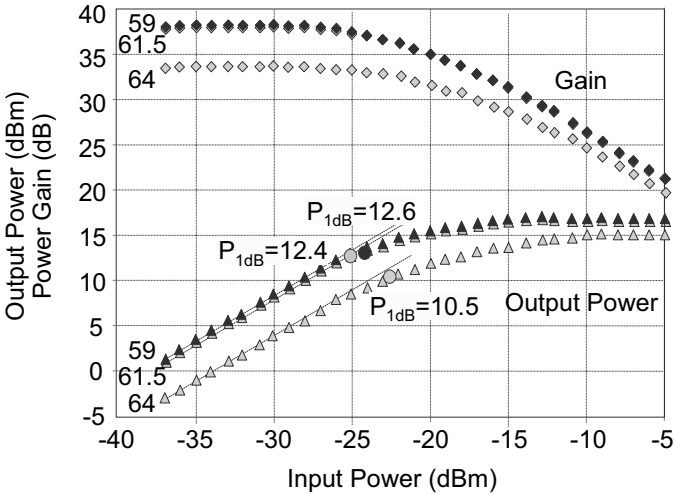


Fig. 4.52 Measured power gain and output power of transmitter [14] (© IEEE 2006).

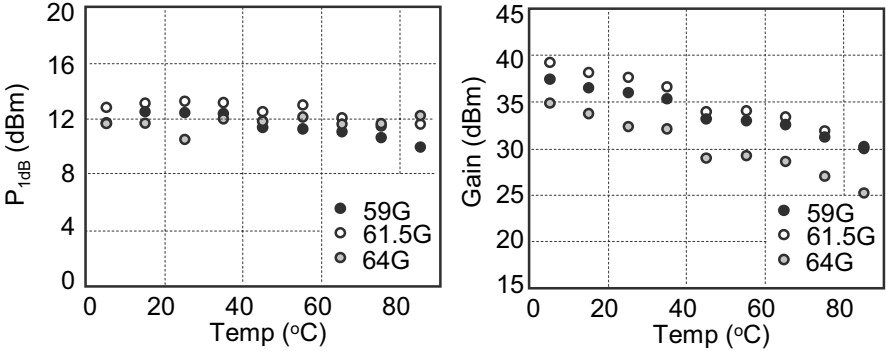


Fig. 4.53 Measured gain and output power of transmitter over temperature [14] (© IEEE 2006).

spur from $3\times\text{LO}$ feedthrough is < -20 dBm. With no DC offset correction applied, the external I/Q quadrature accuracy is within $\pm 2^\circ$. At P_{1dB} , the Tx consumes 190 mA from 2.7 V and 72 mA from 4 V (PA). Table 4.3 summarizes the Tx measured performance.

In summary, a highly-integrated receiver and transmitter chipset for data communications in the 60 GHz ISM band has been demonstrated. The Rx NF is approximately 6 dB and the Tx output P_{1dB} is 10-12 dBm, with provides ample link budget for many applications. Robust performance is maintained at 85°C . The 500 mW Rx and 800 mW Tx power consumption, combined with the wide available bandwidth at 60 GHz, means that data can be transmitted with good energy efficiency, which is

Table 4.3 Summary of transmitter receiver performance

Metric	Measured Value
Frequency Range	59-64 GHz
Gain	34-37 dB
P_{1dB} (out)	10-12 dBm
P_{sat}	15-17 dBm
Image Rejection	20-30 dB
Carrier Suppression	21-25 dB
3xLO Spur	-25 to -20 dBm
Phase Noise, (1MHz) tripled	-85 to -90 dBc/Hz
	< -130 dBc/Hz floor
I/Q Balance	± 2 deg, ± 0.5 dB
Power Dissipation	190 mA, 2.7 V
	72 mA, 4.0 V

desirable for battery-operated devices. The chips are architecturally flexible enough to work with a range of modulation formats.

4.6 Conclusion

Basic communication building blocks such as low-noise amplifiers and down-conversion mixers have been discussed extensively in this chapter. Many of the considerations given here apply directly in the design of other building blocks involving gain and frequency translation. Using core building blocks such as transmission lines, capacitors, and transistors, the design is relatively straightforward and similar to RF design techniques. Transmission lines are more prominent due to the length scalable reactance and better predictability of ground currents. Compared to RF design, though, mm-wave gain stages have much less available gain, requiring matching networks to trade-off between gain, noise figure, and linearity. For the same reason, stability considerations are much more important due to the reduced isolation in amplifiers. Since gain cannot be sacrificed for stability, the matching networks must be carefully designed to avoid instability over process and temperature. The layout and design of cascode gain stages, the workhorse of RF amplifiers, also requires rethinking due to the role of the second high frequency pole of the amplifier cascade. Layout and tuning techniques were discussed to circumvent these problems. Several mixers designs were highlighted, such as a single gate mixer, a “dual gate” mixer, and a classic Gilbert cell mixer. Due to the low gain of the transconductors at mm-wave frequencies, the choice between the Gilbert cell mixer and other topologies is not obvious. Several simple mixers were highlighted which offer modest conversion loss and good linearity. Finally, case studies of the design of complete building blocks in CMOS and SiGe are provided to illustrate the achievable performance for silicon technology.

References

1. L. W. Couch, *Modern Communication Systems: Principles and Applications*, Englewood Cliffs, NJ: Prentice-Hall, 1995.
2. B. Razavi, *RF Microelectronics*, Upper Saddle River, NJ: Prentice-Hall, 1998.
3. G. Gonzalez, *Microwave Transistor Amplifiers: Analysis and Design*, 2nd edition, Upper Saddle River, NJ: Prentice-Hall, 1997.
4. P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 3rd edition, New York, NY: John Wiley & Sons, Inc., 1993.
5. S. P. Voinigescu, M. C. Maliepaard, J. L. Showell, G. E. Babcock, D. Marchesan, M. Schroter, P. Schvan, and D. L. Hareme, "A Scalable High-Frequency Noise Model for Bipolar Transistors with Application to Optimal Transistor Sizing for Low-Noise Amplifier Design," *IEEE J. Solid-State Circuits*, vol. 32, no. 9, pp. 1430–1438, Sept. 1997.
6. B. A. Floyd, "A CMOS Wireless Interconnect System for Multigigahertz clock distribution," Ph.D. Dissertation, University of Florida, Gainesville, FL, 2001.
7. K. Hartmann and M. Strutt, "Changes of the four noise parameters due to general changes of linear two-port circuits," *IEEE Trans. Electron Devices*, vol. 20, pp. 874–877, Oct. 1973.
8. S. Reynolds, B. Floyd, U. Pfeiffer, and T. Zwick, "60GHz transistor circuits in SiGe bipolar technology," *IEEE ISSCC Dig. Tech. Papers*, Feb. 2004, pp. 442–538.
9. B. A. Floyd, "V-band and W-band SiGe bipolar low-noise amplifiers and voltage-controlled oscillators," *IEEE RFIC*, June 2004, pp. 295–298.
10. B. A. Floyd, S. K. Reynolds, U. R. Pfeiffer, T. Zwick, T. Beukema, and B. Gaucher, "SiGe bipolar transistor circuits operating at 60 GHz," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 156–167.
11. B. Jagannathan, et al., "Self-aligned SiGe NPN transistors with 285GHz f_{MAX} and 207GHz f_T in a manufacturable technology," *IEEE Electron Device Lett.*, vol.23, no.5, 2002.
12. B. Floyd, S. Reynolds, U. Pfeiffer, T. Beukema, J. Grzyb, and C. Haymes, "A silicon 60GHz receiver and transmitter chipset for broadband communications," *IEEE ISSCC Dig. Tech. Papers*, Feb. 2006, pp. 184–185.
13. U. R. Pfeiffer, J. Grzyb, D. Liu, B. Gaucher, T. Beukema, B. A. Floyd, and S. K. Reynolds, "A chip-scale packaging technology for 60-GHz wireless chipsets," *IEEE Trans. Microwave Theory Tech.*, vol. 54, no. 8, pp. 3387–3397, Aug. 2006.
14. S. Reynolds, B. Floyd, U. Pfeiffer, T. Beukema, J. Grzyb, C. Haymes, B. Gaucher, and M. Soyuer, "A silicon 60GHz receiver and transmitter chipset for broadband communications," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2820–2831, Dec. 2006.
15. B. Heydari, M. Bohsali, E. Adabi, A.M. Niknejad, "Low-Power mm-Wave Components up to 104GHz in 90nm CMOS," *IEEE ISSCC Dig. Tech. Papers*, pp. 200–201, 597, Feb. 2007.
16. B. Heydari, M. Bohsali, E. Adabi, A.M. Niknejad, "mm-Wave devices and circuit blocks up to 104 GHz in 90nm CMOS," to appear in *IEEE J. Solid-State Circuits*.
17. C. H. Doan, S. Emami, A. M. Niknejad, R. W. Brodersen, "Millimeter-Wave CMOS Design," *IEEE Journal of Solid-State Circuits*, vol. 40, pp.144– 155, Jan. 2005.
18. T. Yao, M. Gordon, K. Yau, M.T. Yang, and S.P. Voinigescu, "60-GHz PA and LNA in 90- nm RF-CMOS," *IEEE RFIC Symposium Digest*, pp. 147–150, June 2006.
19. C.-M. Lo, C.-S. Lin, H. Wang, "A Miniature V-band 3-Stage Cascode LNA in 0.13 μ m CMOS," *IEEE ISSCC Dig. Tech. Papers*, pp. 322–323, Feb. 2006.
20. X. Guan and A. Hajimiri, "A 24-GHz CMOS front-end," *IEEE J. Solid-State Circuits*, vol. 39, no. 2, pp. 368–373, Feb. 2004.
21. Mitomo et al., "A 60GHz Receiver with frequency Synthesizer," presented at the *VLSI Symposium*, 2007.
22. D. Huang, R. Wong, C. Chien, M.-C.F. Chang, "1.2V and 8.6mW CMOS differential receiver front-end with 24 dB gain and -11dBm IRCP," *Electronics Letters*, vol. 42, issue 25, pp. 1449–1450, December 2006.
23. B. Heydari, P. Reynaert, E. Adabi, M. Bohsali, B. Afshar, M. A. Arbabian and A. M. Niknejad, "A 60-GHz 90-nm CMOS cascode amplifier with interstage matching," to be presented at *EU Microwave Conference (EuMic)*, 2007.

24. B. Afshar, A. M. Niknejad, "X/Ku Band CMOS LNA Design Techniques," *Proceedings of CICC*, pp. 389-392, Sept. 2006.
25. Payam Heydari, "Design and Analysis of a Performance-Optimized CMOS UWB Distributed LNA," *IEEE J. Solid-State Circuits*, vol. 42, no. 9, pp. 1892-1905, Sept. 2007.
26. E. Adabi, B. Heydari, M. Bohsali and A. M. Niknejad, "30 GHz CMOS Low Noise Amplifier," *IEEE RFIC Symposium Dig.*, pp. 625-628, June 2007.
27. S. Emami, C. H. Doan, A. M. Niknejad, and R. W. Brodersen, "A 60-GHz down-converting CMOS single-gate mixer," *IEEE RFIC Symposium Dig.*, pp. 163-166, June 2005.
28. M. Schefer, U. Lott, H. Benedickter, Hp Meier, W. Patrick, and W. Bachtold, "Active, monolithically integrated coplanar V-band mixer," in *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 1043-1046, June 1997.
29. S. A. Maas, *Microwave Mixers*, 2nd edition, Boston: Artech House.
30. A. M. Niknejad, *Electromagnetics for High-Speed Analog and Digital Communication Circuits*, Cambridge University Press, 2007.
31. S. Emami, C. H. Doan, A. M. Niknejad, R.W. Brodersen, "A 60GHz CMOS Front-End Receiver," *ISSCC Dig. Tech. Papers*, pp. 190-191, Feb. 2007.

Chapter 5

Voltage-Controlled Oscillators and Frequency Dividers

Jri Lee

5.1 Considerations of VCOs

Voltage-controlled oscillators (VCOs) and frequency dividers play critical roles in all synchronous circuits. They comprise the core components in phase-locked systems, sometimes necessitating co-design and having great influence on the overall performance. Even though we have witnessed a proliferation of VCO and divider topologies in the past two decades of Si RF, high performance oscillators and dividers operating in the mm-wave range continue to pose difficult challenges even in today's technology. We begin our discussion with basic oscillation properties as well as a VCO figure of merit.

Oscillation behavior can be explained in different ways. A well-known model is shown in Fig. 5.1(a), where an amplifier $A(s)$ is placed in a feedback loop. The overall transfer function is given by

$$H(s) = \frac{V_{out}}{V_{in}}(s) = \frac{A(s)}{1 + A(s)}. \quad (5.1)$$

To make it oscillate at certain frequency ω_{osc} , we must satisfy the Barkhausen criteria

$$|A(j\omega_{osc})| = 1 \quad (5.2)$$

$$\angle A(j\omega_{osc}) = 180^\circ, \quad (5.3)$$

which govern all kinds of oscillators. These conditions are actually self-proven: a “stable oscillation” exactly implies “unity loop gain” and “total phase shift of 0° ” (or a multiple of 360°). Physically, oscillation initiates from white noise, an initial condition (charge on a capacitor), or a “kick”, an input waveform with sufficient spectral content to excite the right half plane poles of the system. We set a loop gain greater than unity in actual designs so as to start up the oscillation. While the energy at other frequencies decays eventually, the component at ω_{osc} survives and grows up until the effective loop gain drops to unity. Since all the coefficients of $H(s)$ are real, the poles of $H(s)$ should be either real or complex conjugate pairs. In steady-state oscillation, all of the poles of $H(s)$ are pushed to the left-hand side except that one dominant pair locates at the $j\omega$ -axis [Fig. 5.1(b)]. The output in frequency domain can be represented as

$$V_{out}(s) = \frac{a_0 s + b_0}{s^2 + \omega_{osc}^2} + \sum_i \frac{c_i s + d_i}{(s + \alpha_i)^2 + \beta_i^2} + \sum_j \frac{e_j}{s + \gamma_j}, \quad (5.4)$$

where the coefficients $\{\alpha_i\}$ and $\{\gamma_j\}$ are positive. Since the second and the third terms of Eq. (5.4) produce transient responses only, the time-domain output in steady state is indeed a sinusoid.

By the same token, we can depict the loop gain $A(s)$ in Gaussian plane, arriving at a Nyquist plot as shown in Fig. 5.1(c). The curve passes through $(-1, 0)$ whenever a stable oscillation occurs.

In most cases, an oscillator could become useful only if its frequency is tunable. The tuning range must be large enough to cover the overall bandwidth of interest with sufficient margin for process, temperature and supply (PVT) variations. Also, in a PLL, the VCO gain (K_{VCO}) must be kept as constant as possible, otherwise the loop phase margin suffers. Compared to wireless LAN systems at 2.4 or 5.2 GHz, the 60-GHz indoor RF band, for example, has 7-GHz unlicensed band available. The VCO must accommodate as wide as 15% of the center frequency, across which no serious deviation is allowed.

The output purity of VCOs is quantified as phase noise, i.e., the cyclic uncertainty induced by the noise of the active and passive devices. The phase noise is defined as “the relative noise power per unit bandwidth at certain offset with respect to the carrier power”. That is,

$$L(\Delta\omega) = 10 \log_{10} \left(\frac{P_{noise}}{P_{signal}} \right) \text{ (dBc/Hz)}. \quad (5.5)$$

Phase noise has been studied extensively and a great volume of references can be found in the literature. Here we state without proof that for an inductor-based oscillator with quality factor Q , the phase noise can be represented as [1]

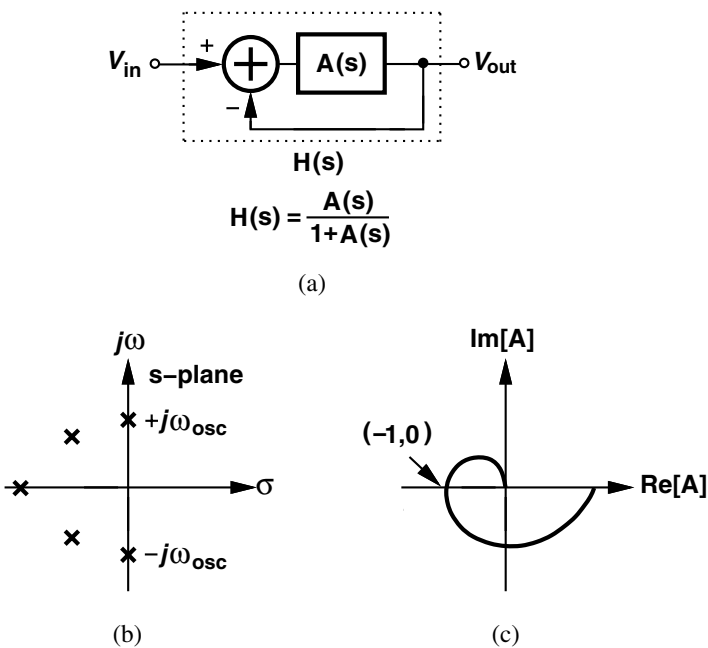


Fig. 5.1 (a) Feedback system, (b) poles of a certain oscillator in steady state, (c) its Nyquist plot.

$$L(\Delta\omega) \propto \frac{1}{Q^2} \cdot \left(\frac{\omega_0}{\Delta\omega}\right)^2 \quad (5.6)$$

at moderate frequency offset. At very small offset, phase noise drops faster due to some complex effects, presenting a steeper falling in this region. Figure 5.2 reveals a general phase noise spectrum.

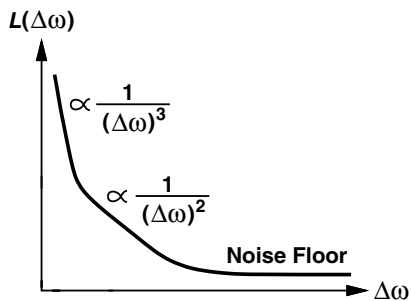


Fig. 5.2 Typical phase noise plot.

A popular figure of merit (FOM) for oscillators summarizes the important performance parameters, i.e., phase noise and power consumption P , to make a fair comparison:

$$\text{FOM} = L(\Delta\omega) + 10\log_{10} \left(\frac{\Delta\omega}{\omega_0} \right)^2 + 10\log_{10} \left(\frac{P}{1\text{mW}} \right). \quad (5.7)$$

The second term is to neutralize the effect of offset in $L(\Delta\omega)$ while taking the center frequency into account. The power consumption is calculated as dBm such that the unit of FOM remains the same as that of $L(\Delta\omega)$. For example, for a 40-GHz VCO with phase noise of -90 dBc/Hz while consuming 1 mW, the FOM is equal to -182 dBc/Hz.

5.2 Cross-Coupled Oscillators

LC-tank VCOs are probably the most pervasive ones in high-speed systems, providing numerous merits that satisfy the requirements of different applications. The simple yet symmetric configuration facilitates high-speed and differential designs with large swing, reasonable tuning range, and low power consumption. The plain structure also allows low-supply operation, even below 1 V. Depending on the inductor Q , the cross-coupled oscillators can achieve sufficiently low phase noise for most applications. We begin our discussion with the fundamental properties.

A typical realization of cross-coupled VCOs can be found in Fig. 5.3(a), where the pair M_1 - M_2 provides negative resistance $-2/g_{m1,2}$ (differentially) to compensate for the inductor loss R_P . At resonance, these two resistances cancel each other and the oscillation frequency is given by

$$\omega_{osc} = \frac{1}{\sqrt{LC_P}}, \quad (5.8)$$

where L and C_P denote the loading inductor and parasitic capacitance at output nodes, respectively, and M_3 - M_4 the MOS varactors. Barkhausen criteria imply that we must have $g_{m1,2} \geq (1/R_P)$ to make the circuit oscillate, while practical design would choose a higher value (≈ 3) to ensure oscillation over PVT variations.

It is instructive to derive an alternative expression for ω_{osc} with simplified conditions to examine what factors actually limit the operation frequency. Modeling the VCO as Fig. 5.3(b), we obtain R_P and ω_{osc} in stable oscillation as

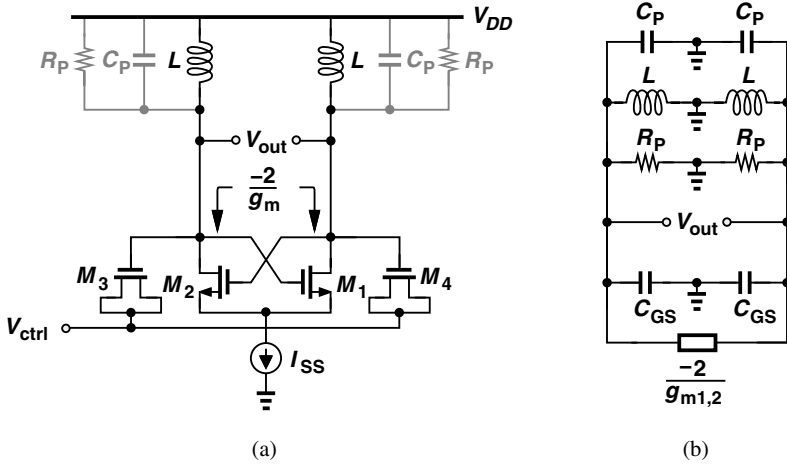


Fig. 5.3 (a) Typical LC oscillator with cross-coupled pair providing negative resistance, (b) its equivalent model with varactors neglected.

$$R_P = Q \cdot \omega_{osc} L = \frac{1}{g_{m1,2}} \quad (5.9)$$

$$\omega_{osc} = \frac{1}{\sqrt{2L\left(\frac{C_P}{2} + \frac{C_{GS}}{2}\right)}}, \quad (5.10)$$

where Q represents the quality factor of the tank, and C_{GS} the average gate-source capacitance contributed by $M_{1,2}$. Here, we take off the varactors for simplicity. If C_P is negligible as compared with C_{GS} (which is basically true at high frequencies), we arrive at

$$\omega_{osc} \approx \frac{1}{\sqrt{LC_{GS}}} \quad (5.11)$$

$$= \frac{1}{\sqrt{\frac{1}{g_m Q \omega_{osc}} C_{GS}}} \quad (5.12)$$

$$= \sqrt{Q \omega_T \omega_{osc}}, \quad (5.13)$$

where ω_T denotes the transit frequency of $M_{1,2}$. It follows that

$$\omega_{osc} = Q \cdot \omega_T. \quad (5.14)$$

In other words, a cross-coupled oscillator can operate at very high frequencies, given that the inductors provide a sufficiently high Q . In reality, however, several issues discourage ultra high-speed oscillation: (1) the on-chip inductors usually have a self-resonance frequency (f_{SR}) of only a few hundred GHz; (2) the varactors present significant loss at high frequencies, and it could eventually dominate the Q of the tank; (3) even (2) is not a concern, the on-chip inductors can never reach a very high Q due to the physical limitations; (4) C_P may not be negligible in comparison with other parasitics. Nonetheless, cross-coupled VCOs are still expected to operate at frequencies close to device f_T . For example, 50- and 96- and 140-GHz realizations¹ have been reported in 0.25- μm , 0.13- μm , and 90-nm CMOS technologies [2][3][4]. The final example illustrates oscillation above f_T .

It is important to know that, if the varactor's capacitance is much greater than other parasitics, the tuning range of an LC VCO approaches a constant and has nothing to do with the inductance. Figure. 5.4 illustrates such an effect. On the other hand,

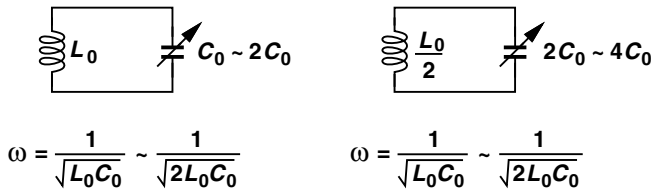


Fig. 5.4 Two LC networks with equal resonance frequency and tuning range. (The varactors are assumed to have 100% variation.)

lowering the inductance leads to smaller swing and puts the oscillator in danger of failing unless the current is increased (assuming that $R_P = Q\omega_0 L$ decreases). Consequently, it is always desirable to use inductors as large as possible.

Several techniques have been developed to improve the performance of LC oscillators. Figure 5.5(a) shows a popular topology moving the tail current to the top to setup the output common-mode around $V_{DD}/2$. The noise of the top current source may disturb the voltage at node P and hence modulate the frequency, resulting in higher phase noise. Figure 5.5(b) incorporates PMOS devices, but the oscillation frequency (or tuning range) degrades due to the extra parasitic capacitance. Note that the tail current plays important role here, because it defines the bias current (and hence the output amplitude if the inductor Q is known) while giving a high impedance to ground so as to maintain a more constant quality factor for oscillation. The tail current in Fig. 5.5(a) and (b) can be removed to accommodate low-supply operation but at a cost of higher supply sensitivity. An approach preserving these merits while eliminating its noise contribution is illustrated in Fig. 5.5(c). A large bypass capacitor C_p absorbs the noise of the current source M_1 . With the L_s - C_s network resonating at twice the output frequency, the common-source node P still experiences a high

¹ Here we consider only the fundamental frequency.

impedance to ground. Differential voltage control is also achievable by adding two sets of varactors with opposite direction, as illustrated in Fig. 5.5(c). The differential operation improves the common-mode rejection by 10-20 dB.

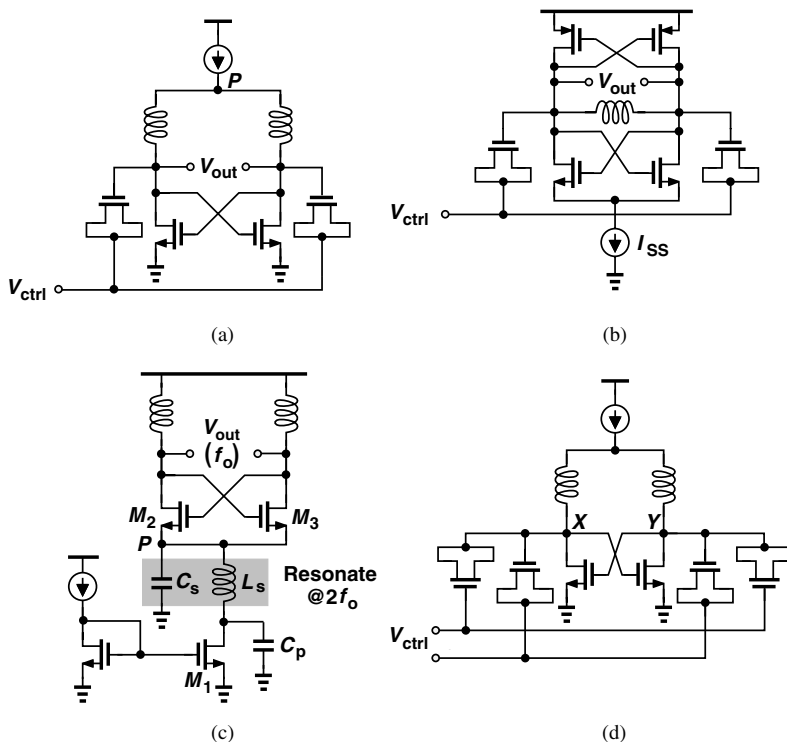


Fig. 5.5 LC VCO modification with (a) top biasing, (b) dual pairs, (c) differential control.

Inductors can also be improved by changing the physical shapes. One useful modification for differential circuit is to wrap the spiral symmetrically [Fig. 5.6(a)]. Due to the differential operation, the effective substrate loss is reduced by a factor of 2, leading to a higher Q . The only side effect is that the spacing between turns needs to be wider so as to minimize the interwinding capacitance. Vertical stacking is another useful technique to shrink the occupied area for a given inductance [Fig. 5.6(b)]. Depending on the mutual coupling factor, the inductance of a two-layer structure is around 3.5 ~ 4 times larger than that of a single layer one with the same area. Note that the two layers should be kept as far as possible in order to maximize the self-resonance frequency. A method combining these two techniques is depicted in Fig. 5.6(c). Recognized as a differentially-stacked inductor, it preserves the benefits from both structures. More details can be found in [5][6][7].

A few techniques are commonly-used to extend the tuning range. A straightforward method is to employ a capacitor array (preferably binary-weighted for better

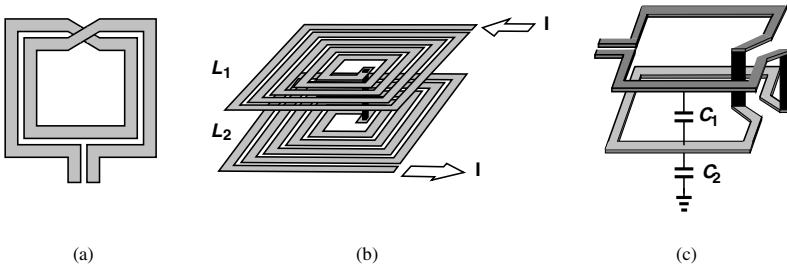


Fig. 5.6 (a) Differential inductor, (b) stacked inductor, (c) differentially-stacked inductor.

efficiency) to tune the VCO coarsely [Fig. 5.7(a)][8]. Such a band selection mechanism sometimes benefits the PLL design because the VCO gain becomes smaller. Similar approach can be applied to inductors. Figure 5.7(b) shows an example where segmented inductors with switches accomplish the digital tuning [9]. Another interesting realization of variable inductors is depicted in Fig. 5.7(c), where the auxiliary inductor L_2 is tuned gradually by means of $M_{3,4}$ and V_{ctrl} . Nevertheless, all of the

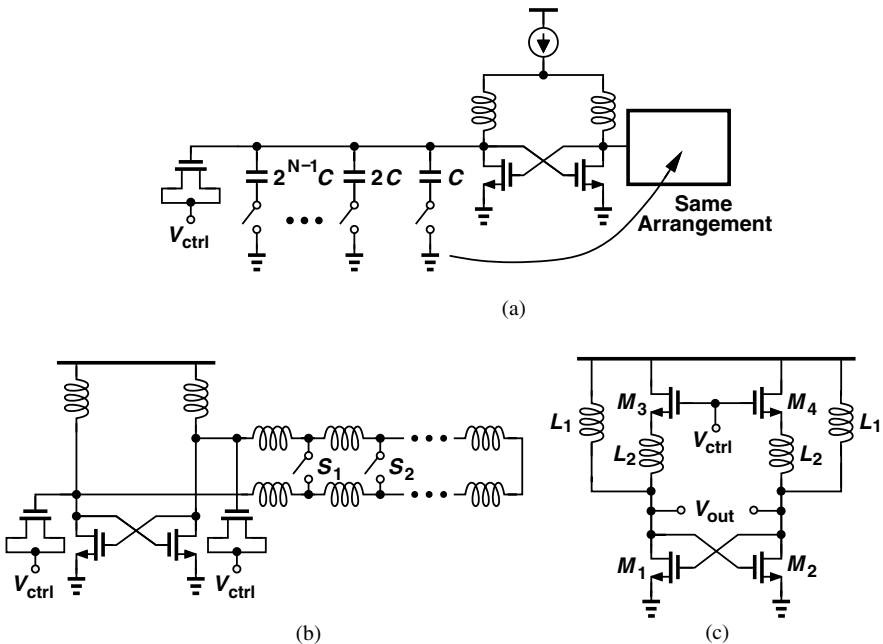


Fig. 5.7 Methods to enlarge the tuning range using (a) Capacitor array, (b) switching inductors, (c) auxiliary inductors.

circuits in Fig. 5.7 suffer from degradation on the tank Q (and hence the phase noise) because of the finite resistance of the MOS switches.

It is noteworthy that the basic cross-coupled oscillators are very efficient, and careless modification of the structure can lead to unpredictable results. One example using capacitive degeneration is illustrated in Fig. 5.8(a). Like a relaxation oscillator, the impedance seen looking into the cross-coupled pair is given by

$$R_{eq} = -\frac{2}{g_{m1,2}} - \frac{1}{sC_E}, \quad (5.15)$$

and the equivalent small-signal model is shown in Fig. 5.8(b). Intuitively, such a

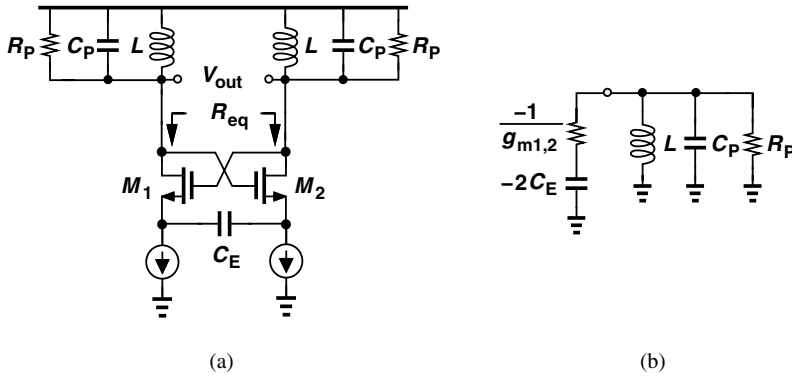


Fig. 5.8 Cross-coupled VCO with capacitive degeneration.

degeneration provides a negative capacitor to cancel out part of the positive capacitor C_P , raising the oscillation frequency. In reality, however, this frequency boosting is accomplished at a cost of weakening the negative resistance, making the circuit harder to oscillate. To see why, let us first consider a general transformation between series and parallel networks. As shown in Fig. 5.9(a), a series circuit containing R_1 and C_1 can be converted to a parallel one by equating the impedance. Defining Q_d as $1/(R_1 C_1 \omega)$, we arrive at

$$\frac{R_2}{R_1} = 1 + Q_d^2 \quad (5.16)$$

$$\frac{C_2}{C_1} = \frac{Q_d^2}{1 + Q_d^2} \quad (5.17)$$

where R_2 and C_2 are of the equivalent parallel combination. Figure 5.9(b) plots R_2 and C_2 as a function of Q_d . Obviously, depending on Q_d , the transformed network behaves differently. For $Q_d \gg 1$, $C_2 \approx C_1$ and $R_2 \approx Q_d^2 R_1$; whereas for lower Q_d , both R_2 and C_2 degrade.

Applying this result into Fig. 5.8(b), we arrive at the small-signal model in Fig. 5.10. The resonance frequency now becomes

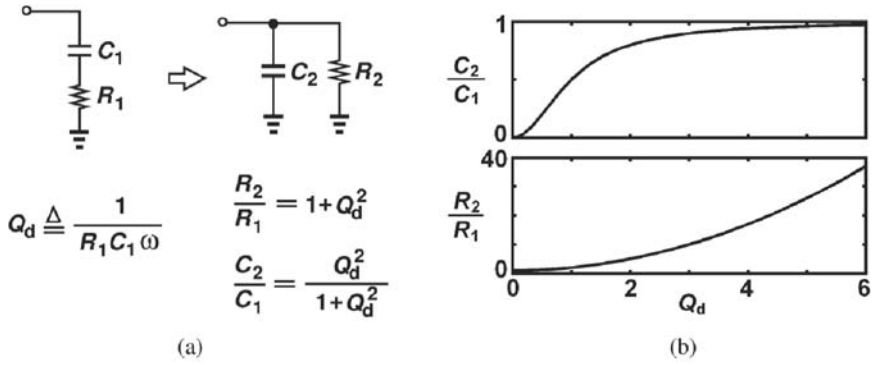


Fig. 5.9 Conversion between series and parallel RC networks.

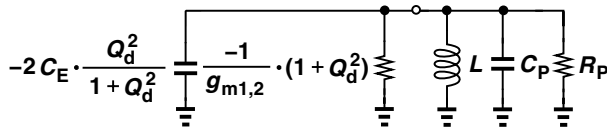


Fig. 5.10 Modification of Fig. 5.8(b).

$$\omega_{osc} = \frac{1}{\sqrt{L(C_P - 2C_E \cdot \frac{Q_d^2}{1 + Q_d^2})}}. \quad (5.18)$$

Although degraded, it is indeed a boost in the frequency. However, a more difficult condition is imposed on the start-up oscillation:

$$g_{m1,2} \geq \frac{1 + Q_d^2}{R_P}. \quad (5.19)$$

For a Q_d of 3, this circuit needs a transconductance 10 times larger in order to ignite (and maintain) the oscillation. As a result, wider devices may be required to implement $M_{1,2}$, leading to less improvement or even deterioration in oscillation frequency. The circuit may consume more power as well.

5.3 Colpitts Oscillator

Another important VCO topology that has been widely used in high-speed systems is Colpitts oscillator. First proposed in 1920's [10], this type of oscillator could be operated with only one transistor. In modern times, the abundance of transistors and the desire for differential circuits favors a symmetric Colpitts oscillators.

A Colpitts VCO can be easily understood by examining a resonating circuit shown in Fig. 5.11(a), where an inductor is sitting across the drain and gate of a MOS with two capacitors C_1 and C_2 connected to these nodes. In order to oscillate, the signal in the feedback path through the C - L - C network must satisfy Barkhausen criteria. Breaking the loop and exciting it with an input V_1 [Fig. 5.11(b)], we obtain the loop

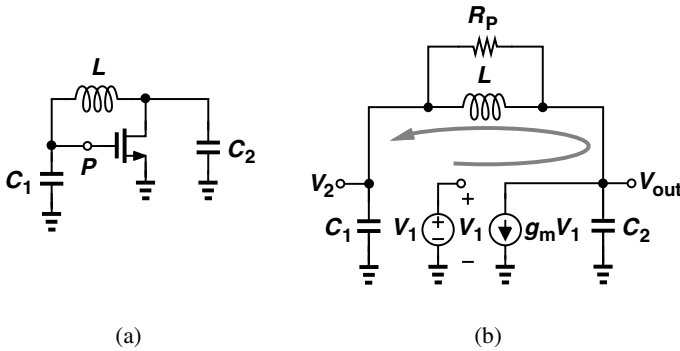


Fig. 5.11 (a) Colpitts oscillator, (b) its linear model with feedback broken at node P .

gain as

$$\frac{V_2}{V_1}(s) = -g_m \cdot \frac{R_P + sL}{s^3 LC_1 C_2 R_P + s^2 L(C_1 + C_2) + s R_P(C_1 + C_2)}. \quad (5.20)$$

To make the oscillation happen at a frequency ω_{osc} , we have $|V_2/V_1| \geq 1$ and $\angle(V_2/V_1) = 0^\circ$. In other words, at $\omega = \omega_{osc}$, the ratio of the real and imaginary parts of the numerator must be equal to that of the denominator:

$$\frac{R_P}{\omega_{osc} L} = \frac{-\omega_{osc}^2 L(C_1 + C_2)}{\omega_{osc} R_P(C_1 + C_2) - \omega_{osc}^3 L R_P C_1 C_2}. \quad (5.21)$$

It follows that

$$\omega_{osc} = \frac{1}{\sqrt{\frac{LC_1C_2}{C_1+C_2}}} \cdot \sqrt{1 + \frac{1}{Q^2}} \approx \sqrt{\frac{1}{L} \left(\frac{1}{C_1} + \frac{1}{C_2} \right)}. \quad (5.22)$$

Here, Q denotes the quality factor of the inductor and $R_P = \omega_0 LQ$. The loop gain requirement yields

$$\left| \frac{V_2}{V_1} \right| (j\omega_{osc}) = \frac{g_m R_P}{\omega_{osc}^2 L (C_1 + C_2)} \geq 1. \quad (5.23)$$

As a result, we have the following condition for oscillation:

$$g_m R_P \geq \frac{(C_1 + C_2)^2}{C_1 C_2} \geq 4. \quad (5.24)$$

An alternative explanation of a Colpitts oscillator is to investigate the impedance seen looking into the gate-drain port of such a circuit [Fig. 5.12(a)]. It can be shown that R_{eq} is given by

$$R_{eq} = \frac{g_m}{C_1 C_2 s^2} + \frac{1}{C_2 s} + \frac{1}{C_1 s}, \quad (5.25)$$

which is equivalent to a negative resistance $-g_m/(C_1 C_2 \omega^2)$ in series with a capacitor $C_1 C_2/(C_1 + C_2)$. If the quality factor of this RC network is high, we can approximate it as a parallel combination as shown in Fig. 5.12(b). Obviously the circuit may oscillate if the negative resistance is strong enough to cancel out the inductor loss R_P . As expected, the oscillation frequency is equal to

$$\omega_{osc} = \sqrt{\frac{1}{L} \left(\frac{1}{C_1} + \frac{1}{C_2} \right)}, \quad (5.26)$$

same result as Eq. (5.22). Equation. (5.24) can be obtained with a similar approach.

Depending on the bias, the prototype in Fig. 5.12(a) provides three topologies of Colpitts oscillators, as shown in Fig. 5.13 [11]. Among them, Fig. 5.13(a) reveals the greatest potential for high-speed operation, since C_1 can be realized by the intrinsic capacitance C_{GS} of M_1 . The capacitor C_2 is replaced by a varactor M_2 to accomplish the frequency tuning. At resonance, all the components oscillate at the same frequency ω_{osc} , including the drain current of M_1 . That allows us to place a loading R_D at drain and take the voltage output from this node. Inductive peaking could be an option here if the output needs to drive large capacitance. Figure 5.13(b) provides another

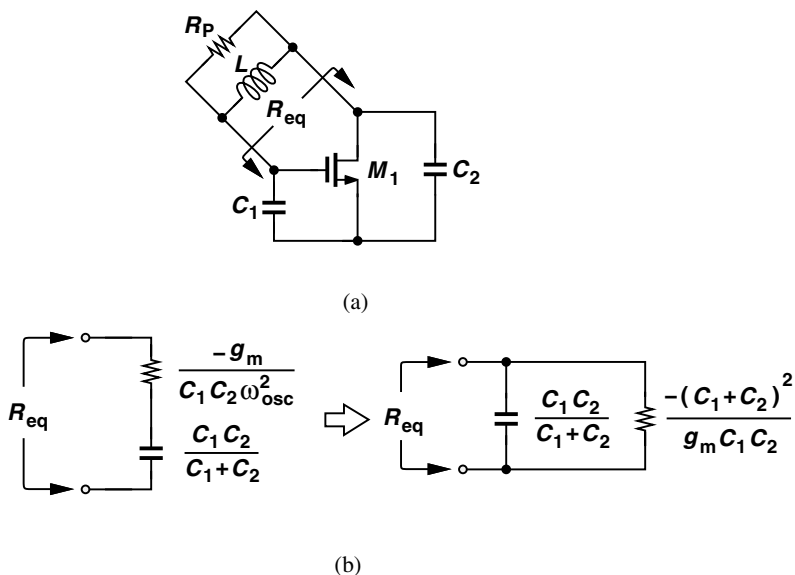


Fig. 5.12 Alternative approach to analyze Colpitts oscillators by examining the equivalent impedance.

structure with C_2 connecting to the source and drain of M_1 . The parasitic capacitance C_{GS} would limit the oscillation frequency or equivalently the tuning range, making this topology less attractive. The structure in Fig. 5.13(c) requires L to be floating at both ends, presenting more parasitic to the circuit. For this reason, the topology in Fig. 5.13(c) is rarely used.

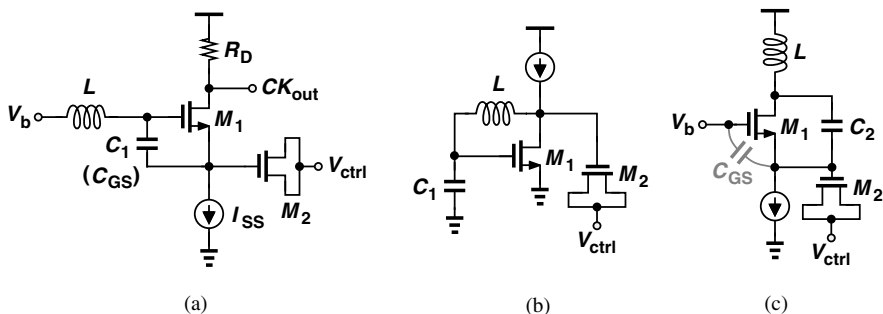


Fig. 5.13 Three realizations of Colpitts oscillator, (a) common-drain, (b) common-source, (c) common-gate.

Despite many advantages, the circuit in Fig. 5.13(a) still suffers from two drawbacks: the single-ended operation makes the oscillator vulnerable to supply noise,

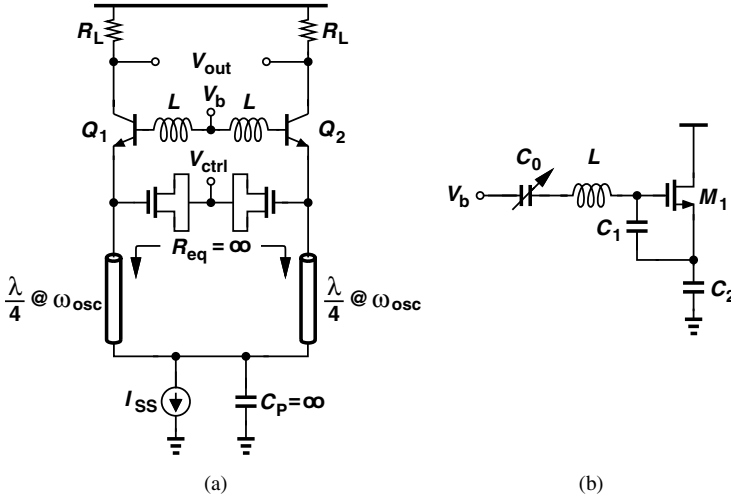


Fig. 5.14 (a) Differential Colpitts oscillator, (b) Clapp oscillator.

and the capacitance contributed by the tail current source degrades the oscillation frequency. To remedy these issues, we usually implement the Colpitts oscillator as a differential configuration with $\lambda/4$ -lines between $Q_{1,2}$ and I_{SS} . Figure 5.14(a) illustrates such a realization. The combined bias points of the symmetric circuit facilitate differential operation, and $\lambda/4$ -transmission lines make the equivalent impedance looking down (R_{eq}) become infinity.² Colpitts VCOs operating at 60 GHz and beyond with silicon compound technologies have been reported extensively [12][13][14]. In fact, the current source can be replaced with a “choke” inductor, or a sufficiently large inductor such that the impedance to ground is dominated by the capacitance for proper feedback. A Colpitts oscillator taking this approach is presented in [15], which demonstrates 104 GHz operation in a 90-nm CMOS technology.

The circuit in Fig. 5.13(a) tunes the frequency at the risk of losing stability or failing the oscillation. According to Eq. (5.24), $g_m R_p$ must be greater than $(C_1 + C_2)^2 / (C_1 C_2)$, which varies as the control voltage changes. To guarantee safe margin for oscillation, one can introduce another capacitor C_0 (which is variable) in series with L and level C_1 and C_2 fixed as depicted in Fig. 5.14(b). The oscillation frequency therefore becomes

$$\omega_{osc} = \sqrt{\frac{1}{L} \left(\frac{1}{C_0} + \frac{1}{C_1} + \frac{1}{C_2} \right)}. \quad (5.27)$$

Also known as “Clapp oscillator”, this circuit inevitably suffers from less tuning range.

² Here we assume a very large bypass capacitor C_P is used. Line length other than $\lambda/4$ could be chosen for finite C_P .

One important application of Colpitts oscillators is the so-called “Pierce oscillator”. As shown in Fig. 5.15, it incorporates a piezoelectric crystal (serving as an inductor) and two capacitors C_1 and C_2 to form a Colpitts oscillator. Here, the crystal

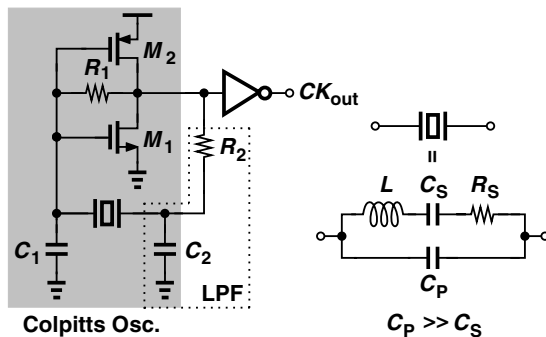


Fig. 5.15 Example of Pierce oscillator.

can be modeled as a series RLC network (i.e., L , C_S and R_S) in parallel with another capacitor C_P and $C_P \gg C_S$. Similar to M_1 in Fig. 5.12(a), the inverter-like amplifier M_1 and M_2 provides negative resistance to compensate for the loss. Note that the circuit is self-biased through R_1 such that both M_1 and M_2 are in saturation. The reader can easily prove that the oscillation frequency is equal to

$$\omega_{osc} \approx \frac{1}{\sqrt{LC_S}}, \quad (5.28)$$

which is an unchangeable value for a given crystal. To increase oscillation stability, R_2 can be added in the loop to dampen the higher order harmonics. Such a crystal-based oscillator achieves marvelous frequency stability in the presence of temperature variation, and is extensively used as a reference clock in various applications.

5.4 Other Topologies

In addition to the cross-coupled and Colpitts oscillators, many other VCO topologies have been proposed to achieve high-speed operation. We look at some representative structures in this section.

5.4.1 mm-Wave Oscillators

Similar to Colpitts VCOs, mm-wave oscillators can also create periodic signals with only one active device. The approach of a mm-wave oscillator is illustrated in Fig.

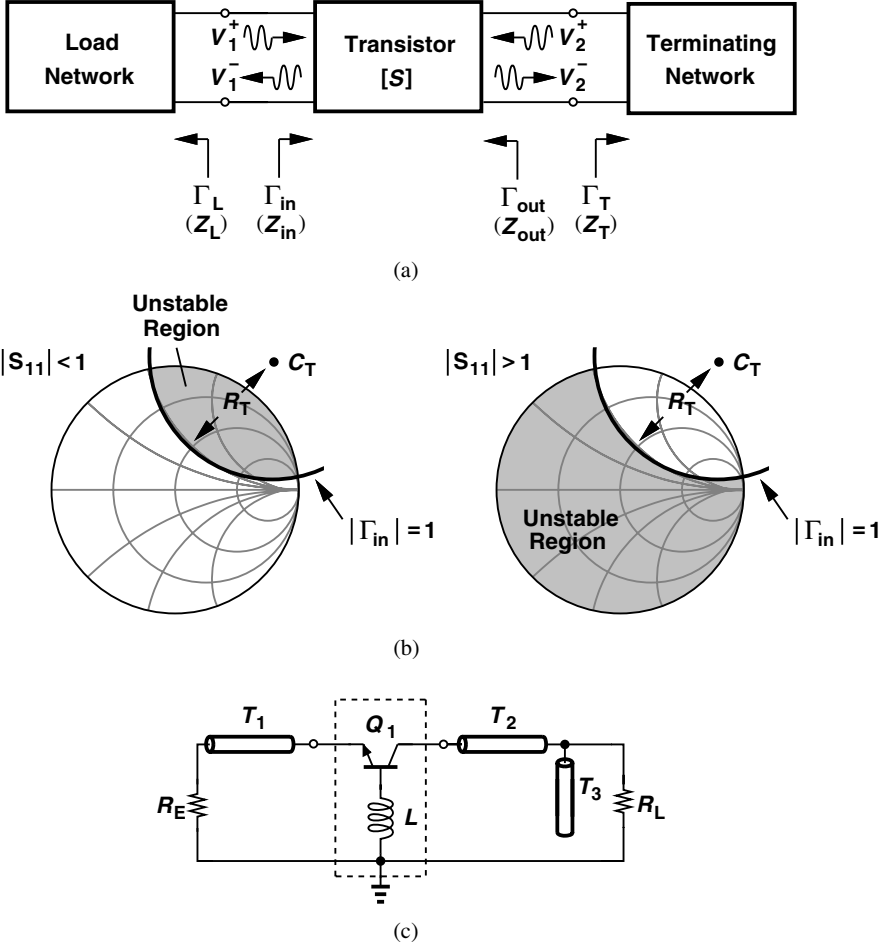


Fig. 5.16 (a)Design of mm-wave oscillators, (b)Smith chart for determining Γ_T , (c)example of mm-wave oscillators.

5.16, where Γ_L , Γ_{in} , Γ_{out} , and Γ_T denote the corresponding reflection coefficients. In order to make the circuit oscillate, we need to build up a device which becomes highly unstable in the vicinity of the desired frequency. It is clear that

$$\Gamma_{in} = \frac{V_1^-}{V_1^+} = S_{11} + \frac{S_{12}S_{21}\Gamma_T}{1 - S_{22}\Gamma_T}, \quad (5.29)$$

and oscillation would occur only if the input impedance Z_{in} possesses negative real part. Assuming all the connections are implemented as 50- Ω transmission lines, we have $|\Gamma_{in}| > 1$ and the boundary condition $|\Gamma_{in}| = 1$ can be further derived to correlate with Γ_T . It can be shown that [16]

$$|\Gamma_T - C_T| = R_T, \quad (5.30)$$

where

$$C_T = \frac{(S_{22} - \Delta S_{11}^*)^*}{|S_{22}|^2 - |\Delta|^2} \quad (5.31)$$

$$R_T = \left| \frac{S_{12} - S_{21}}{|S_{22}|^2 - |\Delta|^2} \right|, \quad (5.32)$$

and Δ represents the determinant of the scattering matrix:

$$\Delta = S_{11}S_{22} - S_{12}S_{21}. \quad (5.33)$$

The selection of Γ_T for different $|S_{11}|$ conditions is illustrated in Fig. 5.16(b). For $|S_{11}| < 1$, the unstable region is the inside region of the $|\Gamma_{in}| = 1$ circle that intersects the Smith chart, whereas for $|S_{11}| > 1$ it is the whole Smith chart outside the $|\Gamma_{in}| = 1$ circle. Γ_T must locate in the unstable region in order to oscillate. In mm-wave oscillator design, common-source or common-gate devices with positive feedback are often incorporated. After the transistor configuration is selected, the $|\Gamma_{in}| = 1$ circle becomes readily available and Γ_T can be properly chosen to produce a large negative resistance at the input of the transistor. The load impedance Z_L needs to match Z_{in} as well. Typical design would require a $3 \times$ margin (i.e., $R_L = -R_{in}/3$) to secure oscillation. The imaginary matching $X_L = -X_{in}$ determines the oscillation frequency. Figure 5.16(c) depicts a design example. Hence, Q_1 - L network serves as a two-port and matching techniques are employed. With a good matching, the output could deliver a high power. Due to the simplicity, the mm-wave oscillator can operate at high frequency as well. The reader can refer to [16] for more details.

The major drawback of this circuit is the difficulty of frequency tuning. The length and characteristic impedance of a transmission line are almost inalterable because of physical limitation and matching requirements. Single-ended operation is another minus. The reader can see that a differential mm-wave oscillator resembles the Colpitts VCO [e.g., Fig. 5.14(a)] in many aspects.

5.4.2 Push-Push Oscillators

One important application of $\lambda/4$ transmission line technique is the push-push oscillators. As suggested by its name, this type of oscillator takes the 2nd-order harmonic from the common-mode node, and amplifies it properly as an output. Note that second-order harmonic is generated by the nonlinearity of the circuit, which manifests itself in large-signal operation. Figure 5.17 reveals an example, where V_P needs

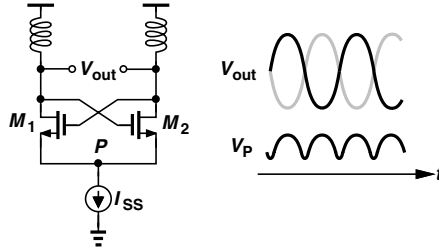


Fig. 5.17 Generation of 2nd-order harmonic.

to swing up and down at twice the fundamental frequency so as to maintain a constant I_{SS} . Similar to a frequency doubler, the desired harmonic can be extracted while the others are suppressed.

Since node P suffers from large parasitic capacitance, we usually resort to other common-mode points to obtain the output. Two examples of circuit-level realization based on cross-coupled and Colpitts structures are illustrated in Fig. 5.18. The $\lambda/4$ lines in both cases reinforce the $2\omega_{osc}$ signal by providing an equivalent open at node P when looking into it, and the output power could be quite large if proper matching

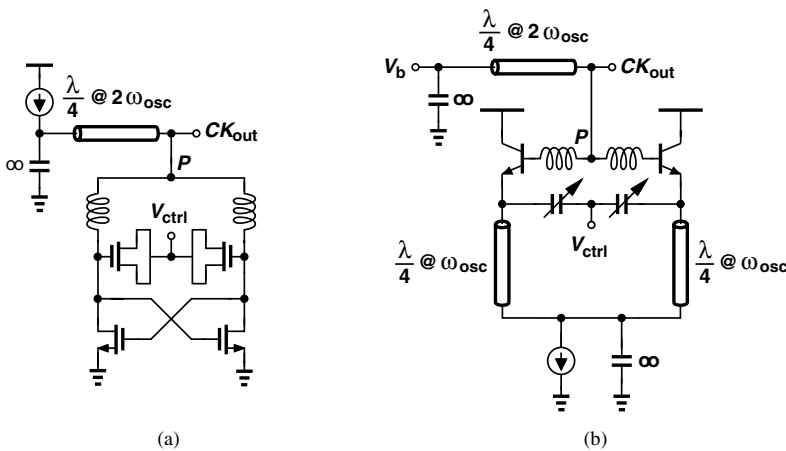


Fig. 5.18 Push-push VCOs based on (a) cross-coupled, (b) Colpitts topologies.

is achieved. Compared with typical frequency doublers, this topology consumes less power and area, resulting in a more efficient approach. More details are described in [17][18].

The push-push oscillator can only provide a single-ended output. In addition, tuning the fundamental frequency could result in a mismatch in the $\lambda/4$ lines, potentially leading to lower output power.

5.4.3 Distributed Oscillators

Another distinctive VCO topology shooting for high-speed operation is the distributed oscillator. As shown in Fig. 5.19, the output of a distributed amplifier is returned back to the input, yielding wave circulation along the loop. Oscillation is therefore obtained at any point along the transmission line. Here, the transmission line loss is

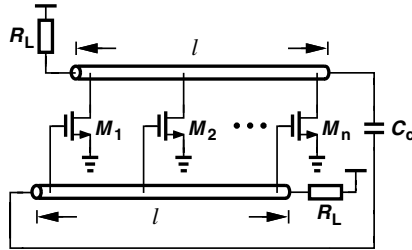


Fig. 5.19 Distributed oscillator.

overcome by the gain generated along the line. To be more specific, we assume the two propagation lines in Fig. 5.19 to be identical, i.e., the characteristic impedances, group velocities, and physical lengths are the same. The oscillation period under such circumstances is nothing more than twice the propagation time along the length l :

$$f_{osc} = \frac{1}{2l\sqrt{L_0C_0}} \quad (5.34)$$

where L_0 , C_0 denote the equivalent inductance and capacitance (with the MOS capacitance included) per unit length. It can be shown that the oscillation frequency is commensurate with the device f_T [8].

While looking attractive, the distributed oscillator suffers from a number of drawbacks: (1) the group velocities along the two lines may deviate from each other due to the difference between the gate and drain capacitance; (2) the circuit would need larger area and higher power dissipation, (3) the frequency tuning could be difficult. The third point becomes clear if we realize that adding any varactors to the lines can

cause significant degradation on the oscillation frequency and the quality factor Q . Varying the bias voltage of the transistors may change the intrinsic parasitics (and therefore the oscillation frequency) to some extent, but the imbalanced swing and the mismatch between the lines could make things worse. The circuit may even stop oscillating in case of serious deviation. Note that placing a “short-cut” on the lines by steering the current of two adjacent transistors is plausible as well [19]: it is hard to guarantee that the wave still propagates appropriately along the lines while both devices are partially on.

A modification of distributed oscillators can be found if we terminate a transmission line by itself. The circuit is based on the concept of the differential stimulus of a closed-loop transmission line at evenly-spaced points, as illustrated conceptually in Fig. 5.20(a). In contrast to regular distributed oscillators, the transmission line requires no termination resistors, lowering phase noise and enlarging voltage swings. The circuit can be approximated by lumped inductors and capacitors, and one example is shown in Fig. 5.20(b). Here, eight inductors form a loop with four differential negative- Gm cells driving diagonally opposite nodes. In steady state, the eight nodes are equally separated by 45° , providing multiphase output if necessary.

The oscillation frequency of the circuit is uniquely given by the travel time of the wave around the loop. We write the oscillation frequency of this topology as

$$f = \frac{1}{8\sqrt{LC}} \quad (5.35)$$

where L and C , respectively, denote the lumped inductance and capacitance of each of the eight sections. The circuit can be further modified as shown in Fig. 5.20(c) to avoid long routing, and the negative- Gm cell can be simply implemented as Fig. 5.20(d). The PMOS transistors help to shape the rising and falling edges while providing lower $1/f$ noise.

One interesting issue in such a VCO is that, due to symmetry, the wave may propagate clockwise rather than counterclockwise. To achieve a more robust design, a means of detecting the wave direction is necessary. Since nodes that are 90° apart in one case exhibit a phase difference of -90° in the other case, a flipflop sensing such nodes generates a constant high or low level, thereby providing a dc quantity indicating the wave direction. Other approach to avoid direction ambiguity can be found in [20].

5.5 Considerations of Dividers

Frequency dividers are also of great concern in communication systems. A good divider must provide correct frequency division over the whole band of interest while contributing negligible noise. At frequency above 10 GHz, designers begin to face a

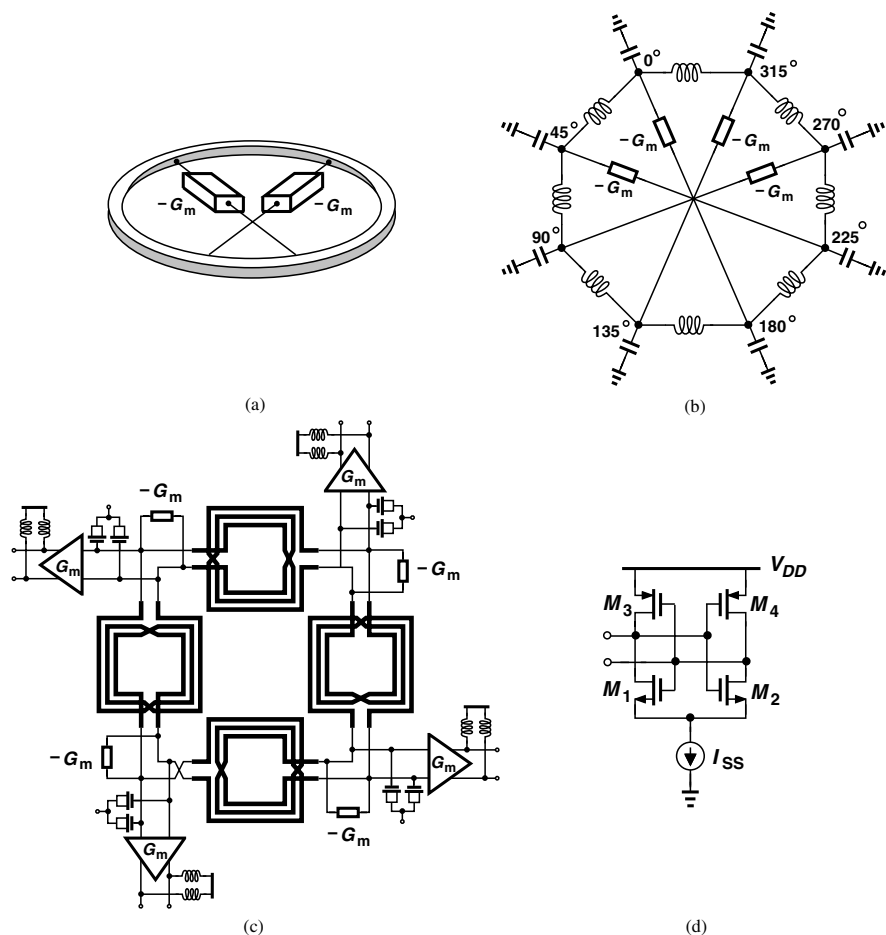


Fig. 5.20 (a) Oscillator based on closed-loop transmission line, (b) half-quadrature realization, (c) modification of (b), (d) implementation of $-G_m$ cell.

tradeoff between the input frequency and operation range. Generally speaking, the injection-locked dividers achieves the highest operation frequency due to the simplest structure while providing the narrowest locking range. Static dividers reveal a relatively wide range of operation but only for low frequencies. Regenerative dividers, also known as Miller dividers, act as a compromise between the two. Figure 5.21 plots the simulated operation ranges for three dividers targeting 80, 40 and 20 GHz with injection-locked, Miller and static topologies. The three curves can be roughly aligned (as the bold dash line), suggesting a direct tradeoff between the input frequency and the operation range since the product of the two is approximately a constant.

The noise induced by the dividers directly affects the overall signal purity. An easy way to examine how much the overall performance degradation is to check the output spectrum: for an ideal divider chain with total modulus of N , the output

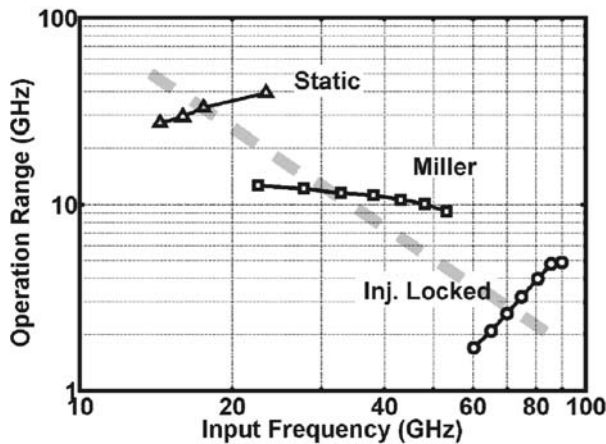


Fig. 5.21 Simulated operation ranges of 3 dividers in 90-nm CMOS.

spectrum should be approximately $20 \log_{10} N$ dB lower than that of the input. Power consumption is another important parameters. In a PLL, for example, the power dissipated by the first few dividers begins to dominate the overall power consumption as the operation frequency approaches 10 GHz. We analyze the three topologies in detail in the following sections.

5.6 Static Dividers

Frequency division of a periodic signal could be achieved in different ways. One of the simplest $\div 2$ realizations is to place an edge-triggered flipflop (composed of two latches) in a negative feedback loop, as illustrated in Fig. 5.22(a). Differentially driven by the input clock, the two latches provide quadrature outputs running at half the input frequency³ [Figure 5.22(b)]. Since the stored information can be held in the latches forever, the static frequency dividers can theoretically operate at arbitrarily low frequencies. Such a simple yet robust configuration manifests itself in low to moderate speed applications.

Although almost any type of latch can be adopted in a static divider, tradeoffs exist among bandwidth, power, robustness and signal integrity. For pure digital implementations such as C^2 MOS or TSPC latches, the stacking of devices as well as the

³ $CK_{out,I}$ and $CK_{out,Q}$ are separated by exactly 90° if the two latches experience the same loading.

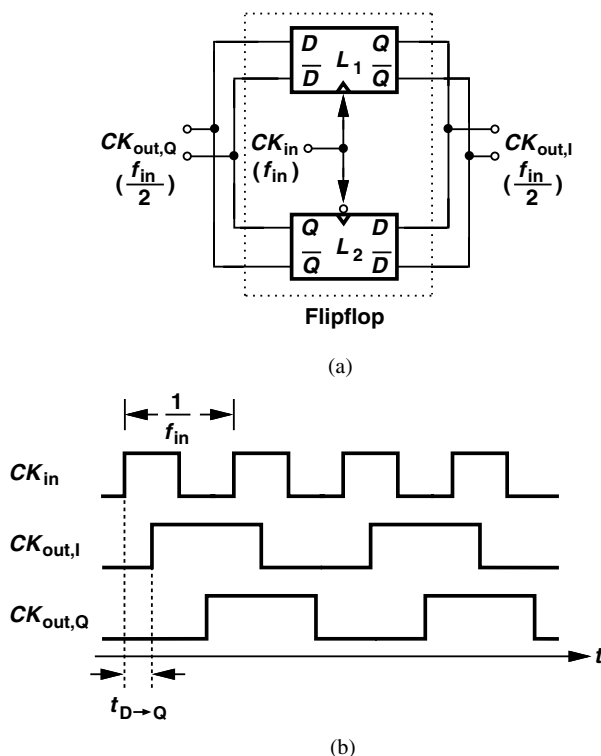


Fig. 5.22 (a) Typical static divider, (b) its waveforms.

rail-to-rail operation lead to long rising and falling times. The single-ended structure also suffers from supply noise coupling, potentially introducing jitter in the output. Meanwhile, the abrupt switching of the circuit would pull significant current from V_{DD} momentarily during transitions, which induces voltage bounce and perturbs the quiet analog region nearby. Even if supplies are separated, unwanted coupling can occur through the substrate or package. A better choice is to use current-mode logic (CML), and Fig. 5.23 depicts two examples for CMOS and bipolar realizations. Controlled by CK_{in} and \overline{CK}_{in} , it samples (amplifies) the input while $M_{1,2}$ ($Q_{1,2}$) pair is activated, and holds (regenerates) the data by means of the cross-coupled pair $M_{3,4}$ (Q_3 - Q_6). The emitter followers (Q_5 and Q_6) in Fig. 5.23(b) serve as level shifters for Q_3 and Q_4 . The constant tail current and differential operation alleviate a number of design issues.

Now let us consider the operation of a static divider with CML latches. At low frequencies, the latches lock the sampled data and wait until the next clock phase comes in. Apparently, the loop gain of the positive feedback (e.g., M_3 - M_4 pair and R_D) must exceed unity, and the output looks like square wave under such a condition. As frequency goes up, the idle time decreases, and the divider would work properly as long as the input pair M_5 - M_6 switches the current completely. Afterwards, the divider

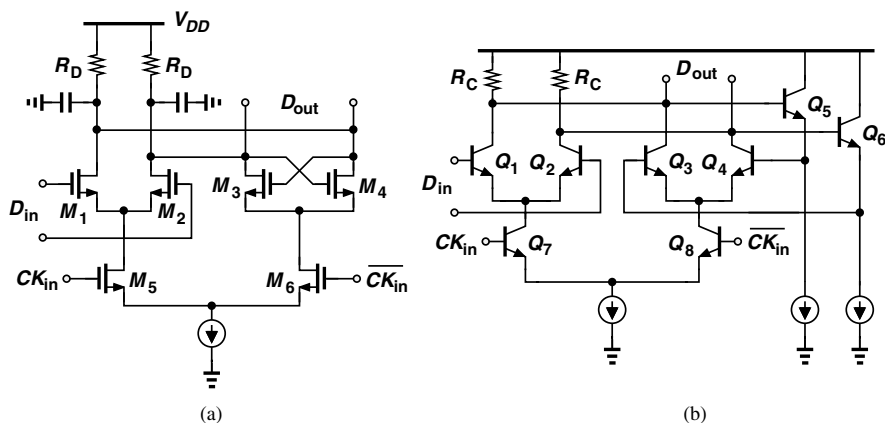


Fig. 5.23 CML latch of (a) CMOS, (b) bipolar technologies.

encounters a self-resonance frequency, where the divider operates as a two-stage ring oscillator. At this moment, the regenerative pairs provide sufficient hysteresis such that each latch contributes 90° of phase shift, and no input power is required. Beyond this frequency, the divider acts as a driven circuit again. It hits a limit as the frequency reaches the bandwidth of the circuit. That is, the D-to-Q delay of the latches ($t_{D \rightarrow Q}$) approaches half input cycle [$1/(2f_{in})$]. As can be clearly explained in Fig. 5.22(b), the timing sequence becomes out of order in such a circumstance, failing the division no matter how large the input power is. Figure 5.24(a) reveals the simulated input sensitivity (i.e., minimum required power) as a function of input frequency of a typical static divider in 90-nm CMOS technology.

It is worth noting that at very low speed, the CML-based static divider may not function properly if the input is sinusoidal. It is because the slow transition of CK_{in} and $\overline{CK_{in}}$ would turn on both latches simultaneously, making the loop transparent for a short period of time. As a result, “racing” phenomenon occurs, and the output toggles rapidly while CK_{in} and $\overline{CK_{in}}$ are transitioning. The simulated waveforms are demonstrated in Fig. 5.24(b). Here we operate the same divider with 20-MHz sinusoidal input, and the unwanted output switching is clearly observed in the inset. Fortunately, the low-speed inputs in most cases have sharp transitions (i.e., square-wave-like) to avoid such an incorrect operation.

It is always possible to extend the bandwidth by introducing inductive peaking or shrinking the size of the cross-coupled pair. However, the price to pay is the degradation of stability at low frequencies and the reduction of overall operation range.

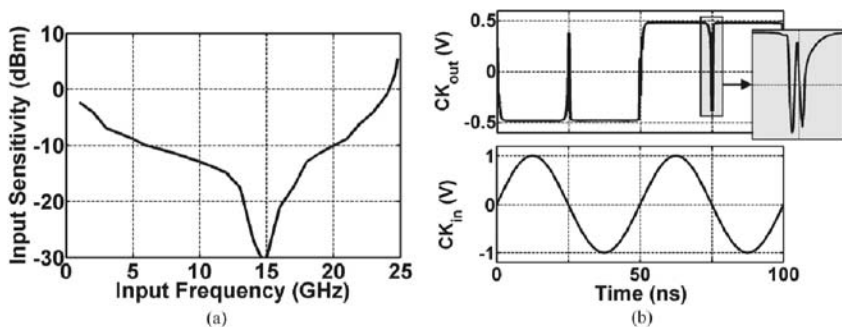


Fig. 5.24 (a) Simulated input sensitivity of a CML static divider in 90-nm CMOS, (b) erroneous switching caused by slow sinusoidal input.

5.7 Regenerative (Miller) Dividers

Originally proposed by Miller in 1939 [21], the regenerative divider is based on mixing the output with the input and applying the result to a low-pass filter (Fig. 5.25). Under proper phase and gain conditions, the component at $\omega_{in}/2$ survives and circulates around the loop, achieving $\div 2$ operation. Such a configuration allows joint design of the mixer and the low-pass filter to arrive at a high speed, since the device capacitance of the former can be absorbed as part of the latter. This topology thus becomes attractive and popular at moderate to high frequencies.

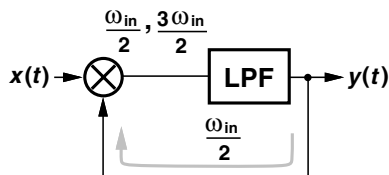


Fig. 5.25 Regenerative divider.

Let us first examine the division behavior and estimate the operation range. For the circuit to divide properly, the loop gain at $\omega_{in}/2$ must exceed unity. Redrawing the divider in Fig. 5.26(a) with simple RC filter and input amplitude A , we have

$$\frac{\beta A}{2} \left| H(j\frac{\omega_{in}}{2}) \right| \geq 1, \quad (5.36)$$

where β denotes the conversion gain of the mixer. It can be further derived that

$$A \geq \frac{2}{\beta} \sqrt{1 + \left(\frac{\omega_{in}}{2\omega_c} \right)^2} \geq \frac{2}{\beta}. \quad (5.37)$$

Here, the corner frequency $\omega_c = (R_1 C_1)^{-1}$. Equation (5.37) implies that a minimum level of at least $2/\beta$ is required for the input. Unlike the static or the injection-locked dividers, the regenerative ones present no self-resonance frequency, resulting in a relatively flat input sensitivity.

Realizing that the LPF is to filter out the component at $3\omega_{in}/2$ and preserve that at $\omega_{in}/2$, we examine two cases to determine the operation range. As illustrated in

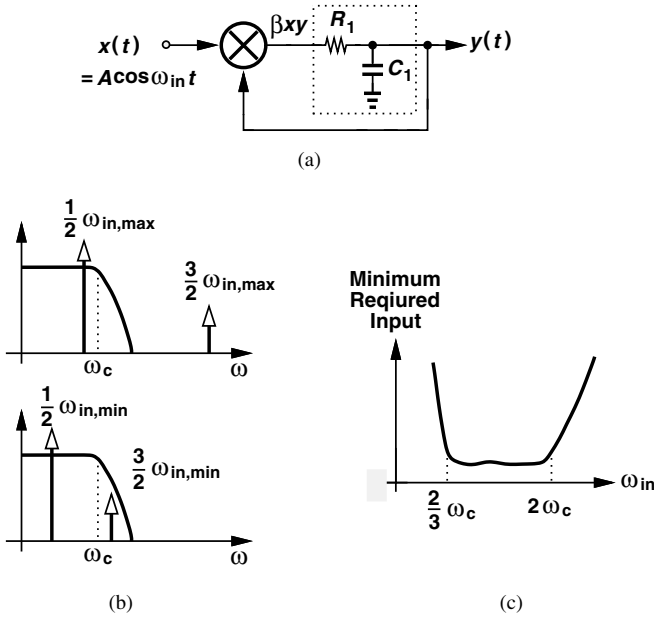


Fig. 5.26 (a) Regenerative divider with an RC filter, (b) operation range determination, (c) typical input sensitivity.

Fig. 5.26(b), the rule of thumb is to keep $\omega_{in}/2$ inside the passband while rejecting $3\omega_{in}/2$ and other harmonics. In other words, we can roughly estimate the operation range as

$$\frac{\omega_{in,max}}{2} \leq \omega_c \quad \text{and} \quad \frac{3\omega_{in,min}}{2} \geq \omega_c, \quad (5.38)$$

and hence

$$\frac{2\omega_c}{3} \leq \omega_{in} \leq 2\omega_c. \quad (5.39)$$

Figure 5.26(c) illustrates the sensitivity of a typical regenerative divider.

While providing an intuitive understanding of the circuit's behavior, the model in Fig. 5.26(a) fails to stipulate the condition for proper division. Neglecting nonlinearities in the mixer, we have

$$R_1 C_1 \frac{dy}{dt} + y = \beta y A \cos \omega_{in} t. \quad (5.40)$$

An explicit solution can be obtained as

$$y(t) = y(0) \exp \left(-\frac{t}{R_1 C_1} + \frac{\beta A}{R_1 C_1 \omega_{in}} \sin \omega_{in} t \right). \quad (5.41)$$

That is, $y(t)$ decays to zero with a time constant of $R_1 C_1$, i.e., the circuit fails to divide regardless of the ratio of ω_{in} and ω_c . Such a contradiction implies that the model in Fig. 5.26(a) is oversimplified. A more accurate model can be obtained by introducing a delay ΔT following the LPF to represent the broadband phase shift around the loop [Fig. 5.27(a)]. Indeed, as shown in Fig. 5.27(b), a typical bipolar

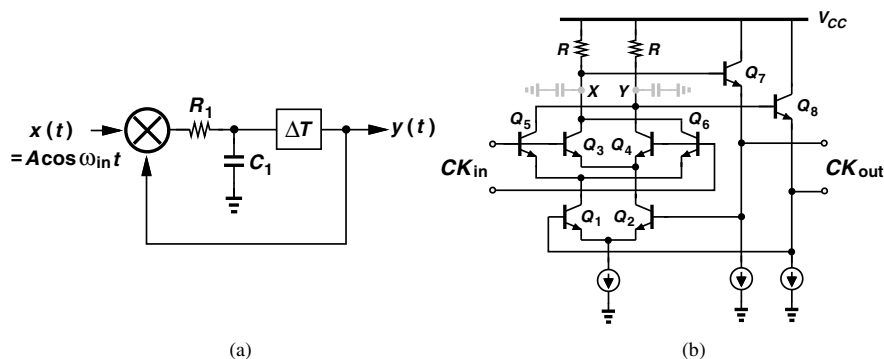


Fig. 5.27 (a) Modified Miller divider model, (b) bipolar realization.

implementation introduces delay at nodes X and Y through the emitter followers and at the collectors of Q_1 and Q_2 . This delay is indispensable and almost all the bipolar Miller dividers belong to this topology.

However, the configuration of Fig. 5.27(b) is difficult to realize in CMOS technologies because the relatively low transconductance of CMOS devices arrives at a source follower with poor performance. It may consume substantial voltage headroom while attenuating the signal and discouraging the divider from high-speed operation. An alternative approach of Miller divider suitable for CMOS devices has been proposed by employing an LC tank (or equivalently, a bandpass filter) as the

load in the mixer [22]. Shown in Fig. 5.28(a), a bandpass filter replaces the low-pass one to suppress the higher-order harmonic. Again, the loop gain at $\omega_{in}/2$ must be greater than unity. Following the same derivation for Eq. (5.37), we obtain the minimum input level necessary for correct division as

$$A \geq \frac{2}{\beta} \sqrt{1 + \frac{\left(1 - \frac{\omega_{in}^2}{4\omega_n^2}\right)^2}{\xi^2 \frac{\omega_{in}^2}{\omega_n^2}}}, \quad (5.42)$$

given that $2\xi\omega_n = (RC)^{-1}$ and $\omega_n^2 = (LC)^{-1}$. For $\Delta\omega = |\omega_{in} - 2\omega_n| \ll 2\omega_n$, we have

$$1 - \frac{\omega_{in}^2}{4\omega_n^2} \approx \frac{\Delta\omega}{\omega_n}. \quad (5.43)$$

Realizing that $\xi = (2Q)^{-1}$, we reduce the fraction under the square root in Eq. (5.42) to $(Q\Delta\omega/\omega_n)^2$ and arrive at

$$A \geq \frac{2}{\beta} \sqrt{1 + \left(\frac{Q\Delta\omega}{\omega_n}\right)^2}. \quad (5.44)$$

Figure 5.28(b) plots a typical input sensitivity as a function of ω_{in} . One CMOS example of such a BPF-based divider is depicted in Fig. 5.28(c).

It is interesting to note that a mixer has two input ports, that leads to two possible configurations of Miller dividers. As illustrated in Fig. 5.29, the output could either return to the RF port (type I) or the LO port (type II) of the mixer. Although conceptually indistinguishable, these two approaches still make difference in circuit implementation. Figure 5.30(a) shows a CMOS Miller divider with the output directly applied to the LO port. The M_3 - M_6 quad of the double-balanced mixer can be redrawn as that in Fig. 5.30(b). It in fact resembles an injection-locked divider (which will be discussed in the next section): M_3 and M_4 form a cross-coupled pair, and M_5 and M_6 appear as diode-connected transistors to lower the Q of the tank and increase the locking range. The differential injection in such a manner is believed to help enlarge the range of operation to some extent. It is possible to find a self-resonance frequency of the circuit if $(W/L)_{3,4} > (W/L)_{5,6}$ [22].

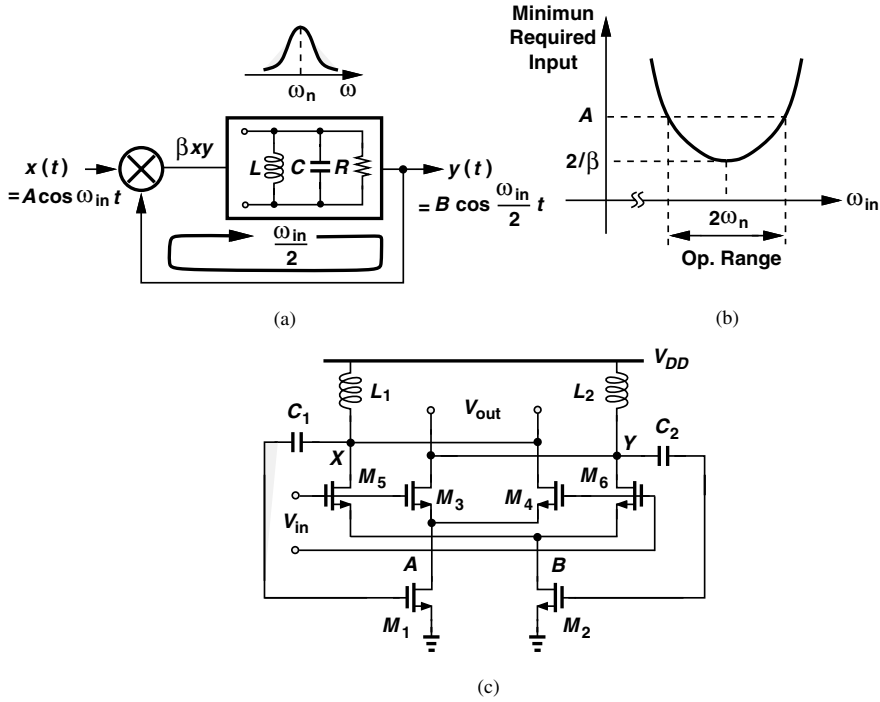


Fig. 5.28 (a) Regenerative divider with bandpass filter; (b) its input sensitivity, (c) typical CMOS realization.

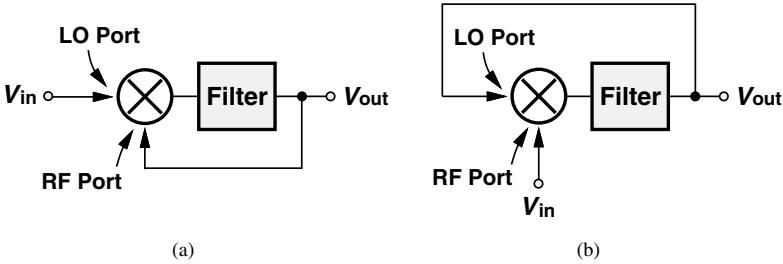


Fig. 5.29 Regenerative divider with the output fed back to (a) RF port, (b) LO port.

5.8 Injection-Locked Dividers

The operation speed of dividers can be further boosted up if we simplify the structure at the circuit level. Since a cross-coupled VCO provides ultimate simplicity in generating differential oscillation, one may think of injecting a periodic signal (approximately twice the VCO free-running frequency) into the common-mode point of it and forcing the VCO to lock. Recognized as an injection-locked divider, this

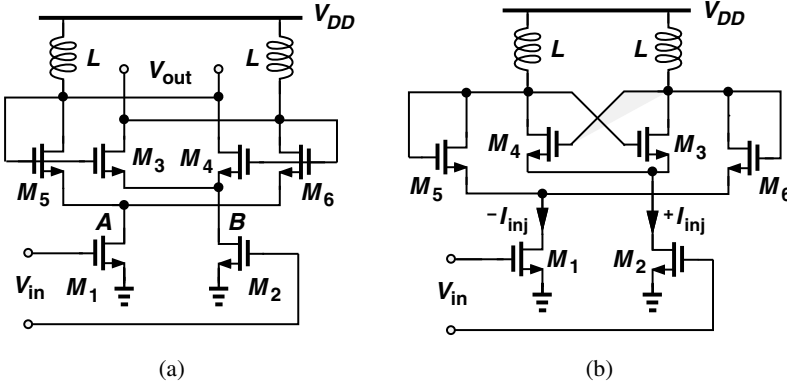


Fig. 5.30 (a) Type II regenerative divider, (b) redrawn to show injection locking.

approach is indeed an inverse operation of push-push oscillators. Among the existing divider topologies, it basically reaches the highest speed.

The injection locking phenomenon can be explained as adding an external sinusoidal current I_{inj} to a well-behaved oscillator [Figure 5.31(a)]. If the amplitude and frequency of I_{inj} are chosen properly, the circuit oscillates at the injection frequency of ω_{inj} rather than the tank resonance frequency ω_0 . The key point here is that, to accommodate the phase shift contributed by the tank at ω_{inj} , I_{osc} (the intrinsic oscillation current) and I_{inj} must sustain a certain phase difference such that the total phase shift maintains 0° . Intuitively, the injection locking would occur only in the vicinity of ω_0 and the locking range is limited. In fact, it can be analytically derived from different approaches [23][24] that the normalized locking range is given by

$$\frac{\Delta\omega}{\omega_0} = \frac{1}{Q} \cdot \frac{I_{inj}}{I_{osc}} \cdot \frac{1}{\sqrt{1 - \frac{I_{inj}^2}{I_{osc}^2}}}, \quad (5.45)$$

where Q denotes the quality factor of the tank. It degenerates to a simple form as $I_{inj} \ll I_{osc}$:

$$\frac{\Delta\omega}{\omega_0} \approx \frac{1}{Q} \cdot \frac{I_{inj}}{I_{osc}}. \quad (5.46)$$

The injection locking technique can be easily applied to dividers. Figure 5.31(b) shows such a circuit with the injection input I_{inj} of approximately twice the tank resonance frequency. The cross-coupled pair M_1 - M_2 can be considered as a mixer down-converting ω_{inj} into $\omega_{inj} - \omega_0$. With abrupt switching,⁴ it is equivalent to injecting a current of $I_{inj} \cdot (2/\pi)$ at $\omega_{inj} - \omega_0$ into the LC tank. Since $\omega_{inj} - \omega_0 \approx \omega_0$,

⁴ It is expectable if the tail current and the inductors create enough swing at V_{out} .

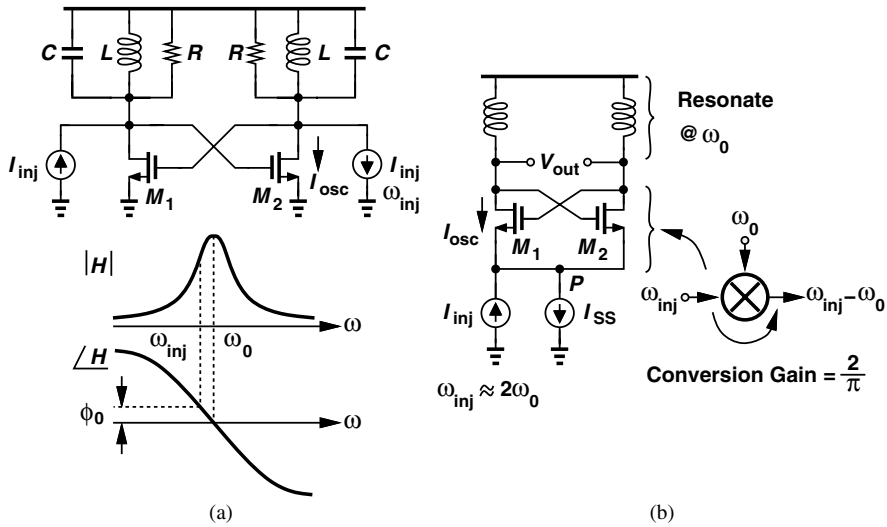


Fig. 5.31 (a) Operation of injection locking, (b) injection-locked divider.

we achieve the $\div 2$ operation with the following locking range

$$\frac{\Delta\omega}{\omega_0} = \frac{1}{Q} \cdot \frac{2}{\pi} \cdot \frac{I_{inj}}{I_{osc}}. \quad (5.47)$$

Note that Eqs. (5.45)-(5.47) define the *relative* locking range, which makes no difference whether it is observed from the input or the output.

A few modifications can be made to improve the performance of the divider in Fig. 5.31(b). One issue of the circuit in Fig. 5.31(b) stems from the parasitic capacitance associated with node P . At high speed, it creates a path to ground, robbing significant portion of I_{inj} and undermining the injection. To modify it, an inductor L can be added to resonate out the capacitance C_P [Fig. 5.32(a)], enlarging lock range without extra power consumption [25]. Other than the parasitic, the circuit in Fig. 5.31(b) is driven single-endedly, wasting 50% of the injection power. Another topology called “direct injection” is shown in Fig. 5.32(b) [26]. Here, the signal injection is accomplished by driving the two switches M_5 and M_6 differentially, which are sitting across the two outputs of the oscillator made of M_1 - M_4 and L . Note that M_5 and M_6 are turned on and off almost simultaneously. Here, the input signals still drive the common-mode points, i.e., gates of M_5 and M_6 . With proper design and biasing, the quasi-differential operation is expected to achieve a wider locking range.

The injection locking technique can be also utilized to implement dividers with modulus other than 2. Figure 5.33(a) reveals a possible realization of $\div 3$ circuit [27]. Here, transistors M_1 - M_3 form a ring oscillator, and the input signal (approximately 3 times of the ring oscillation frequency) is injected into the “common-mode” point

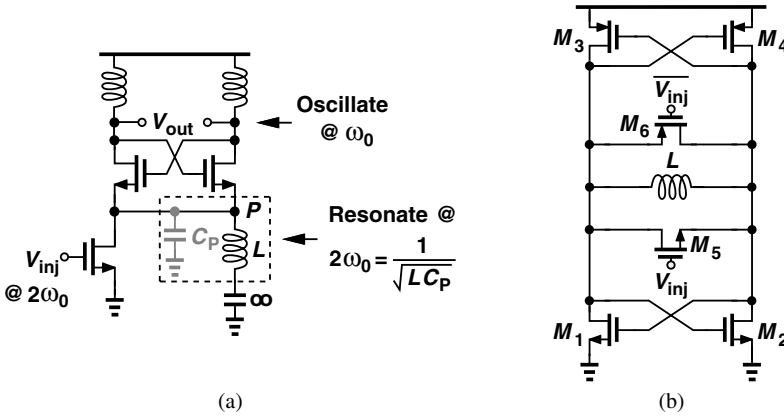


Fig. 5.32 Modified injection-locked dividers with (a) shunt peaking inductor, (b) direct injection.

by means of M_4 . Again with proper design, the ring would lock to one-third of the input frequency. Figure 5.33(b) shows another divider example that performs $\div 4$ operation [28]. The circuit is nothing more than a direct injection divider but with M_3 functioning as a 3rd-order harmonic mixer. That is, under proper biasing, the 3rd-order harmonic of M_3 becomes comparable with the fundamental component [Fig. 5.33(b)]. In other words, the circuit mixes the input ($\approx 4\omega_0$) with the 3rd-order harmonic of the output, while the LC tank provides bandpass filtering at ω_0 . Such an arrangement reaches twofold power efficiency as regular $\div 2$ circuits because the modulus is doubled.

The narrow locking range of injection-locked dividers usually necessitates careful design, skillful layout, as well as meticulous EM simulations. It is especially true at high speed since the deviation of natural frequency caused by PVT variations may destroy the locking. Any tuning technique intended to dynamically adjust the locking range would prove futile because the varactor's overhead may have degraded the natural frequency considerably.

5.9 Case Study

To reinforce the design ideas described above, we analyze three mm-wave PLLs that can be incorporated in future wireless systems. Emphasizing the VCO and dividers, we present the circuit details and measurement results of these works in ascending order of frequency and sophistication.

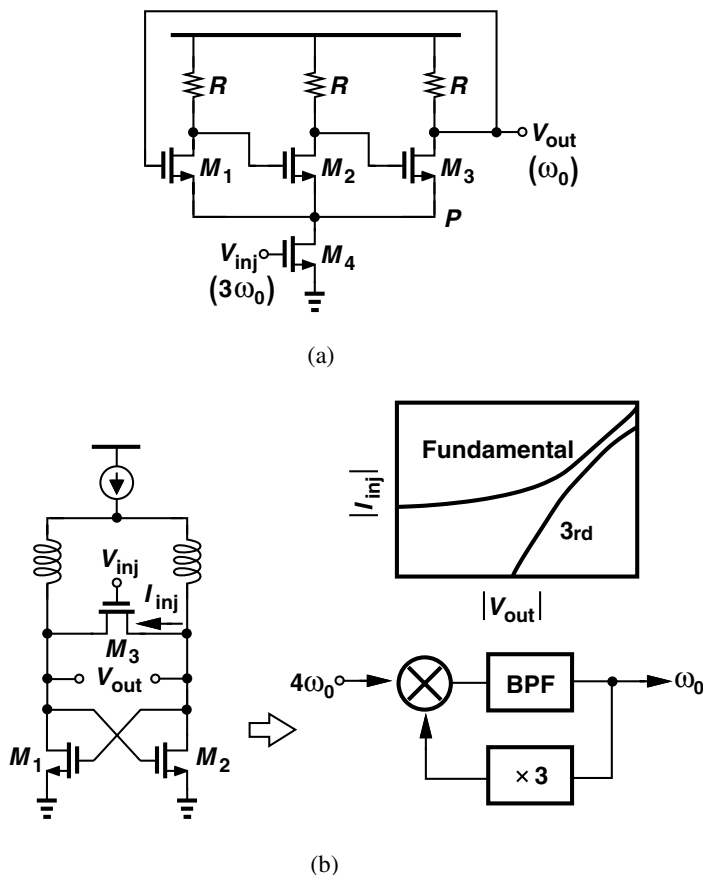


Fig. 5.33 Higher modulus dividers utilizing injection-locking technique, (a) $\div 3$, (b) $\div 4$.

5.9.1 52-GHz LO Signal Generator

We first see the design of a 52-GHz PLL in a wireless transceiver [29]. It contains an on-chip VCO, an injection-locked divider as the first stage, a subsequent divider chain of modulus 512, an off-chip phase and frequency detectors, and a loop filter. The VCO is shown in Fig. 5.34(a), where the cross-coupled topology with ac coupling achieves a 10% tuning range around 52 GHz. In order to generate quadrature clocks, the first $\div 2$ circuit [Fig. 5.34(b)] is implemented as two identical injection-locked dividers coupling to each other. The transmission line and the varactor reveal a quality factor of 24 and 40, respectively. Measurement verifies that the VCO and its buffers consume 25 mW, and the divider presents a locking range of 3.1 GHz while drawing 3.1 mA from a 2.5-V supply. The circuit is implemented in SiGe BiCMOS process with an f_T of 200 GHz. Figure 5.35 depicts the VCO tuning curve, phase noise performance, and divider sensitivity. The phase noise at 1-MHz offset is -95 dBc/Hz.

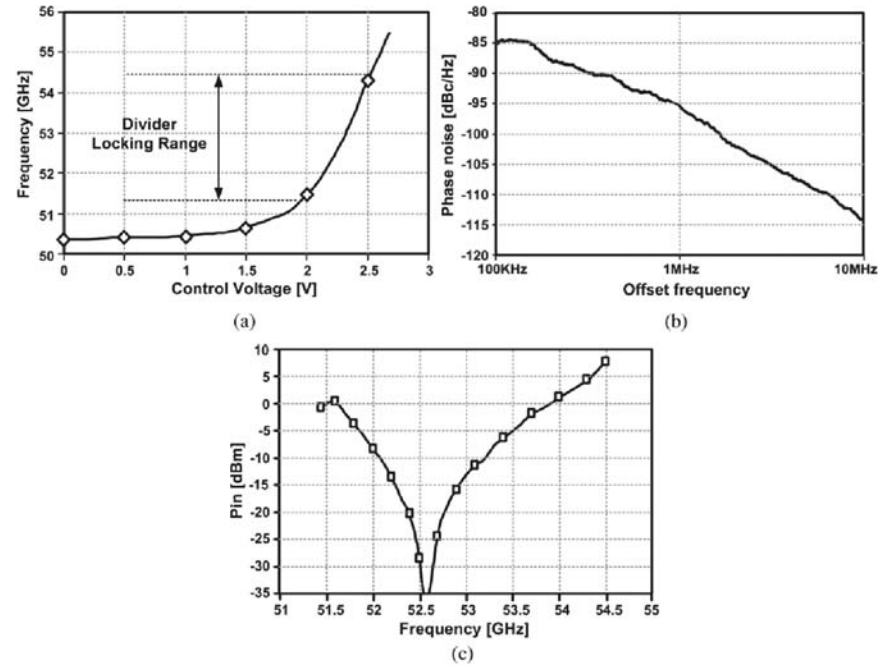


Fig. 5.35 (a)VCO tuning range, (b)VCO phase noise, (c)divider sensitivity of [29] (©IEEE 2006).

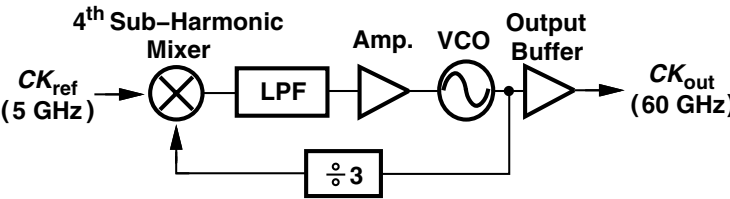
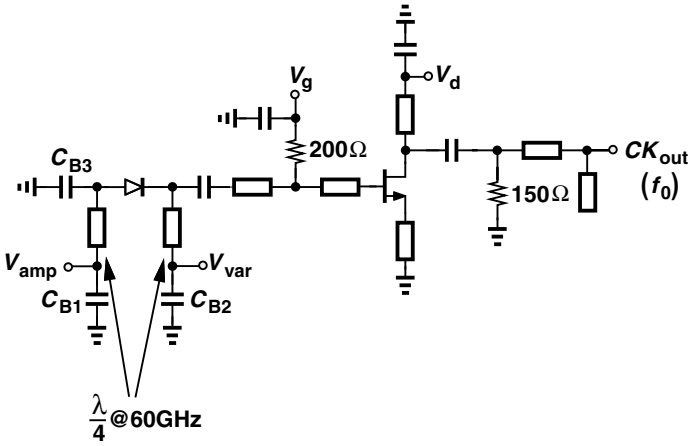
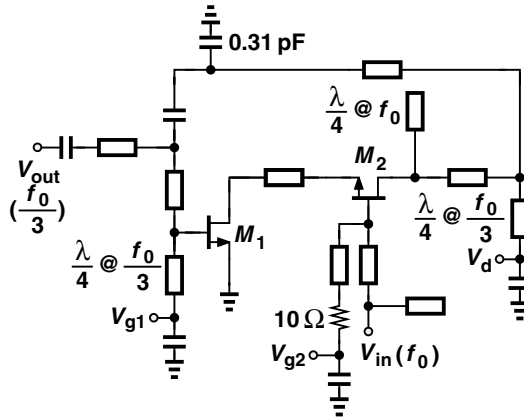


Fig. 5.36 PLL architecture of [30].

The $\div 3$ divider is designed based on a cascode injection locking topology. As shown in Fig. 5.37(b), it consists of cascode FETs and phase-delay lines tending to resonate at $1/3$ of the input frequency. The inter-modulation terms are generated through nonlinear multiplication, which along with the feedback loop arrives at $\div 3$ operation. Two gate biases V_{g1} and V_{g2} are adjusted externally to optimize the performance. The output is coupled out of the circuit at a low-impedance point in the feedback loop. The feedback line is realized with impedance matching at both ends to maintain sufficient loop gain at $f_0/3$. The matching circuit at the gate of M_2 is implemented as the lowest impedance at $f_0/3$ for the same reason. The quarter wavelength open stub at f_0 is placed at the drain of M_2 to provide short circuit to



(a)



(b)

Fig. 5.37 (a)VCO, (b) $\div 3$ circuit in [30].

the input signal. These stubs also act as band selecting filters, reducing the unwanted harmonics and spurs.

The VCO achieves a tuning range of 2.2 GHz and a phase noise of -87.7 dBc/Hz at 1-MHz offset. With $V_g = -0.4$ V and $V_d = 1.5$ V, the dc current of the VCO equals 28 mA. The $\div 3$ circuit reveals a bandwidth of 2.3 GHz around 60.9 GHz with an input power of 1 dBm. The power consumption is 7 mW. Figure 5.38 shows the VCO tuning curve and divider locking bandwidth.

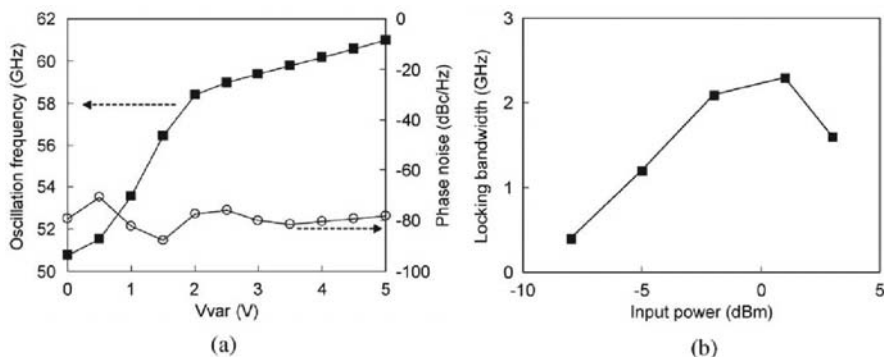


Fig. 5.38 Performance of [30]: (a) VCO tuning range and phase noise, (b) locking range of $\div 3$ circuit (©IEEE 2006).

5.9.3 A 75-GHz PLL in 90-nm CMOS

In the last case we study a fully-integrated CMOS design recently published [33]. Figure 5.39(a) shows the PLL architecture. It consists of a differential VCO running at full rate, a divider chain with total modulus of 64, a phase and frequency detector, and a third-order loop filter. Different divider topologies: injection-locked, Miller, and static, are employed and placed in descendant order of frequency to accommodate the severe tradeoffs between the input frequency and operation range.

Figure 5.39(b) depicts the VCO and the first divider stage. Transmission lines equivalent to $3/4$ wavelength of a 75-GHz clock is introduced here to distribute the capacitive loading and boost the oscillation frequency. Having one end short-circuited and the other open-circuited, these lines resonate differentially with the cross-coupled pair M_1-M_2 providing negative resistance. Connecting to the $1/3$ points of the lines (nodes A and A'), this pair forces the transmission lines to create peak swings at these nodes. The waves thus propagate and reflect along the lines, forming the second maximum swings with opposite polarities at nodes B and B' . That is, node A (A') and node B (B') are 180° out of phase. As a result, the buffers (M_3-M_4), dividers (M_5-M_6), and varactors (M_7-M_8) can be removed to these ends, making the two zenith positions bear approximately equal capacitance. This arrangement absorbs the loading into the transmission lines and raises the oscillation frequency. To achieve high Q and compact layout, the transmission lines are actually realized as three identical inductors in series. The natural biasing established by M_1-M_2 pair facilitates dc coupling between the VCO and subsequent blocks. Careful layout arrives at perfect symmetry between the loadings at nodes B and B' . Inductor L_R is added to resonate out the parasitic capacitance associated with nodes C and C' , allowing stronger signal injection through M_5 and M_6 .

The VCO is biased with a supply-independent circuit M_9-M_{12} and R_S . To further reject the supply noise, M_{13} is introduced to absorb extra current variation caused by channel-length modulation. That is, $|\partial I_{SS} / \partial V_{DD}| = |\partial I_C / \partial V_{DD}|$ and the

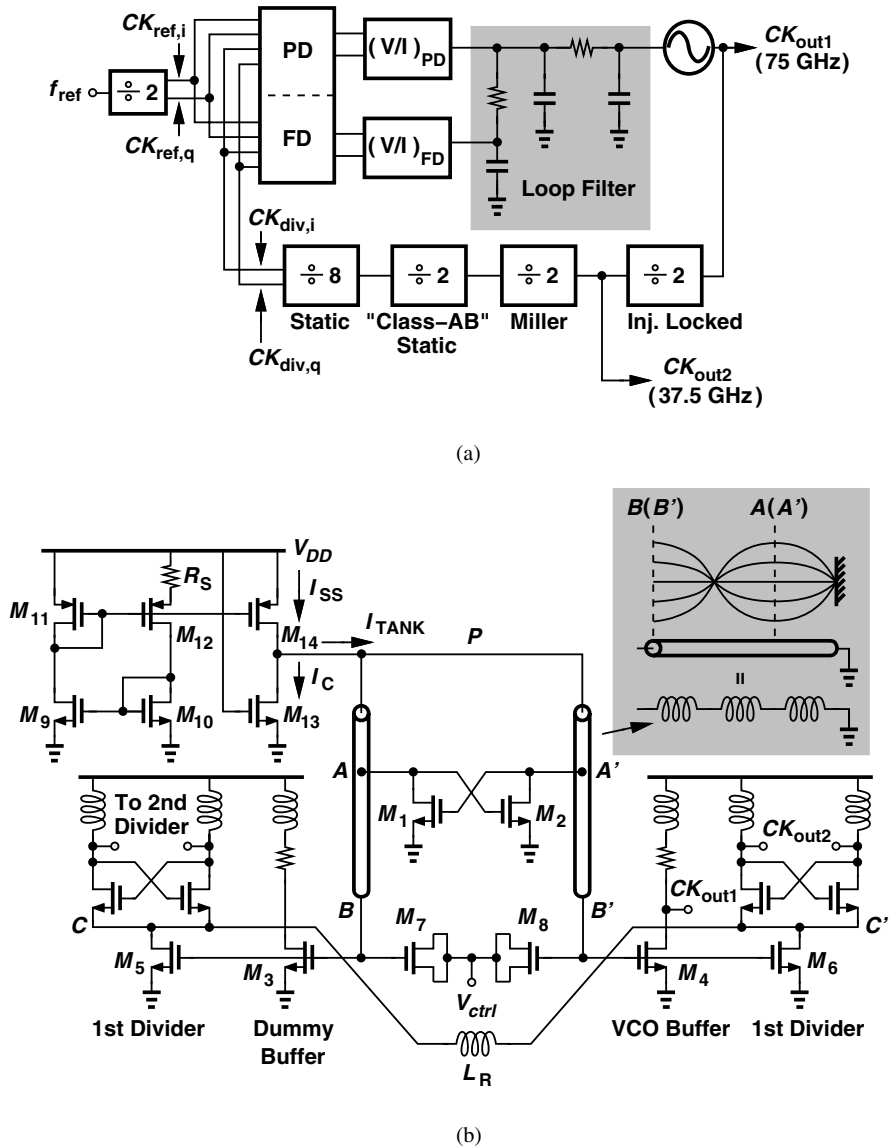


Fig. 5.39 (a) PLL architecture, (b) VCO and first divider of [33].

current flowing into M_1 - M_2 pair (I_{TANK}) remains constant. It leads to a fixed voltage at node P , leaving the resonance frequency insensitive to supply perturbation.

The PLL has been fabricated in 90-nm CMOS technology. The total power consumption with a 1.45-V supply is 88 mW, of which 8 mW is dissipated in the VCO, 66 mW in the divider chain, and 14 mW in the PFD and V/I converters. Figure 5.40 depicts the output spectra of the VCO and the first divider under locked condition.

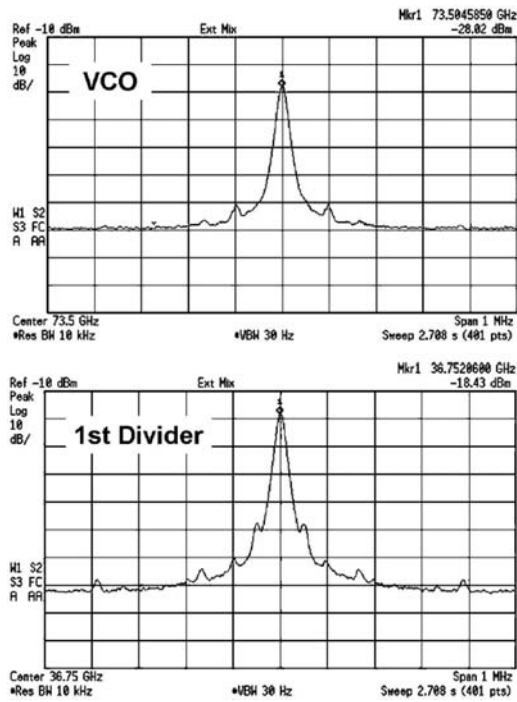
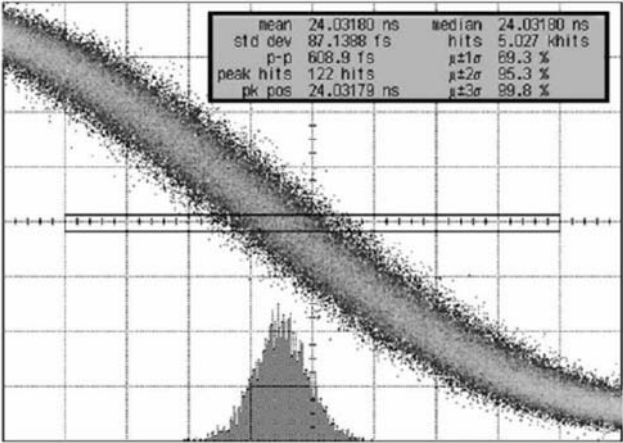
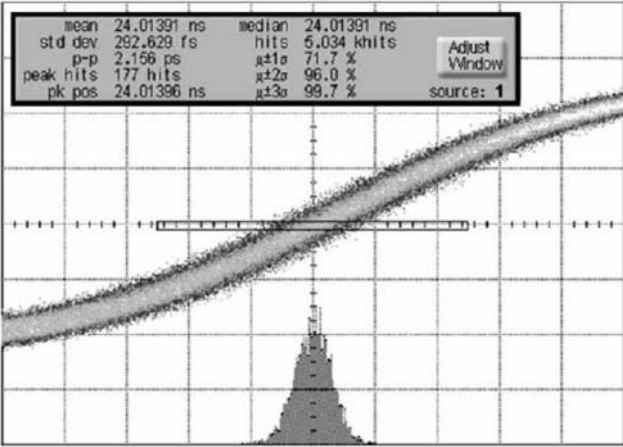


Fig. 5.40 Output spectra of VCO and first divider in [33].

The phase noise at 100-kHz offset measures -88 and -94 dBc/Hz, respectively. The time domain performance are recorded as shown in Fig. 5.40. The 75-GHz output presents peak-to-peak and rms jitter of 609 fs,pp and 87 fs,rms, respectively, whereas the 37.5-GHz output reveals jitter of 2.15 ps,pp and 293 fs,rms.



(a)



(b)

Fig. 5.41 Waveforms in [33] for (a) 75-GHz, (b) 37.5-GHz outputs.

References

1. D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," *Proc. IEEE*, vol.54, pp.329-330, Feb. 1966.
2. H. Wang *et al.*, "A 50 GHz VCO in 0,25 μ m CMOS," *ISSCC Dig. of Tech. Papers*, pp. 372-373, Feb. 2001.
3. C. Cao *et al.*, "192 GHz push-push VCO in 0.13 μ m CMOS," *Electron. Lett.*, vol. 42, pp. 208-210, Feb. 2006.
4. C. Cao *et al.*, "A 140-GHz Fundamental Mode Voltage-Controlled Oscillator in 90-nm CMOS Technology," *Microwave and Wireless Components. Lett.*, vol. 16, pp.555-557, Oct. 2006.
5. M. Danesh *et al.*, "A Q-factor ehancement technique for MMIC inductors," *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium Dig. Paper*, pp.217-220, June 1998.
6. A. Zolfaghari *et al.*, "Stacked inductors and transformers in CMOS technology," *IEEE J. Solid-State Circuits*, vol. 36, no. 4, pp. 620-628, Apr. 2001.
7. J. Lee, "High-speed circuit designs for transmitters in broadband data links," *IEEE J. Solid-State Circuits*, vol. 41, no. 5, pp. 1004-1015, May 2006.
8. E. Hegavi *et al.*, "A Filtering Technique to Lower Oscillator Phase Noise," *ISSCC Dig. of Tech. Papers*, pp. 364-365, Feb. 2001.
9. J. Chien *et al.*, "A 40-GHz Wide-Tuning-Range VCO in 0.18- μ m CMOS," *VLSI Dig. of Tech. Papers*, pp. 178-179, 2006.
10. E. H. Colpitts *et al.*, "Carrier current telephony and telegraphy," *Journal AIEE.*, vol. 40, no.4, pp. 301-305, Apr. 1921.
11. B. Razavi, "Design of Integrated Circuits for Optical Communications," New York: McGraw-Hill, 2002.
12. W. Winkler *et al.*, "60 GHz transceiver circuits in SiGe:C BiCMOS technology," *Proc. of European Solid-State Circuits Conf.*, pp. 83-86, Sep. 2004.
13. B.A. Floyd *et al.*, "SiGe Bipolar transceiver circuits operating at 60 GHz," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 156-167, Jan. 2005.
14. S.T. Nicolson *et al.*, "Design and scaling of SiGe BiCMOS VCOs above 100GHz," *Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 1-4, Oct. 2006.
15. B. Heydari *et al.*, "Low-Power mm-Wave Components up to 104GHz in 90nm CMOS," *ISSCC Dig. of Tech. Papers*, pp. 200-201, Feb. 2007.
16. D. Pozar, "Microwave Engineering," John Wiley & Sons, Inc., 1998.
17. P. Huang *et al.*, "A low-power 114-GHz push-push CMOS VCO using LC source degeneration," *IEEE J. Solid-State Circuits*, vol. 42, no. 6, pp. 1230-1239, June 2007.
18. R. Wanner *et al.*, "SiGe integrated mm-wave push-push VCOs with reduced power consumption," *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium Dig. Paper*, pp. 483-486, June. 2006.
19. H. Wu and A. Hajimiri, "Silicon-based distributed voltage-controlled oscillators," *IEEE J. Solid-State Circuits*, vol. 36, no. 3, pp. 493-502, Mar. 2001.
20. N. Tzartzanis *et al.*, "A reversible poly-phase distributed VCO," *ISSCC Dig. of Tech. Papers*, pp. 2452- 2461, Feb. 2006.
21. R. L. Miller, "Fractional-frequency generators utilizing regenerative modulation," *Proc. Inst. Radio Eng.*, vol. 27, pp. 446-456, Jul. 1939.
22. J. Lee and B. Razavi, "A 40-GHz frequency divider in 0.18- μ m CMOS technology," *IEEE J. Solid-State Circuits*, vol. 39, no. 4, pp. 594-601, Apr. 2004.
23. R. Adler, "A study of locking phenomena in oscillators," *Proc. IEEE*, vol. 61, no. 10, pp. 1380-1385, Oct. 1973.
24. B. Razavi, "A study of injection locking and pulling in oscillators," *IEEE J. Solid-State Circuits*, vol. 39, no. 9, pp. 1415-1424, Sep. 2004.
25. H. Wu and A. Hajimiri, "A 19GHz 0.5 μ m CMOS frequency divider with shunt-peaking locking-range enhancement," *ISSCC Dig. of Tech. Papers*, pp. 412-413, Feb. 2001.
26. M. Tiebout, "A CMOS direct injection-locked oscillator topology as high-frequency low-power frequency divider," *IEEE J. Solid-State Circuits*, vol. 39, no. 7, pp. 1170-1174, Jul. 2004.

27. S. Verma *et al.*, "A multiply-by-3 coupled-ring oscillator for low-power frequency synthesis," *IEEE J. Solid-State Circuits*, vol. 39, no. 4, pp. 709-713, Apr. 2004.
28. K. Yamamoto *et al.*, "70GHz CMOS harmonic injection-locked divider," *ISSCC Dig. of Tech. Papers*, pp. 2472-2481, Feb. 2006.
29. A. Natarajan *et al.*, "A 77-GHz Phased-Array Transceiver With On-Chip Antennas in Silicon: Transmitter and Local LO-Path Phase Shifting," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2807-2819, Dec. 2006.
30. J. Jinho *et al.*, "A Fully Integrated V-band PLL MMIC Using 0.15- μ m GaAs pHEMT technology," *IEEE J. Solid-State Circuits*, vol. 41, no. 5, pp. 1042-1050, May 2006.
31. P. J. Howes *et al.*, "Ka-band and MMIC pHEMT-based VCOs with low phase-noise properties," *IEEE Trans. Microw Theory Tech.*, vol. 46, no. 10, pp. 1531-1536, Oct. 1998.
32. B. Piernas *et al.*, "A compact and low-phase-noise Ka-Band pHEMT-based VCOs," *IEEE Trans. Microw Theory Tech.*, vol. 51, no. 3, pp. 778-783, Mar. 2003.
33. J. Lee, "A 75GHz PLL in 90 nm CMOS," *ISSCC Dig. of Tech. Papers*, pp. 432-433, Feb. 2007.

Chapter 6

Power Amplifiers at 60GHz and Beyond

Ehsan Afshari and Abbas Komijani

6.1 Motivation and Challenges

Recently, there has been growing interest in using silicon-based integrated circuits at high microwave and millimeter wave frequencies. The high level of integration offered by silicon enables numerous new topologies and architectures for low-cost reliable SoC applications at microwave and millimeter wave bands, such as broadband wireless access (*e.g.*, WiMax) [1], vehicular radars at 24GHz and 77GHz [2][3], short range communications at 24GHz and 60GHz [4][5][6], and ultra narrow pulse generation for UWB radar [7].

On a silicon substrate, power generation and amplification is one of the major challenges at millimeter wave frequencies. The two most important issues are the low unity power gain frequency, f_{max} , of the transistors, and the loss of on-chip passive elements, such as inductors and transmission lines, required for impedance matching. For narrowband amplifiers, where device capacitance is normally tuned out, f_{max} is a better metric for device speed than f_T . In MOSFETs, f_{max} is limited primarily by the series gate resistance [8][9]. Generally, MOS transistors have lower f_T and f_{max} as compared to SiGe bipolar transistors fabricated with the same feature size [10]. In a typical silicon process, the f_{max} of an NMOS transistors, with an optimum layout, is almost a factor of two smaller than the f_{max} of their SiGe bipolar counterparts [11].

Lossy on-chip passive components present another barrier to the full integration of a high-frequency power amplifier. Skin effect results in larger ohmic losses in inductors and transmission lines at high frequencies. Even at 24GHz, the skin depth in Aluminum is $0.5\mu\text{m}$, which negates some of the advantages of a thick top metal layer, though the lateral sidewalls still help in reducing loss. Although Copper has

Ehsan Afshari

Cornell University, 406 Phillips Hall, Ithaca, NY 14853, e-mail: afshari@cornell.edu

Abbas Komijani

Entropic Communications, San Jose, CA 95131 e-mail: abbas.komijani@entropic.com

better conductivity, in practice its performance can be additionally degraded by the cheese and fill rules necessary for stress relief during fabrication. As an example the cheese rule can increase the sheet resistance of a thick copper trace by a factor of two. Due to the relatively high conductivity of the substrate in most CMOS processes, the inductors and coplanar waveguide transmission line structures have substrate-induced losses as well. The combination of low active gain at high frequencies and high loss in impedance matching networks reduces the power gain of a single-stage amplifier. As a result, it becomes necessary to either cascade an impractically large number of amplifier stages to achieve desired output power levels or combine the output of many smaller amplifiers.

This chapter looks at the design and analysis of millimeter wave power amplifiers. We begin with reviewing passive elements that can lower the substrate loss and reduce on-chip wavelength. Next, we study the power transistors and their characteristics. We then review high efficiency power combining techniques and conclude the chapter with a few recent examples of power amplifiers operating at 60GHz and beyond.

6.2 Passive Components

As we saw, lossy on-chip passive components present a challenge toward the realization of the high-frequency power amplifier on a silicon substrate. In this section, we review a few passive structures that can be used to address this challenge.

6.2.1 *Substrate-Shielded Coplanar Waveguide Structure*

At 60GHz, large capacitive coupling to substrate lowers the quality factor of the inductors, making inductor-based impedance matching networks lossy. On the other hand, this frequency is not high enough for direct application of standard transmission line structures. For example, for silicon dioxide as dielectric, the wavelength (λ) at 60GHz is 2.5mm. Therefore, the transmission lines required for on-chip matching networks will have high loss because of their long length.

As shown in Figure 6.1(a), in coplanar waveguide (CPW) structures designed in CMOS processes with relatively high substrate conductivity ($\sim 10\Omega\cdot\text{cm}$), capacitive coupling to the substrate is often the dominant source of high-frequency loss [12]. On the other hand, in the on-chip microstrip structure, shown in Figure 6.1(b), substrate-induced losses are minimal due to the shielding effect of ground plane. However, the close proximity of the ground plane to the signal line results in a narrow signal line for practical impedance levels. This constraint increases ohmic losses in the signal line. Figure 6.1(c) shows the substrate-shielded coplanar structure that is a combination of the two structures. Slotting the bottom plate forces the return current to be mostly concentrated in the coplanar ground lines. The large separation between signal and return currents causes more magnetic energy to be stored in space, resulting

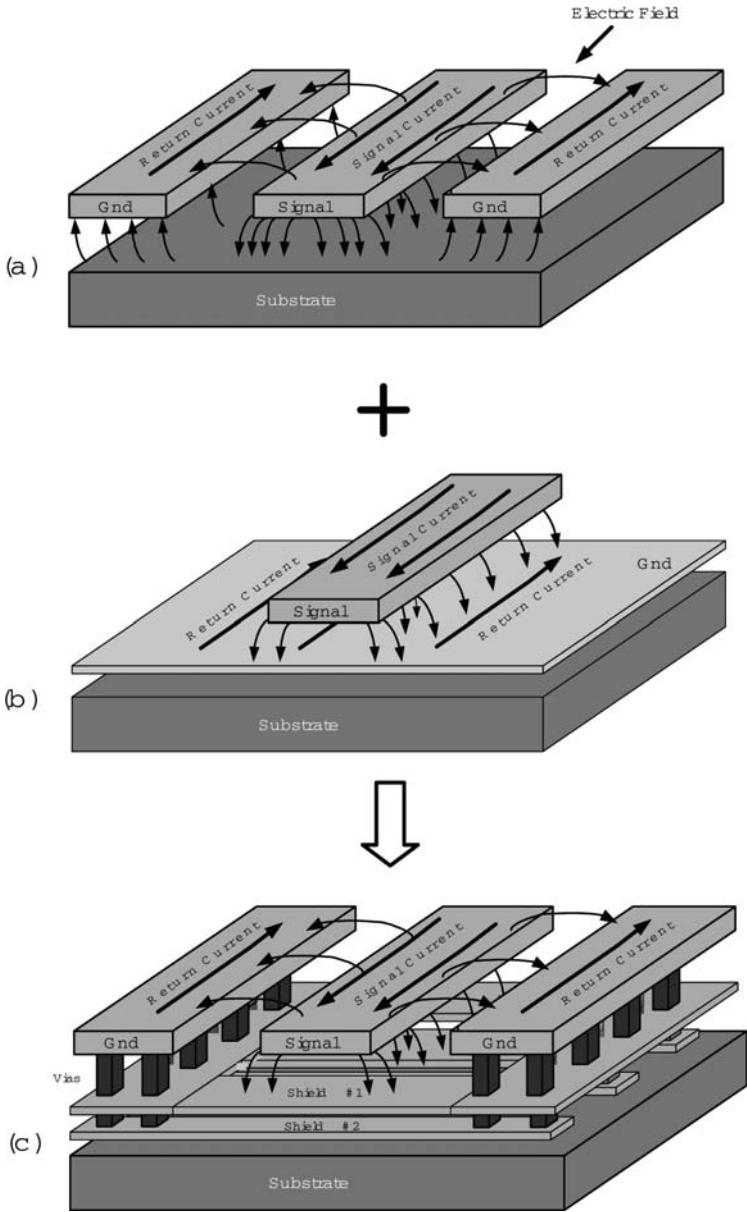


Fig. 6.1 Combination of (a) CPW and (b) Microstrip structures to realize (c) substrate-shielded CPW structure.

in a larger distributed inductance per unit length, L . However, the proximity of the slotted ground line to the signal line results in high capacitance per unit length, C . Simultaneously high values of L and C , lead to slower wave velocity, ($v = 1/\sqrt{LC}$), and hence shorter wavelengths.

Another way to look at this structure is to view it as a CPW structure with periodic capacitive loading. This is similar to the inductive loading concept in [13]. A similar capacitive loading concept is presented in [14]. However, in this case extra capacitance is added by placing the patterned ground beneath the coplanar structure. Therefore, the capacitance per unit length of the structure is increased, thereby slowing down the wave.

In this structure, the velocity is reduced by more than a factor of two, and as a result the wavelength at 60GHz in this structure is reduced to 1.2mm. Furthermore, as opposed to a microstrip structure, reasonable impedance levels are achieved for large signal line widths of $60\mu\text{m}$ thereby decreasing ohmic losses. The combination of lower ohmic losses and the shorter length of the transmission lines leads to a much lower passive loss in the matching networks. As the MOS transistor gain at high frequencies is low, this reduction in passive loss is critical to achieving desired gain and output power.

As shown in Figure 6.1(c), in the substrate-shielded coplanar waveguide structure the two coplanar ground lines are forced to the same potential with vias to the patterned shield. This acts as an air bridge, allowing only one fundamental TEM mode to propagate. Also, a second shield layer is placed beneath the first shield layer, with metal stripes covering slots of the first layer, thereby completely isolating the coplanar structure from the substrate [15].

6.2.2 Characterization of the Substrate-Shielded CPW Structure

The simulated electric and magnetic fields of a cross-section of the substrate-shielded CPW structure with and without slotted shields is shown in Figure 6.2. In the shielded structure the electric fields do not penetrate into the substrate, reducing capacitively-coupled substrate losses. Though the penetration of the magnetic field into the substrate is not affected by the presence of the shield, EM simulations indicate that this does not contribute significantly to the loss as the eddy currents are limited.

6.2.2.1 Characterization

To characterize the substrate-shielded CPW structure, a separate test structure was fabricated in the same process as the amplifier in [15]. The test structure was designed for a characteristic impedance of 27.5Ω , the same impedance used for impedance-matching in the power amplifier. This choice of low characteristic impedance minimizes passive losses in the impedance transformation network. Figure 6.3 shows the die photo of this test structure.

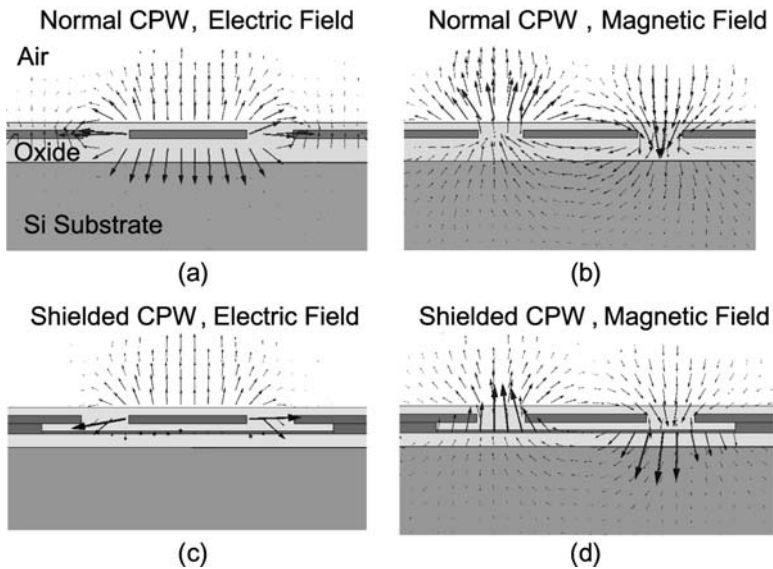


Fig. 6.2 Electric and Magnetic field distributions from 3D EM simulations of (a), (b) a normal CPW structure, and (c), (d) a substrate-shielded CPW structure.

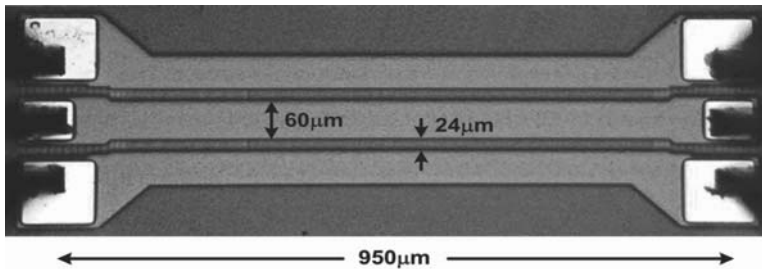


Fig. 6.3 Die photo of the substrate-shielded CPW test structure; shield layer consists of 4 μm -wide stripes with 2 μm spacing.

The top three metal layers were used for the transmission line structure. The top metal layer is 4 μm -thick Aluminum and is located 11.7 μm above the substrate. The two shield layers use 1.25 μm Aluminum and 0.3 μm Copper metal layers placed 5.3 μm and 9.5 μm beneath the bottom of the top metal. 3D electromagnetic simulations with HFSS were performed to accurately simulate a short length of the line, while quasi-planar electromagnetic simulations with IE3D were used as a faster approach to simulate T-junctions and discontinuities [16][17].

Table 6.1 Simulated and measured parameters of the transmission line at 24GHz with wideband fitting.

Parameter	EM Simulation	Measurement
Attenuation constant	0.5dB/mm	1dB/mm
Characteristic impedance	27.5Ω	27.5Ω
Effective relative permittivity	18.7	18

The S-parameters of the line measured in a 50Ω environment are shown in Figure 6.4. A Short-Open-Line-Thru (SOLT) calibration was performed up to the probe tips. A wideband model of the transmission line with parameters shown in Table 6.1 was fitted to the measurement results. Compared to a single-frequency parameter fit in [18], this is a more physical interpretation of the measured data and is less susceptible to measurement errors at a single frequency. To accommodate the skin effect, the loss of the transmission line, in dB, was assumed to be proportional to the square-root of the frequency [19]. As shown in Figure 6.4, this will result in a wideband curve fit.

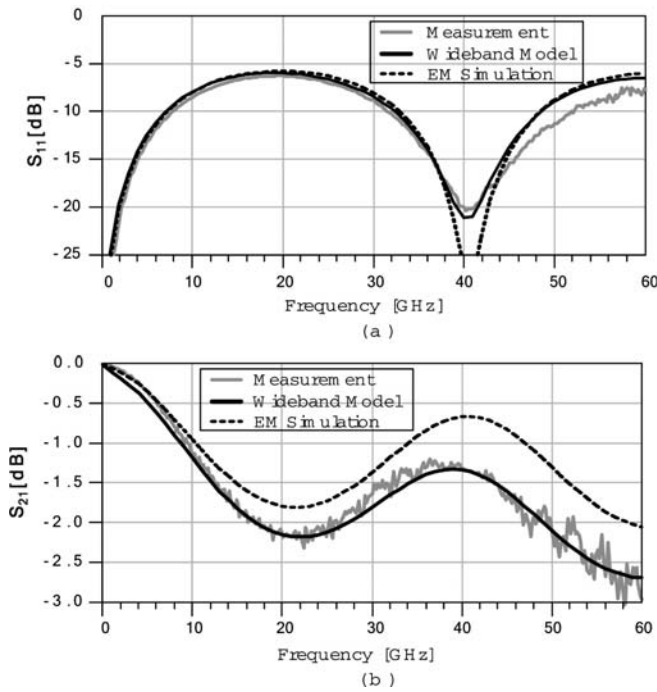


Fig. 6.4 Simulated and measured S-parameters of the transmission line. (a) Reflection parameter (S_{11}) and (b) Transmission parameter (S_{21}).

6.2.3 Conductor-Backed Coplanar Waveguide as the Transmission Line Structure

The conductor-backed coplanar waveguide (CBCPW) structure, shown in Figure 6.5, is another structure that has been used for millimeter wave power amplifiers for impedance matching [20][21]. The use of vias to connect bottom and side ground planes eliminates unwanted parallel-plate modes. Figure 6.5(b) shows the magnetic field distribution in the transmission line, simulated with Ansoft HFSS 3D field solver [16]. The bottom plate carries very little current (small tangential component of the magnetic field), while the side-shield carries most of the return current.

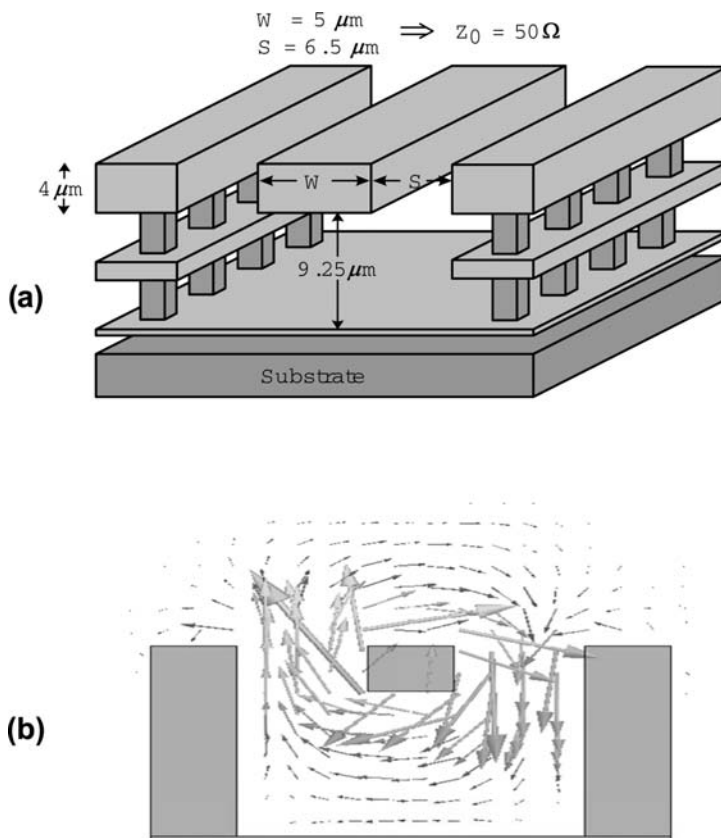


Fig. 6.5 (a) Conductor-backed coplanar waveguide transmission line structure used for matching in the amplifier. (b) The simulated magnetic field distribution of the structure, showing most of the return current coming from the side shields.

The tub shape reduces surface wave propagation in the silicon substrate, improving isolation between lines. Figure 6.6 shows the isolation between two adjacent 50Ω lines, simulated using IE3D at 77GHz, versus their center-to-center spacing. The lines are implemented using the top three metals of the process. The side shields increase isolation by more than 20dB. The coupling in the secondary line is larger in the direction opposite to the wave direction of the primary line.

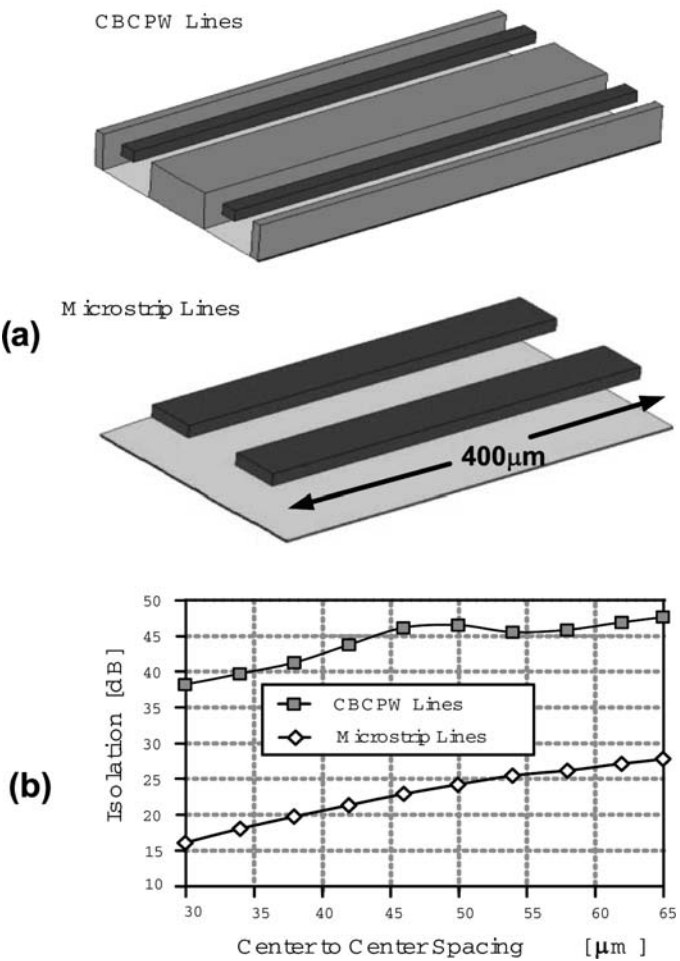


Fig. 6.6 The simulated isolation between two side-by-side 400μm, 50Ω microstrip lines with sideshield ($W = 5\mu\text{m}$, $S = 7.5\mu\text{m}$) and without sideshield ($W = 13\mu\text{m}$).

There is a tradeoff between the isolation of lines and their insertion loss. Since the side-shield increases unit length capacitance, to keep the characteristic impedance constant, the width of the line should be reduced. This increases loss of the t-line. The

50 Ω line without shield has a loss of 0.5dB/mm, while the loss for the line with side-shield is 0.75dB/mm. Since often times, power amplifier is intended to be used in a single-chip transceiver, it is imperative to minimize the interference generated by the high-power power amplifier to sensitive elements such as on-chip VCO. Therefore, the transmission lines were always used with side-shield. The unloaded quality factor of the transmission line can be found by

$$Q = \frac{\pi}{\lambda\alpha} \quad (6.1)$$

where λ is the guided wavelength, and α is the attenuation in nepers per meter. The corresponding quality factor for the CBCPW line at 60GHz is around 10.

6.2.4 Wirebond and Pad Parasitic Effects

The wirebond inductance can change the response of the matching network. The change in inductance is caused by variations in the length and curvature of the wirebond. 3D electromagnetic simulations for the intended test board reveal a range of 0.2nH-0.5nH for the inductance, depending on different wirebond curvatures.

A way to solve this problem placing capacitors in series with the input and output pads to resonate out this inductance. In the large-inductance mode (wirebond inductance greater than 0.4nH), the voltage swing across the series capacitance can exceed the breakdown voltage of the MIM capacitors(5V). We can use a vertical parallel-plate (VPP) capacitor with a breakdown voltage in excess of 100V to prevent capacitor breakdown [22].

In order to avoid de-embedding the pad capacitance or the need to taking that into account during the design process, we can extend the substrate shield of the transmission line beneath the bondpads, making the bondpads part of the transmission line structure. Furthermore, to eliminate the tapering discontinuity, we can match the width of the transmission line and pads. This is especially important in wideband power amplifier design.

6.3 Power Transistors

Another challenge toward the realization of a fully integrated high-frequency power amplifier is dealing with the low unity power gain of the transistors. Also there is a trade-off between the transistor power gain and its stability. Furthermore, the cut-off frequency of a transistor and its breakdown voltage are correlated: a fast transistor can not handle high voltages. In this section, we look at the limitations of a power transistor and the trade-offs between its gain, stability, and breakdown voltage.

6.3.1 Single-Transistor Power Gain and Stability

The effect of loss elements in a MOS transistor can be better understood by calculating its maximum available power gain, which is the maximum gain that can be achieved from the transistor and is realized when both transistor input and output are simultaneously conjugate-matched to the source and load impedances, respectively. Although the power gain can be readily derived from the S-parameters of the transistor [23], the relationship between S-parameters and the transistor's physical parameters (such as C_{gs} and g_m) is often complicated and does not provide good insight into the gain-limiting mechanisms in a MOS transistor.

As shown in Figure 6.7, by ignoring the gate-drain capacitance of the transistor (assuming unilaterality), a simple equation for maximum available power gain, G_A , can be derived¹. By choosing source and load impedances as in Figure 6.7 and setting $L_1 = 1/(C_{gs}\omega^2)$ and $L_2 = 1/(C_{ds}\omega^2)$, the reactive parts cancel, and the input and output ports of transistor are conjugate-matched. Therefore, $V_{gs} = V_s/(2jR_gC_{gs}\omega)$, and the resulting (unilateral) power gain at amplifier operating frequency, f , will be:

$$G_{AU} = \frac{\text{output available power}}{\text{source available power}} = \frac{g_m^2 R_{ds}}{4\omega^2 R_g C_{gs}^2} \approx \frac{R_{ds}}{4R_g} \left(\frac{f_T}{f}\right)^2 \quad (6.2)$$

where cutoff frequency, f_T is $g_m/(2\pi C_{gs})$.

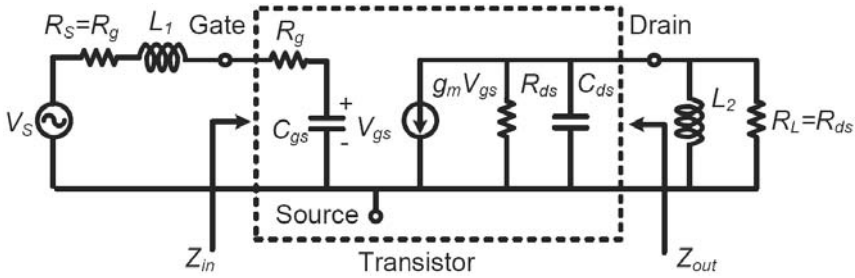


Fig. 6.7 Unilateral model of MOS transistor with conjugate-matched source and load terminations is used to calculate the maximum unilateral power gain.

The gain of a conjugate-matched FET drops off as $1/f^2$. To maximize the power gain, the gate resistance should be reduced by having smaller finger gate lengths and increasing the number of fingers. However, in practice, the gate-drain capacitance, C_{gd} , cannot be ignored, as it's the source of feedback and can cause instability. Figure 6.8 shows the model of the transistor which includes C_{gd} .

¹ To see the relationship between the transducer power gain (G_T), and the unilateral power gain (G_{TU}), calculated by ignoring C_{gd} , look at [23].

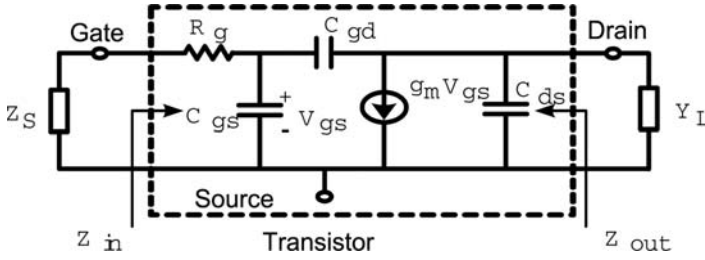


Fig. 6.8 Model of MOS amplifier used to derive stability criterion.

This model is used to analyze the stability of the transistor by looking at the input impedance. In order to simplify the analysis, R_{ds} is ignored. The input and output impedance of the amplifying stage are:

$$Z_{in} = R_g + \frac{1}{sC_{gs} + sC_{gd} + sC_{gd} \frac{g_m - sC_{gd}}{sC_{gd} + sC_{ds} + Y_L}} \quad (6.3a)$$

$$Y_{out} = sC_{ds} + sC_{gd} + sC_{gd} \frac{g_m - sC_{gd}}{sC_{gd} + sC_{gs} + \frac{1}{R_g + Z_s}} \quad (6.3b)$$

When the conjugate match conditions of $Y_{out} = Y_L^*$ and $Z_{in} = Z_s^*$ are imposed, the real part of the input impedance will be:

$$R_{in}^2 = R_g^2 + \frac{R_g C_{gd}}{g_m (C_{gs} + C_{gd})} - \frac{1}{4\omega^2 (C_{gs} + C_{gd})^2} \quad (6.4)$$

Since the right side of the Equation (6.4) has a negative component, in order to be able to have input conjugate match the following condition should hold:

$$R_g > \frac{f_T}{2g_m f} \quad (6.5a)$$

$$\alpha = \frac{C_{gd}}{C_{gs}} \ll 1 \quad (6.5b)$$

Therefore for having an amplifier with a power-matched input R_g should be comparable to $1/g_m$ of the transistor. A more detailed analysis shows that for small values of R_g , the presence of feedback capacitor C_{gd} can make the amplifier unstable. Equations (6.2) and (6.5) demonstrate the conflicting requirements that arise when

a single transistor is used as the amplifying element. As can be seen from Equation (6.5), in order to maximize the power gain through maximum power transfer to the transistor, R_g should be large, in which case the power gain of the transistor will be significantly reduced as per Equation (6.2). This conflict can be resolved by using a cascode design for the amplifying stages.

6.3.2 Stability of the Cascode Pair

As discussed in the previous section, a single MOS transistor designed for maximum power gain in a common source configuration and conjugate matched at input and output can be unstable. The cascode structure makes the device more unilateral, and hence, unconditionally stable. Also, as the cascode pair has a higher drain-source breakdown voltage, for example a 2.8V supply can be used for $0.18\mu\text{m}$ devices that have a drain-source breakdown voltage of 2.5V. The use of cascode amplifying stage in power amplifier, with higher gain and better stability, has the disadvantage of lower drain efficiency. This is because the two transistors in series increase the series resistance when the transistors operate in switching mode. Figure 6.9 shows the maximum stable gain of a single transistor and a cascode stage in a $0.12\mu\text{m}$ process.

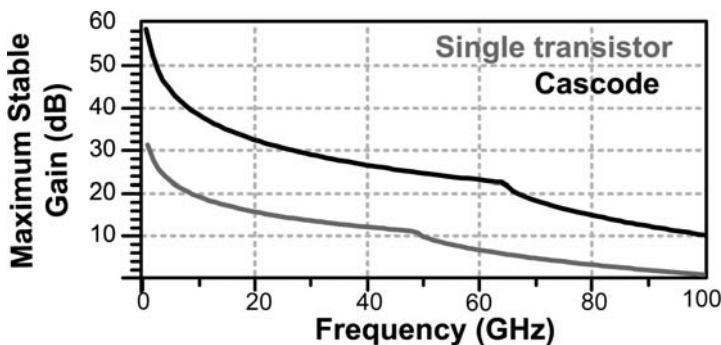


Fig. 6.9 Maximum stable power gain of a cascode stage vs. a single transistor

Let us look at the stability of the cascode stage in more details as shown in Figure 6.10. The gate of M_2 is self-biased by R_2 and bypassed by C_1 . In [24], a self-biased cascode structure has been proposed in which the gate of the cascode device is not grounded at RF. Although such a structure reduces the stress on the cascode transistor, the amplifier has a non-optimal gain performance. Due to the limited gain at millimeter wave frequencies, the gate of the cascode device should be RF-grounded with a large bypass capacitor, C_1 [18]. Careful layout can minimize L_1 , the parasitic series inductance of C_1 . When L_1 is large, there remains a potential

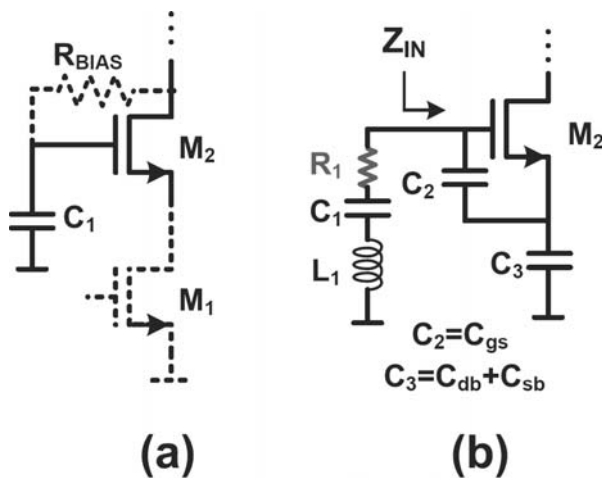


Fig. 6.10 (a) Self-bias of cascode transistor pair. (b) Equivalent circuit for the analysis of stability.

for high-frequency instability. A simple model for the circuit is shown in Figure 6.10. Neglecting gate-drain capacitance of M_4 , the impedance looking into gate of M_4 is:

$$Z_{in} = \frac{1}{sC_2} + \frac{1}{sC_3} + \frac{g_m}{C_2C_3s^2} \quad (6.6)$$

The real part of this impedance has a negative component equal to $-g_m/(C_2C_3\omega^2)$, indicating that the circuit can oscillate if there is a parasitic inductance between the gate and ground. By introducing the series resistance R_1 , the circuit can be stabilized. The value of R_1 should be chosen such that the amplifier remains stable for the largest estimated value of L_1 . Using (6.6), the condition for the stability can be expressed as

$$R_1 > \frac{g_m}{C_2C_3\omega_{osc}^2} \quad (6.7a)$$

$$\omega_{osc} = \frac{1}{L_1C_{eq}} \quad (6.7b)$$

$$C_{eq} = \left(\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} \right)^{-1} \quad (6.7c)$$

6.3.3 Relationship of Breakdown Voltage and Cut-off Frequency

When the power amplifier is driven into saturation, the collector voltage of the output transistor can go up to more than twice the supply voltage. Therefore, the breakdown voltage of the transistor will determine the maximum value of the supply voltage. The two relevant breakdown parameters for the bipolar transistor are the collector-emitter breakdown voltage with the base terminal open circuited (BV_{CEO}), and the collector-base breakdown voltage with an open-circuited emitter (BV_{CBO}), as illustrated in Figure 6.11.

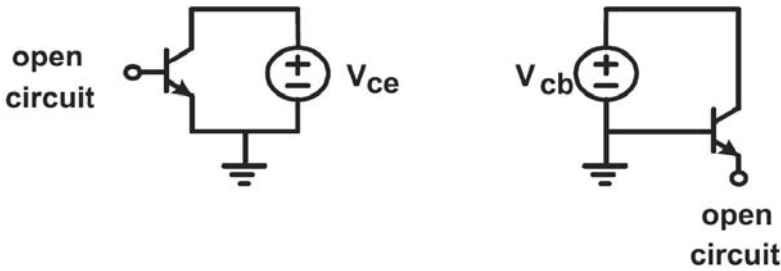


Fig. 6.11 Transistor in the open base and open emitter configurations.

In a normal silicon transistor, maximum dielectric breakdown field and velocity saturation pose a rather fundamental breakdown voltage vs. speed trade-off [25][26]. In Figure 6.12, the f_T versus breakdown voltage relationship of several SiGe HBTs is plotted [11]. For voltages larger than the collector-base breakdown voltage BV_{CBO} , the collector-base junction breaks down. This is independent of the impedance connected between the base and emitter. Therefore, BV_{CBO} is the absolute maximum voltage that can be applied between collector and base. Under all operating conditions, this condition should be taken into account. From the data in Figure 6.12 it seems that the smaller open-base collector-emitter breakdown voltage, BV_{CEO} , will limit the maximum supply voltage of the PA [11]. As discussed in [27], a small supply voltage will reduce the drain efficiency of the PA. It will also necessitate a larger impedance transfer ratio in the output matching network, which in turn will increase the passive losses [28]. In a typical $0.12\mu\text{m}$ process, the BV_{CEO} is just around 1.8V, which will limit the maximum supply voltage to about 0.9V.

Nevertheless, the BV_{CEO} limitation is set by the impact ionization effect, in which the generation of electron-hole pairs by accelerated electrons constitute the necessary base recombination current. If the base is driven with lower source impedance, the extra generated majority carriers will be extracted from the base, and hence the breakdown voltage will increase [29]. In this case, the voltage swing is limited by BV_{CER} rather than BV_{CEO} . In a typical $0.12\mu\text{m}$ process, for R_B equal to 300Ω ,

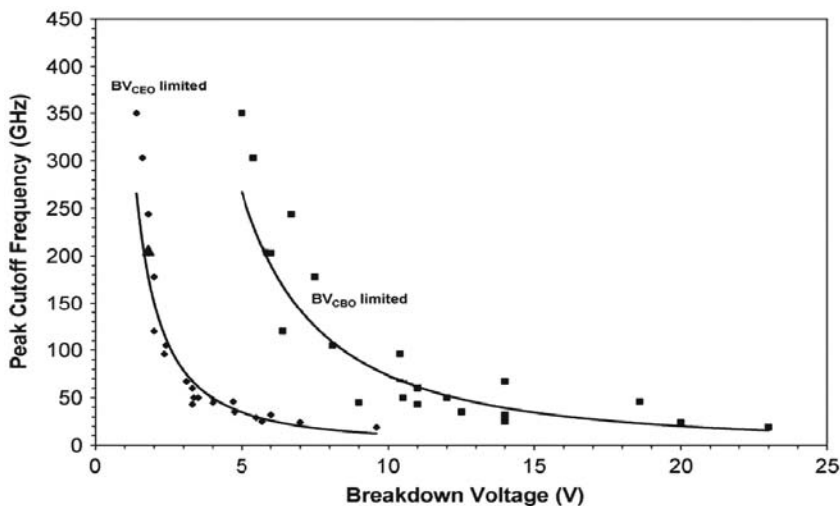


Fig. 6.12 f_T versus breakdown voltage relationship of SiGe HBTs [11].

the BV_{CER} is about 4V [30]. Consequently, higher supply voltages can be used for the PA.

6.4 Power Combining Techniques

6.4.1 Basic Principle

When the power level of the output stage of a PA is increased, many parallel transistors can be used to generate the output power. Efficient power combining is particularly useful in silicon where a large number of smaller power sources and/or amplifiers can generate large output power levels reliably. This would be most beneficial if the power combining function is merged with impedance transformation that will allow individual transistors to drive more current at lower voltage swings to avoid breakdown issues [28][31]. Furthermore, for a constant output power, using several parallel transistors reduces the size of each transistor and hence the compact lumped model of the transistor becomes more accurate. Division of the power generation core into smaller cells has additional advantages in terms of uniform on-chip heat distribution and also relaxed impedance transformation ratio [28], but necessitates the use of a power-combining structure.

In power combining circuits with hybrid or corporate combiners, the power combining network is matched to each transistor cell [32]. In this case the output power degradation due to individual device failure will be graceful [33]. In a low-yield com-

pound III-V process, there is a chance that one of the transistor cells in the power combining network fails to operate properly. In a silicon process with a high yield the extra constraint of individual match can be traded for a simpler power combining network with lower loss and hence higher amplifier efficiency.

As shown in Figure 6.13, as long as there is a global match between load and effective parallel impedance of all the branches, there won't be any reflection at the combining node. In this figure, $E_{i,j}$ is the incident wave in branch j , and $E_{r,ji}$ is the reflected wave in branch j caused by the incident wave in branch i . Using KCL it can be shown that:

$$\sum_{i=1}^n E_{r,ji} = 0 \quad (6.8)$$

Therefore, when all the branches are driven in-phase, due to superposition, reflection of each branch is canceled out. In other words, simply by connecting different branches and having a global power match there would be no power loss due to reflection. By eliminating the complex corporate power combining network, the passive loss is significantly reduced.

6.4.2 Distributed Active Transformer

Distributed active transformers (DATs) were invented for power combining and impedance transformation at lower giga-hertz frequency range [28]. The first fully integrated power amplifier for cellular phone applications was implemented on a silicon substrate using a DAT. The DAT topology provides high efficiency power combining and impedance transformation at the same time.

A conventional transformer is often used in different power amplifier architectures. The primary purpose of the transformer is to change the impedance and also increase the power enhancement ration (PER). PER is defined as the ratio of the delivered power to the load with the transformer to the power delivered to the load without it. The analysis in [28] suggests that if we can increase the transformer turn ratio with a constant quality factor, we can achieve a larger PER. Unfortunately, in reality as we increase the turn ratio of the transformer, the quality factor suffers. Also the impedance level is proportional to the square of the turn ratio, which means that the impedance will be impractically small for higher turn ratios.

This argument leads us to lowering the turn ratio which means results in a small PER. To solve this problem, we can use N independent 1:1 transformers by connecting the secondaries as shown in Figure 6.14. Here, the voltage on the secondaries add while the primaries are driven by low voltage power sources. Since each power source drives $1/N$ th of the total primary inductor, the impedance transformation ratio is N . In ideal lossless case, N power sources are being combined and therefore the PER is N^2 .

One problem with the structure of Figure 6.14 is that the two terminals of the slab inductors are not in close proximity of each other. Hence, if one wants to connect the

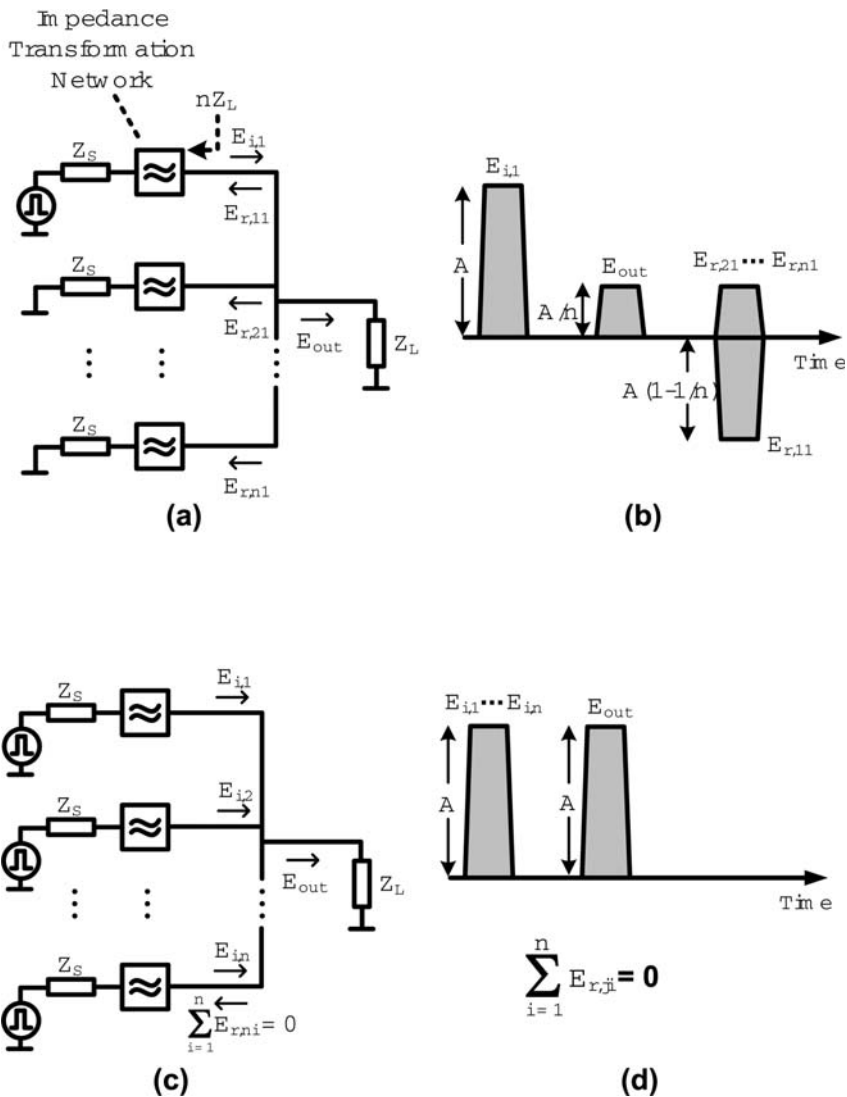


Fig. 6.13 (a) Power combining without individual branch match, but satisfying global match to the load. (b) Scattering behavior for one of the incident waves at the combining point. (c) Scattering behavior when all the branches are driven in-phase. (d) Cancellation of branch reflection through superposition and symmetry.

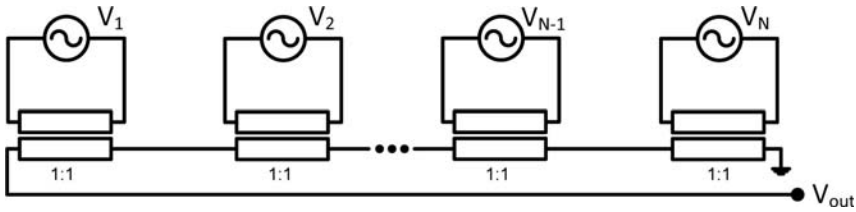


Fig. 6.14 Basic idea of a DAT.

signal sources to the slab inductors, the inductance of the wires will be comparable with the inductance of the slab inductors. To address this problem, one can use a double differential drive. This will create a virtual ground in the middle of the slab inductors that can be used for dc power supply: the impedance of the dc bias does not affect the circuit in a differential mode. Furthermore, to make sure that the ground connection of the driving transistors are in close proximity, it is necessary to use a double differential drive which creates a virtual ac ground at the ground connection of each power source. This ac ground between the transistors, also stops the ac current from flowing through the ground line. This double differential drive is shown in Figure 6.15.

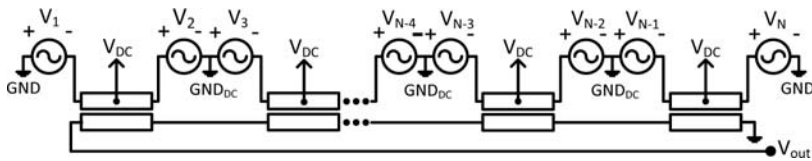


Fig. 6.15 A DAT with double differential drive and shared grounds between power sources except for the first and last one.

The only problem in the topology shown in Figure 6.15 is that the two power sources at the two ends of this structure do not have the virtual ground. To solve this problem, one can wound the structure as shown in Figure 6.16. This winding reduces the quality factor of the inductors due to the negative magnetic coupling between opposite sides of the structure. Fortunately this reduction in the quality factor is smaller than winding each transformer individually as the distance between the opposite sides in a DAT is much larger.

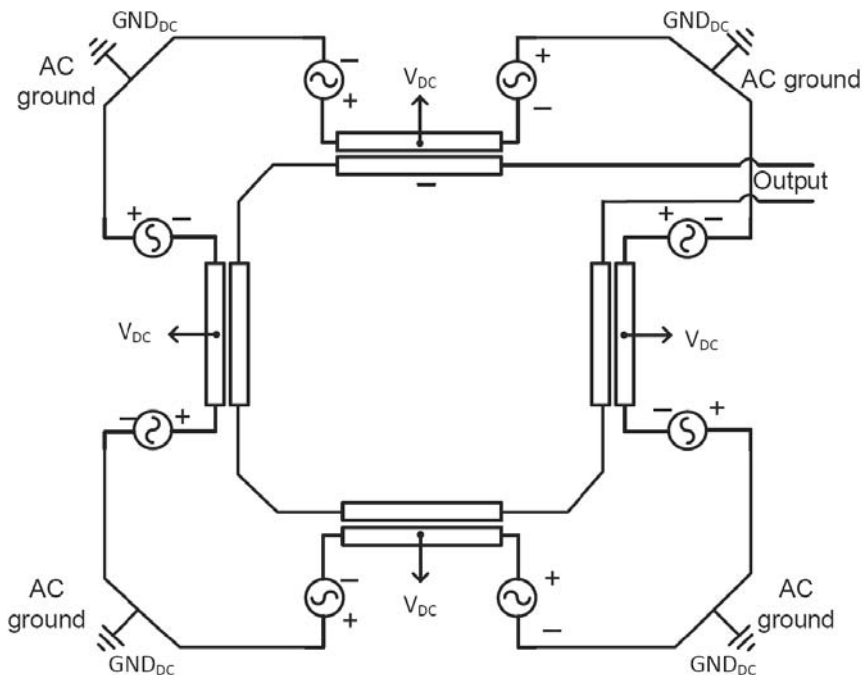


Fig. 6.16 The topology of a DAT with double differential drive.

To summarize, the DAT structure has three main advantages:

- High efficiency power combining,
- Impedance transformation,
- Providing the dc current through virtual grounds which makes the amplifier insensitive to the bias network.

Using the DAT for millimeter-wave frequencies imposes a few challenges. The main challenge is that as the frequency goes up, the value of the primary inductors should go down. This means that the DAT is more sensitive to the value of the parasitics. Furthermore, the synchronization of the double differential drivers is more challenging. To address these challenges, a stacked transformer has been proposed [34] as shown in Figure 6.17.

Also to predict the parasitics more accurately, side bars have been added to this structure. These side bars help the accurate E/M modeling of the transformers. The complete DAT topology at 60GHz is shown in Figure 6.18. In this work, the authors, have demonstrated a 23dBm power amplifier at 60GHz.

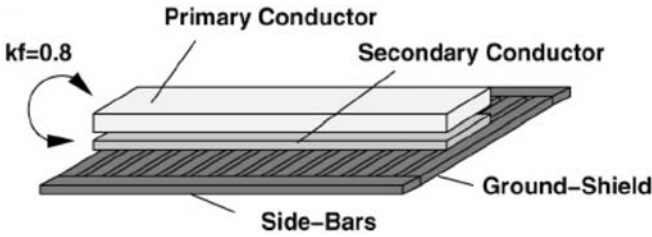


Fig. 6.17 The stacked transformer structure used in a DAT which can operate at millimeter-wave frequencies.

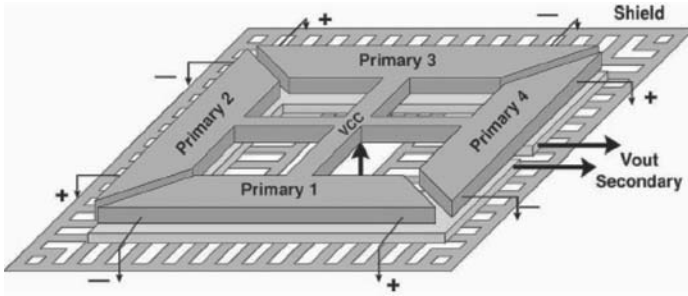


Fig. 6.18 The stacked transformer structure used in a DAT which can operate at millimeter-wave frequencies.

6.4.3 Electrical Funnel

6.4.3.1 Background

One-dimensional LC ladders have been extensively studied before. In fact, the idea of a nonuniform linear transmission line goes back to the work of Heaviside in the nineteenth century and numerous references prior to 1955 can be found in Kaufman's bibliography [35]. A homogeneous 1-D LC ladder consists of identical LC blocks repeated multiple times and can support wave propagation up to a certain cut-off frequency. It can also be used for broadband delay generation and low ripple filtering. An inhomogeneous linear 1-D line can be used to introduce controlled amounts of dispersion to a signal for filtering and equalization.

A 1-D LC ladder can be generalized to a 2-D propagation medium by forming a lattice consisting of inductors (L) and capacitors (C) with triangular, square, or hexagonal tiling formation lattices. Figure 6.19 shows a square lattice.

Generally, this lattice can be inhomogeneous where the L 's and C 's vary in space or nonlinear where they are current and/or voltage dependent [36][37][38][39]. When the L 's and C 's do not change too abruptly between adjacent cells, it is possible

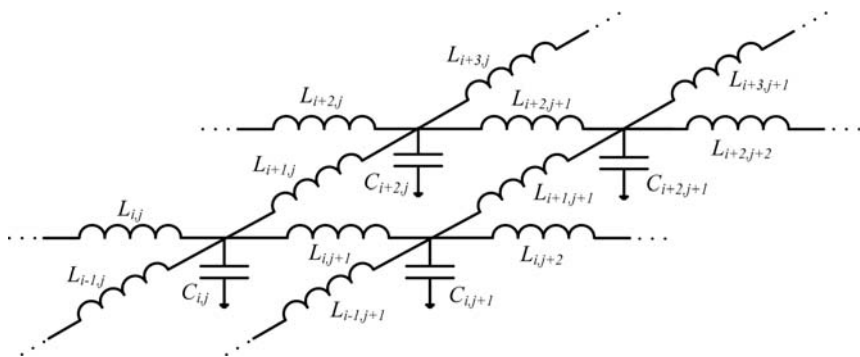


Fig. 6.19 2D square electrical lattice

to define local propagation delay ($\tau_d = \sqrt{LC}$) and local characteristic impedances ($Z_0 = \sqrt{L/C}$) at each node.² This allows us to define local impedance and velocity as functions of x and y , which can be engineered to achieve the desired transmission and reflection properties [40]. One application of these 2-D lattices as a means for simultaneous power combining and impedance transformation.

6.4.3.2 Funnel and Lens

One way these surfaces can be engineered is by keeping the propagation velocity constant vertically (constant LC product for a given y), while increasing the characteristic impedance at the top and bottom of the lattice at a faster rate as we move horizontally along the x -axis to the right, as illustrated in Figure 6.20. A planar wave propagating in the x direction from left to right gradually experiences higher impedances at the edges, creating a lower resistance path for the current in the middle; this funnels more power to the center as the wave propagates to the right, while we can perform a gradual impedance transformation from the left to the right. By keeping the propagation velocity independent of y as we move along the x axis, we can maintain a plane wave keeping the lattice response essentially frequency independent for the frequencies lower than its natural cut-off frequency [40]. This structure is called an *electrical funnel* due to the way it combines and channels the power to the center at the output.

There is a dual to the funnel where the local characteristic impedance is kept independent of y while the propagation velocity is modified to increase at the top and bottom of the combiner as the wave front moves to the right. The input sources on the left boundary add coherently at the focal point which should occur in middle of the right boundary where the output is taken. This resembles the behavior of an

² In the most general case, τ_d and Z_0 can be represented as tensors to capture the non-isotropy of the lattice structure in question.

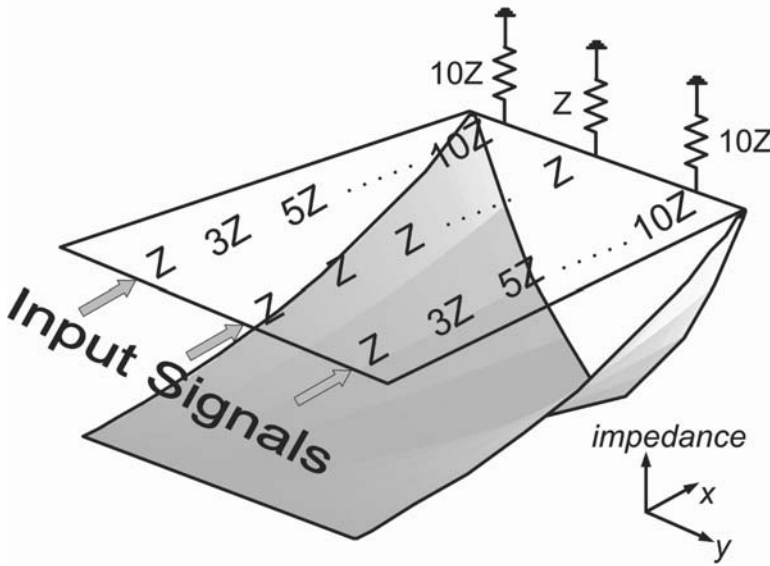


Fig. 6.20 Basic idea of the funnel

optical lens and is thus called an *electrical lens*, due to its focusing nature. However, this focusing behavior is frequency dependent and hence works perfectly only at one frequency. For other frequencies, the phase shift from the input to the output is different, resulting in a different focal length. For this reason, electrical lens might not be an ideal power combiner, however it could perform spatial Fourier transform of the input signal [41]. Here, we only focus on the properties and implementation of the *electrical funnel*.

6.4.3.3 Mathematical Analysis

Consider a section of the two-dimensional transmission lattice shown in Figure 6.21. Using only regular polygons, there are three possible lattice blocks that can be used to tile the two-dimensional plane: triangular, rectangular, and hexagonal. Though the governing equations in each case will be different, at the continuum limit, they will have the same physical properties. Therefore, for mathematical simplicity, we analyze only the rectangular case. We suppose that the lattice is nonuniform, meaning

$$\nabla L(x, y) \neq \mathbf{0}, \quad \nabla C(x, y) \neq \mathbf{0}.$$

where $L(x, y)$ is a continuous function whose value at $x = ih$ and $y = jh$ is l_{ij} and $C(x, y)$ is a continuous function whose value at $x = ih$ and $y = jh$ is c_{ij} . Then

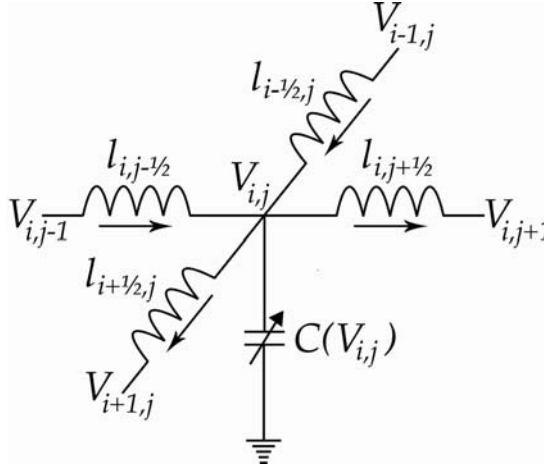


Fig. 6.21 2D transmission lattice

Kirchoff's laws yield the semi-discrete system:

$$I_{i,j-1/2} + I_{i-1/2,j} - I_{i+1/2,j} - I_{i,j+1/2} = c_{ij} \frac{dV_{ij}}{dt} \quad (6.9a)$$

$$V_{ij} - V_{i,j-1} = -\ell_{i,j-1/2} \frac{d}{dt} I_{i,j-1/2} \quad (6.9b)$$

$$V_{ij} - V_{i+1,j} = \ell_{i+1/2,j} \frac{d}{dt} I_{i+1/2,j} \quad (6.9c)$$

Differentiating (6.9a) with respect to time, we substitute (6.9b-6.9c), yielding

$$\frac{V_{ij} - V_{i,j-1}}{\ell_{i,j-1/2}} + \frac{V_{ij} - V_{i-1,j}}{\ell_{i-1/2,j}} + \frac{V_{ij} - V_{i+1,j}}{\ell_{i+1/2,j}} + \frac{V_{ij} - V_{i,j+1}}{\ell_{i,j+1/2}} = -c_{ij} \frac{d^2 V_{ij}}{dt^2}. \quad (6.10)$$

We assume that the nodes are equispaced and that the node spacing, h , is small. Taking the continuum limit, we obtain the $\mathcal{O}(h^0)$ lattice model

$$\nabla^2 V - L C V_{tt} = \frac{\nabla V \cdot \nabla L}{L}, \quad (6.11)$$

where

$$\nabla^2 V = V_{xx} + V_{yy}.$$

Here, indexes indicate derivative with respect to x and y . Equation (6.11) is a 2-D wave propagation equation with a correction term on the right hand side due to the inhomogeneity of the medium. A second order approximation, capturing the dispersive nature of the lattice can be written by keeping all the terms at order h^2 . This approximation is given in Appendix A.

Considering an extremely large $M \times N$ lattice, i.e., the case when M and N are both large, we may ignore the $\mathcal{O}(h^2)$ terms and use (6.11) as our governing equation.

The transmission lattice is at rest (no voltage, no current) at $t = 0$, at which point a sinusoidal voltage source with amplitude A and frequency f is switched on at the left boundary. We assume that the transmission lattice is long in the x direction, and that it is terminated at its (physical) right boundary in such a way that the reflection coefficients there are very small. Hence we model the transmission lattice as semi-infinite in the x coordinate, but bounded in the y coordinate.

There are at least two methods for finding exact solution of (6.11):

1) General Exact Solution:

Suppose we insist on a solution V that is separable in the following sense:

$$V(x, y, t) = p(x, y)g(x, t). \quad (6.12)$$

Here g satisfies the constant impedance equation $\omega_0^{-2}g_{tt} = \nabla^2 g$, i.e.,

$$g(x, t) = -A \sin \left[\frac{\omega}{\omega_0} (x - \nu_0 t) \right]. \quad (6.13)$$

In words, V represents a sinusoidal wave front g that propagates to the right along perfectly horizontal rays, with spatially dependent amplitude p . For the case of funnel (constant LC product) the solution of this initial-boundary-value problem could be written as [40]:

$$V(x, y, t) = \frac{-A\kappa_1(y)}{\kappa_2(y)x + \kappa_1(y)} \sin \left[\frac{\omega}{\omega_0} (x - \nu_0 t) \right]. \quad (6.14)$$

where κ_1 and κ_2 should satisfy:

$$L(x, y) = 4(\kappa_2(y)x + \kappa_1(y))^{-2} \quad (6.15a)$$

$$\kappa_2 = \frac{\kappa_1}{\kappa_{1yy}} \kappa_{2yy}. \quad (6.15b)$$

By proper choosing of κ_1 and κ_2 one could build any desired function for L and find the voltage anywhere in the lattice. In general, this solution could be complicated.

2) Exponential Funnel:

we could assume an exponential shape for the inductance and capacitance functions, resulting in an exponential impedance, Z_0 and constant propagation velocity:

$$L(x, y) = B \exp(\lambda_1 x + \lambda_2 y) \quad (6.16a)$$

$$C(x, y) = \frac{\tau^2}{B} \exp(-\lambda_1 x - \lambda_2 y) \quad (6.16b)$$

next we define a normalized voltage, u , as:

$$u = \frac{V}{\sqrt{Z_0}} \quad (6.17)$$

Defining u and a simple Fourier transform could simplify (6.11) to a well-known Helmholtz differential equation:

$$(\nabla^2 + \frac{1}{4}(\lambda_1^2 + \lambda_2^2))u = -\omega^2 \tau^2 u \quad (6.18)$$

6.4.3.4 Numerical Analysis

Let us now turn to numerical simulations of the lattice dynamics, starting from Kirchoff's equations (6.9). These equations are discrete in space but continuous in time. For a lattice with M elements in the vertical direction and N elements in the horizontal direction, we have a system of $2MN$ ordinary differential equations (ODEs). By prescribing inductance and capacitance functions L and C together with initial and boundary conditions, we may numerically integrate these ODEs and solve for the voltage and current in the lattice. In our studies, we shall suppose that the lattice is initially dead: that is, at $t = 0$, all voltages and currents are zero except at the left boundary. Multiple synchronous signal sources driving the low-impedance left boundary side of the funnel generating a planar wave-front moving along the x axis:

$$V_{1,j}(t) = V_0 \sin \omega t.$$

The output node is at the center of the right boundary. The entire right boundary nodes are terminated with a resistor matched to the local impedance at that node. The up and down boundaries are kept open.

For all subsequent numerical results, we will use the “funnel” inductance function of the form (6.16). Physical limitations for on-chip fabrication of inductors and capacitors limit the maximum and minimum inductance and capacitance in the lattice and therefore we used $30\text{pH} \leq L \leq 150\text{pH}$ and $30\text{fF} \leq C \leq 300\text{fF}$ and quality factor of 20 for all elements in all of the following simulations.

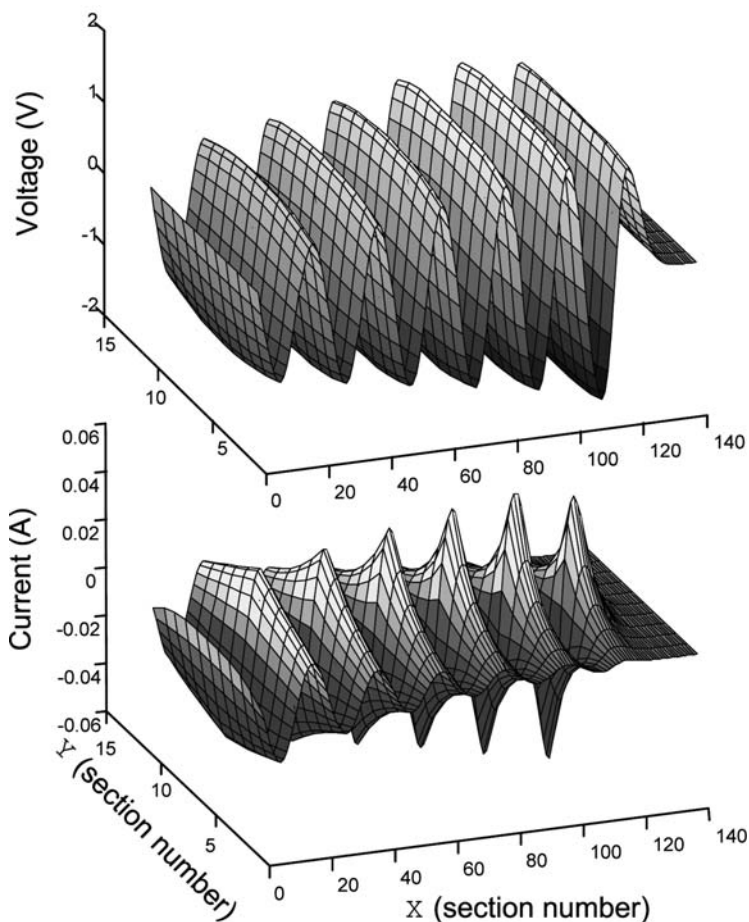


Fig. 6.22 Voltage V_{ij} and current I_{ij} as a function of position for the funnel lattice.

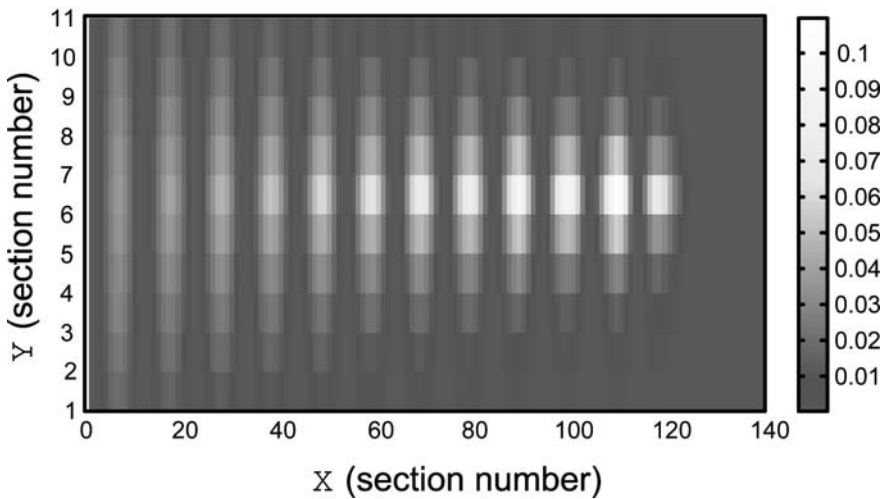


Fig. 6.23 Power P_{ij} as a function of position, demonstrating the funneling effect.

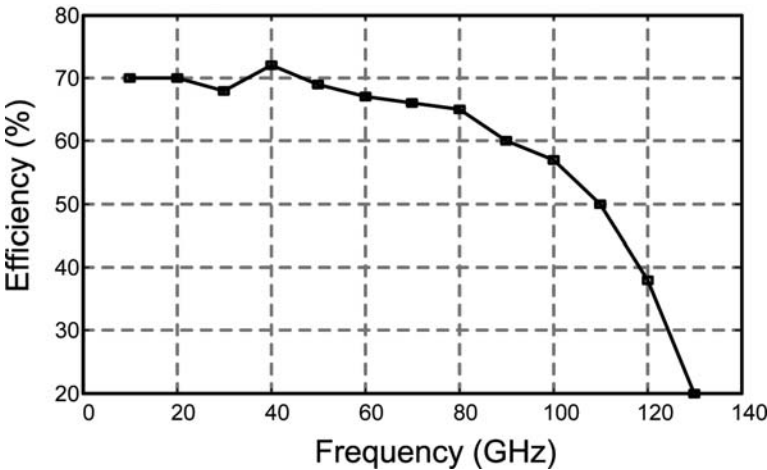


Fig. 6.24 Efficiency as a function of input frequency with $30\text{pH} \leq L \leq 150\text{pH}$ and $30\text{fF} \leq C \leq 300\text{fF}$

Figures 6.22 shows the voltage and current at different nodes of the lattice. There are two intriguing effects in this simulation:

- It is clear that the current focuses near the center of the lattice because of the lower impedance there.
- due to the constant delay ($LC = \text{const}$), we do not see bending of the wave form. we can maintain a plane wave keeping the lattice response frequency independent for the frequencies lower than its natural cut-off frequency.

If we multiply voltage and current, we will have the instantaneous power in the lattice as shown in Figure 6.23. Note that the power is distributed evenly at the left boundary, but at the right boundary it is narrowly focused nearly the central line.

Figure 6.24 shows simulated efficiency of one implementation vs. frequency demonstrating the broadband nature of the electrical funnel. Efficiency is defined by the ratio of the power at the output node to the sum of powers of inputs.

6.4.4 Circuit Implementation

In practice, the characteristic impedance at the edges of the rectangular implementation keeps increasing and hence it is possible to discard the higher impedance parts of the mesh as we move to the right, effectively reducing it to a trapezoid. In order to further reduce the size of the electrical funnel, it is possible to design a continuous version of the electrical funnel: In a silicon process with multiple metals, we can use different metal layers as the ground plane at different points on the y axis. One funnel implementation uses four lower metal layers to form the variable depth ground plane [31]. This leads to different capacitance per unit length that can be used to control the local characteristic impedance across the combiner, as shown in Figure 6.25. Since this does not change the inductance, the propagation delay is not constant vs. y , resulting in a band-pass response. The output is matched to 50Ω while each of the inputs is matched to the output impedance of each power source which is around 15Ω .

While the result might look similar to previously designed multi-way sector shaped power combiners [42], it is different. Here is the reason: The basic principal of this 2-D media (with applications such as power combing in the case of the funnel) is local control of the characteristic impedance and propagation velocity via changing the local inductance and capacitance and hence creating a preferred signal path without introducing an abrupt change in either impedance or propagation velocity. The previous designs are based on planar designs that use equal spacing between the upper and lower metal lines (in a microstrip like setting) [42]. Although the top view of the funnel in the die photo resembles the four-way combiner, it looks completely different underneath. It relies on changing the local characteristic impedance and propagation velocity by changing the capacitance (and/or inductance) per unit length by using different metal layers (with different spacing) underneath. In fact, the funnel

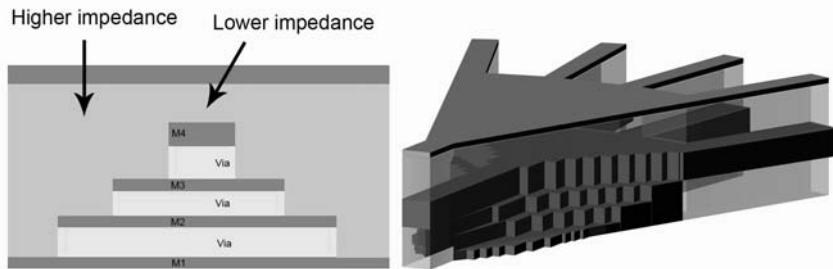


Fig. 6.25 Combiner structure

does not even have to look like a sector and in the more general case, is done as a rectangle, as can be seen in the last section.

In other words, although it is true that the tapering results in a general change of impedance but it is a different kind of impedance change from the point-wise characteristic impedance and propagation delay changes that can be engineered in the case of 2-D medium. Impedance change due to the tapering in a sector power combiner is a global effect which has a lot less flexibility in terms of what can or cannot be done in the design and is a very limited subset of all the possibilities that the 2-D media offer with local control of the characteristic impedance at any point in space.

The difference between this structure and a standard tapered transmission line is a larger bandwidth (45% increase in [31]) over a shorter distance (lower loss) due to the variable-depth ground plane. Intuitively this ground plane makes *soft* boundaries compared to a standard transmission line. As a result, input and output have better matching over a broader range of frequencies.

This version of electrical funnel is not as broadband as Figure 6.24, however it has lower dimensions compared to a lattice of inductors and capacitors.

6.5 Case Studies

In this section, we review two examples of power amplifier design in millimeter wave frequencies. Both power amplifiers have been designed in a $0.12\mu\text{m}$ BiCMOS process featuring SiGe transistors with a cutoff frequency of $f_T \approx 200\text{GHz}$ and $f_{max} \approx 280\text{GHz}$ [43]. The back-end consists of 5 metal layers with three copper bottom layers and two thick $1.25\mu\text{m}$ and $4\mu\text{m}$ aluminum layers as top metals. The breakdown voltages of the bipolar transistors are BV_{CEO} 1.7V and BV_{CBO} 5.5V with a substrate resistivity of $14\ \Omega\cdot\text{cm}$. The first design, is a 77GHz power amplifier with 17.5dBm of output power and a peak power added efficiency (PAE) of 13%, intended for the automotive radar system [21]. The second design is a 85GHz power

amplifier with 3-dB bandwidth of more than 24GHz [31]. We examine these two designs to review some of the design and measurement methodologies. In the next section, we will look at a few other works in this frequency range and summarize the state-of-the-art technology.

6.5.1 A 77GHz Amplifier for Automotive RADAR Application

The schematic of the amplifier is shown in Figure 6.26. The amplifier consists of 4 gain stages, where the output stage is designed for maximum efficiency, and the other stages are designed for maximum gain.

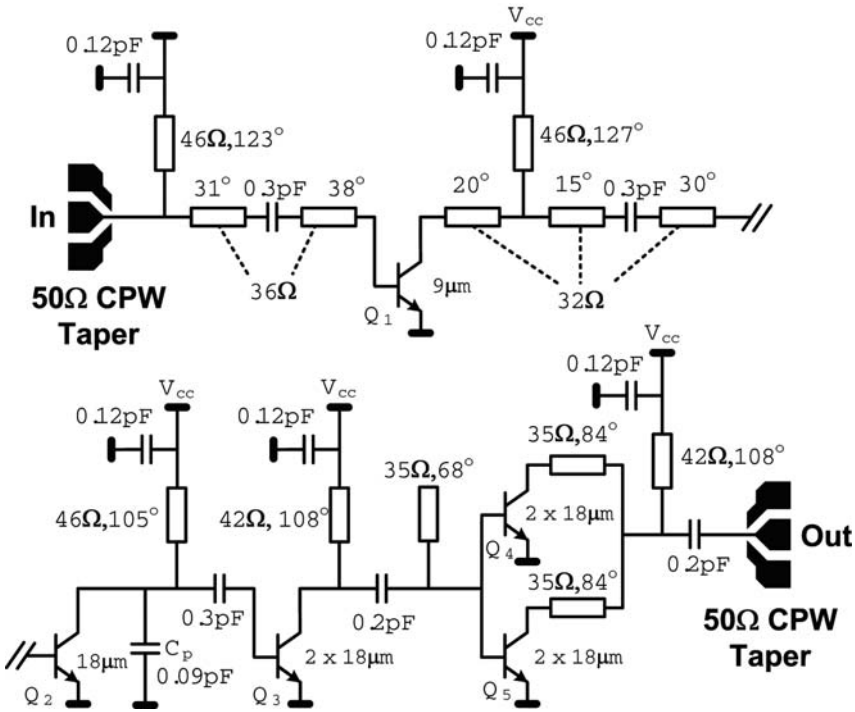


Fig. 6.26 Schematic of the 77GHz power amplifier including element values.

The last three stages use 1, 2, and 4 identical transistor cells, respectively. This geometric scaling of transistor size from each stage to the next ensures that the output transistors will enter compression first as long as the preceding stages have at least 3-dB of gain. All of the transistors have single emitter stripe, use a minimum emitter width of $0.12\mu\text{m}$, and have two base and two collector contacts. For a reliable

operation, the collector junction has more than the minimum number of possible contacts (three rows of long rectangular vias in parallel). The amplifier is biased in class-AB mode. With 1.2mA of current per $1\mu\text{m}$ of emitter length, the transistors are biased at their maximum f_{max} .

6.5.1.1 Design of the Matching Networks

The matching networks use series transmission lines and parallel shorted-stubs for power matching between stages. As shown in Figure 6.26, at the input of the last stage an open stub provides lower matching network loss than a shorted stub does. At the output of second stage, the same objective was achieved with a parallel MIM capacitor (C_p).

The capacitors at the end of shorted parallel stubs are in parallel with a series RC network (which for simplicity is not shown in Figure 6.26. A proper choice of R and C reduces the gain of the amplifier at low frequencies, enhancing stability.

The optimum impedance at the collector of each stage is determined with a large-signal power match. A load-pull simulation is performed to find the best load for the transistor. For the output stage this point is chosen to maximize the efficiency, and for the other stages to maximize the gain. Figure 6.27 shows the result of the load-pull simulations for all of the four stages. These gain and PAE contours have peak values of 6dB and 30% and step sizes of 1dB and 4%, respectively. The contours become denser as we move toward the output stage, indicating larger sensitivity of the amplifier to matching errors. As shown in Figure 6.28, the contour is opened, and this sensitivity is reduced if lower characteristic impedance is used for the transmission lines at the output stage.

By having an initial assessment for the losses in the matching network, the load pull simulations also provide an estimation for the transistor size. The exact size of the transistor is chosen by iterating through the design procedure after the matching network is designed and its corresponding insertion loss is determined. Unlike the linear load-line matching technique described in [44] and used in [45], the large-signal load-pull methodology for choosing transistor size and optimum load impedance captures the large-signal non-linear behavior of the transistor, as depicted in the non-circular shape of the contours of Figure 6.28. The gain and output power contours for a transistor with linear parameters have a circular shape [30].

In Figure 6.27, the realized impedance is not located exactly at the peak of the contours. This is more evident in the output stage, where the realized load provides a PAE that is 4% lower than the maximum possible PAE. This is because the optimum load impedance has not been the only constraint in the design of the matching network. Loss of the matching network also needs to be minimized. A weighted least-mean-square optimization with gradient-descent scheme is utilized to choose the length and characteristic impedance of the lines. The optimization goal is to minimize the weighted sum of the squares of the distance to the optimum load point and the loss in the matching network. Therefore, for having a reasonable passive efficiency, the realized load is not exactly at the center of load pull contours.

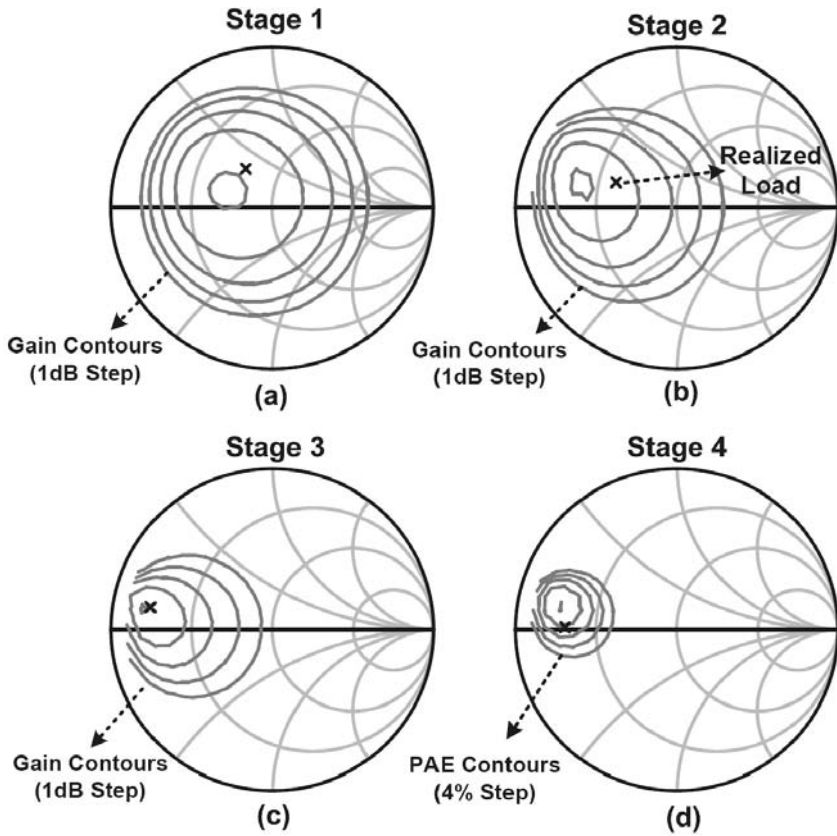


Fig. 6.27 Load-pull simulation of the four stages of the power amplifier, together with the actual realized load impedances.

6.5.1.2 Simulation and Layout Methodology

The die photo of the amplifier is shown in Figure 6.29. Parasitic capacitors are extracted on local nodes where the capacitance is not part of the distributed transmission-line structure. These nodes include connections to transistors, where the signal line is closer to the substrate. The parasitic capacitance is included in the design of the matching networks to make sure the amplifier peak gain and efficiency happen at the desired frequency. Parasitic collector-base capacitance is very important, as it will be multiplied due to the Miller effect and appear at the input or might even cause oscillation. A careful layout minimizes the overlap of the collector and base connections.

The largest ratio of parasitic capacitance to device capacitance (around 60%) occurs at the output of the first stage, with 16.5fF of parasitic capacitance. The layout of one of the output stage parallel branches is shown in Figure 6.30. Each

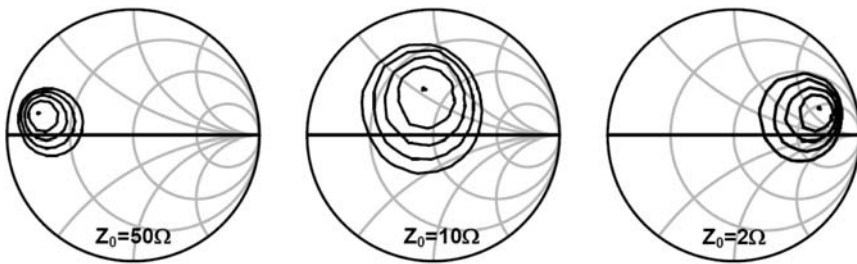


Fig. 6.28 Lowering sensitivity to matching errors in the output stage: load-pull result of the output stage plotted for different reference characteristic impedances in the matching network.

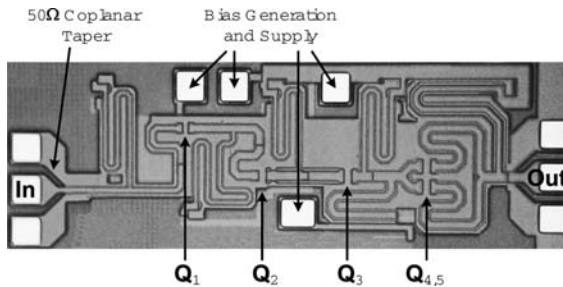


Fig. 6.29 Die micrograph of the 77GHz power amplifier, chip size: $1.35 \times 0.45 \text{ mm}^2$.

transistor has two base and collector contacts, and the spacing between transistors is dictated by design rules. Instead of a larger transistor with $32\mu\text{m}$ emitter length, which has a lower f_{max} , two $18\mu\text{m}$ parallel transistors are used.

6.5.2 A Broadband Amplifier at 85GHz

The second example is a power amplifier that exploits the idea of the electrical funnel as a broadband power combiner [31][46]. Figure 6.31 shows the chip architecture.

Four power amplifier drive the power combiner to maximize the output power. Assume that the voltage at node A in Figure 6.31 is V_{in} , then we could write input and output power as:

$$P_{in} = n \cdot \frac{V_{in}^2}{2Z_1} \quad (6.19a)$$

$$P_{out} = n \cdot \frac{A_v^2 V_{in}^2}{2Z_2} \cdot \eta_{comb.}, \quad (6.19b)$$

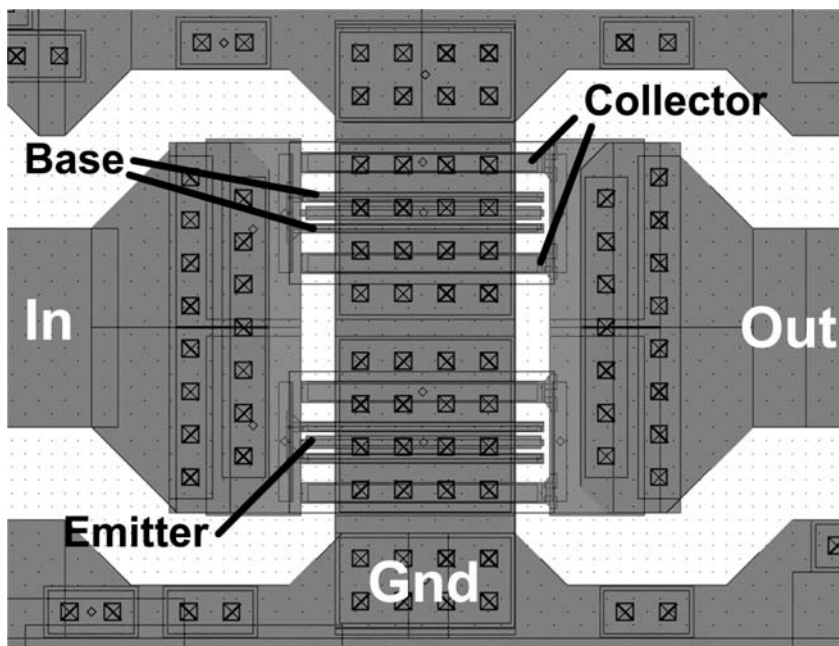


Fig. 6.30 Layout of one of the output parallel branches consisting of two transistors (depicted as Q_5 in the amplifier schematic and layout).

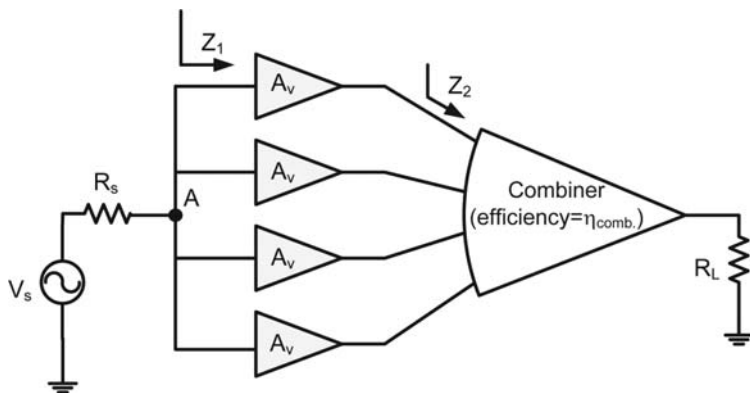


Fig. 6.31 The architecture of the 85GHz power amplifier [31].

where Z_1 and Z_2 are input and output impedance of each amplifier, A_v is its voltage gain, n is the number of amplifiers, and $\eta_{comb.}$ is the combining efficiency. From input and output power, we can find the power gain of the amplifier, G , as:

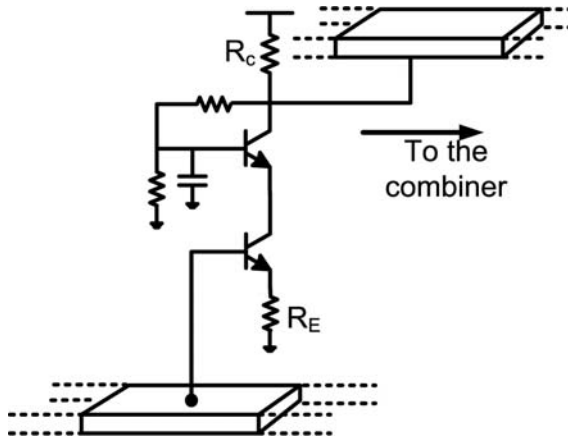


Fig. 6.32 Cascode architecture is used in distributed amplifiers.

$$G = \frac{P_{out}}{P_{in}} = A_v^2 \cdot \eta_{comb.} \cdot \frac{Z_1}{Z_2}. \quad (6.20)$$

In order to obtain a wideband response, a degenerate cascode distributed amplifiers with emitter degeneration as input drivers, shown in Figure 6.32 is used. The main advantage of cascode stage over single transistor is its higher maximum stable power gain. As Figure 6.9 shows, a non-degenerate cascode amplifying stage in this process has a maximum stable power gain of 15dB at 80GHz, as opposed to 7dB for a standard common-emitter. The cascode stages are emitter degenerated to improve bandwidth and avoid thermal runaway.

Each of the four distributed amplifiers consists of eight cascode stages driving the output transmission line, which drive the inputs of the combiner. Figure 6.33 shows the structure of each distributed amplifier.

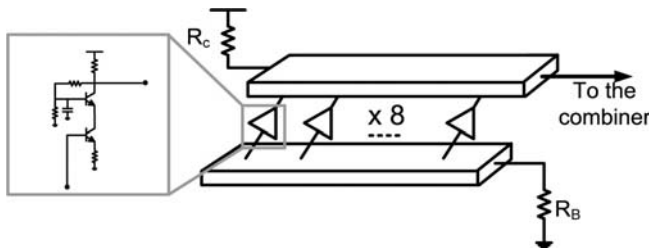


Fig. 6.33 One stand-alone distributed power amplifier.

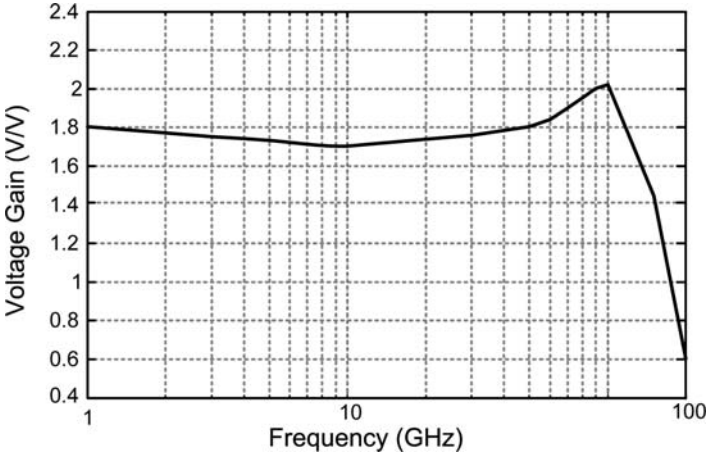


Fig. 6.34 Simulated voltage gain of each distributed amplifier.

Figure 6.34 shows the gain of each stand alone distributed amplifier. Using equation 6.20, for our amplifier, with $A_v \sim 1.8$, $\eta_{comb.} \sim 0.7$, and $Z_1 \sim 4Z_2$ the power gain in 84GHz should be around 9dB, which is close to the measured value of 8dB.

Die photo of the amplifier is shown in Figure 6.35. Figure 6.36 shows the setup used to measure the characteristic of the power amplifier.

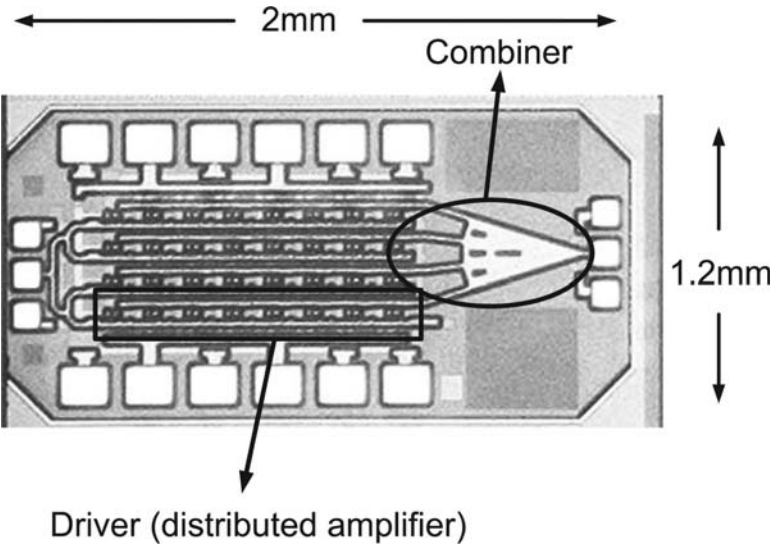


Fig. 6.35 85 GHz power amplifier chip micro photograph.

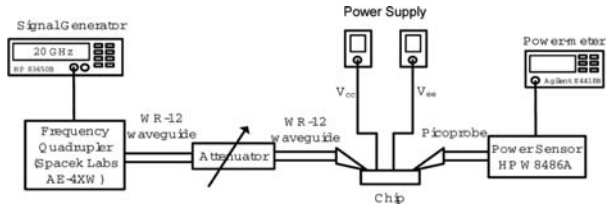


Fig. 6.36 Measurement setup.

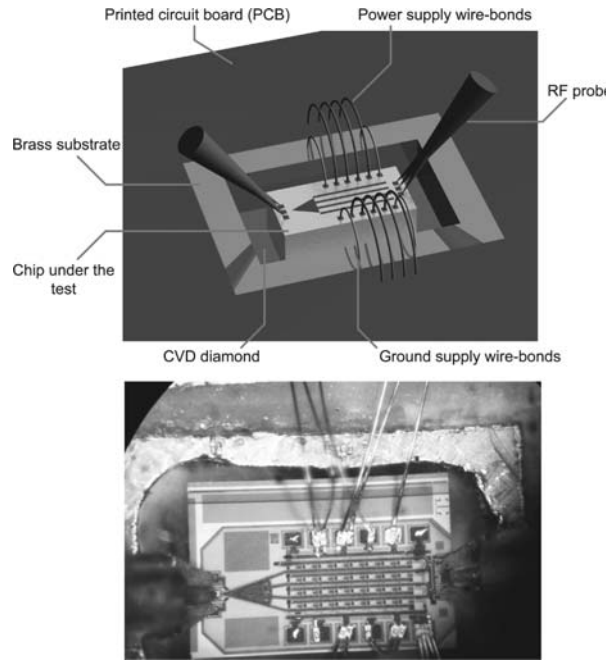


Fig. 6.37 The 85GHz power amplifier chip under test.

The chip is mounted on a brass substrate which is connected to ground. The input is provided by an HP 83650B signal generator and a Spacek frequency multiplier which could generate power from 60GHz to 90GHz. To be able to control input power, a variable attenuation is used before the RF probes. The input and output of the amplifier is probed and the output power is measured using a power-meter. Because the chip has two supplies (-2.5V and 0.8V), the chip substrate (which is at -2.5V) can not directly be connected to the brass. On the other hand it is critical to have a good heat sink for the chip. To solve this problem, a thin low-cost CVD diamond [47] is placed between our chip and brass. Diamond is a superior electrical insulator and is the best isotropic thermal conductor with thermal conductivity of around 10 W/cm/K. Figure 6.37 shows the chip under the test.

6.6 Summary

Table 6.2 summarizes the performance of recently implemented power amplifiers at 60GHz and beyond.

Table 6.2 A comparison between recent power amplifier designs at 60GHz and beyond.

<i>Freq.</i>	Technology (μm)	<i>P_{out}</i> (dBm)	<i>PAE_{max}</i> (%)	Gain(dB)	Ref.
77GHz	0.13 SiGe	13	3.5	6.1	[2]
77GHz	0.13 SiGe	18.5	5.4	-	[45]
60GHz	0.13 SiGe	16	4.3	10.8	[30]
61.5GHz	0.13 SiGe	14	4.2	12	[48]
77GHz	0.13 SiGe	17.5	12.8	17	[21]
58GHz	0.13 SiGe	11.5	20.9	4.2	[49]
85GHz	0.13 SiGe	21	4	8	[31]
60GHz	0.18 SiGe	15.8	16.8	11.5	[50]
60GHz	0.13 SiGe	20	12.7	18	[51]
60GHz	0.13 SiGe	23	6.3	20	[34]

As table 6.2 suggests, it is quite possible to achieve output powers of more than 100mW with an acceptable power added efficiency. The next step seems to be realization of a fully integrated CMOS power amplifier with the same output power level. Limited gain of CMOS transistors at 60GHz makes it extremely challenging to do so. To attack this problem, we need to develop passive structures with lower loss and find ways to efficiently combine the output of a large number of smaller amplifiers. This requires investigation of novel high efficiency power combiners. Electrical funnel and DAT are two successful example of this approach. Of course, there is no reason to stop us from exploiting other physical phenomena to design passive and active structures for 60GHz and beyond.

Appendix A

In equation (6.10), if we keep all terms at order \hbar^2 , we obtain

$$\nabla^2 V - LCV_{tt} = \frac{\nabla V \cdot \nabla L}{L} - \hbar^2 \left[\frac{1}{12} (V_{xxxx} + V_{yyyy}) - \frac{1}{6} \frac{L_x V_{xxx} + L_y V_{yyy}}{L} - \frac{1}{4} \frac{L_x^2 V_{yy} + L_y^2 V_{xx}}{L^2} \right] \quad (6.21)$$

Since we are dealing with slowly varying function L of both x and y , rendering negligible the terms involving squared derivatives of L , i.e., L_x^2/L^2 and L_y^2/L^2 . Our $\mathcal{O}(h^2)$ lattice model is

$$\nabla^2 V - LC V_{tt} = \frac{\nabla V \cdot \nabla L}{L} - \frac{h^2}{12} (V_{xxxx} + V_{yyyy}) + \frac{h^2}{6} \frac{L_x V_{xxx} + L_y V_{yyy}}{L} \quad (6.22)$$

In this equation, the left-hand side is the normal wave propagation equation. On the right, the first term corresponds to the inhomogeneity, the second term is due to the discreteness and the last term represents both inhomogeneity and discreteness. It is noteworthy that this equation captures the dispersion effect of the lattice.

When M and N are both large, we may ignore the $\mathcal{O}(h^2)$ terms and use (6.11) as our governing equation. The reason is simple: suppose we non-dimensionalize (6.22), the third- and fourth-order derivatives of voltage V will be multiplied by factors of $1/N^2$ and $1/M^2$ in the resulting equation; hence these terms are negligible. Physically, by assuming $h \rightarrow 0$, we actually have neglected the effect of dispersion [40].

References

1. A. Hosh, D. R. Wolter, J. G. Andrews, *et al.*, "Broadband Wireless Access with WiMax/802.16: Current Performance Benchmarks and Future Potential," *IEEE Communications Magazine*, vol. 43, pp. 129-136, Feb 2005.
2. U. R. Pfeiffer, *et al.*, "A 77GHz SiGe Power Amplifier for Potential Applications in Automotive Radar Systems," *Proceedings of RFIC*, pp. 91-94, June 2004.
3. A. Natarajan, *et al.*, "A 77-GHz Phased-Array Transceiver with On-chip Antennas in Silicon: Transmitter and Local LO-Path Phase Shifting," *IEEE Journal of Solid-State Circuits*, vol. 41, pp. 2807-2819, Dec 2006.
4. A. Natarajan, A. Komijani and A. Hajimiri, "A Fully Integrated 24-GHz Phased-Array Transmitter in CMOS," *IEEE Journal of Solid-State Circuits*, vol. 40, pp. 2502-2514, Dec 2005.
5. C. H. Doan, S. Emami, A. Niknejad, *et al.*, "Millimeter-Wave CMOS Design," *IEEE Journal of Solid-State Circuits*, vol. 40, pp. 144-155, Jan 2005.
6. B. Razavi, "A 60-GHz CMOS Receiver Front-End," *IEEE Journal of Solid-State Circuits*, vol. 41, pp. 17-22, Jan 2006.
7. S. Vitebskiy, L. Carin, M. A. Ressler, *et al.*, "Ultra-Wideband, Short-Pulse Ground-Penetrating Radar: Simulation and Measurement," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, pp. 762-772, May 1997.
8. C. H. Doan, S. Emami, A. Niknejad, and R. W. Brodersen, "Millimeter-Wave CMOS Design," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 1, pp. 144-155, Jan. 2005.
9. B. Razavi, R. H. Yan, K. F. Lee, "Impact of Distributed Gate Resistance on the Performance of MOS Devices," *IEEE Trans. Circuits and Systems I: Fundamental Theory and Applications*, vol. 41, no. 11, pp. 750-754, Nov. 1994.
10. M. Racanelli and P. Kempf, "SiGe BiCMOS Technology for Communication Products," *Proc. IEEE Custom Integrated Circuits Conf.*, pp. 331-334, Sep. 2003.
11. A. J. Joseph, *et al.*, "Status and Direction of Communication Technologies-SiGe BiCMOS and RFCMOS," *Proceedings of the IEEE*, Vol. 93, no.9, pp. 1539- 1558, Sept. 2005.
12. B. Kleveland, C. H. Diaz, D. Wook, L. Madden, T. H. Lee, and S. S. Wong, "Exploiting CMOS Reverse Interconnect Scaling in Multigigahertz Amplifier and Oscillator Design," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 10, pp. 1480-1488, Oct. 2001.
13. L. Zhu, "Guided-Wave Characteristics of Periodic Coplanar Waveguides with Inductive Loading: Unit-Length Transmission Parameters," *IEEE Trans. on Microwave Theory and Techniques*, vol. 51, no. 10, pp. 2133-2138, Oct. 2003.
14. F. Aryanfar and K. Sarabandi, "Compact millimeter-wave filters using distributed capacitively loaded CPW resonators," *IEEE Trans. on Microwave Theory and Techniques*, vol. 54, no. 3, pp. 1161-1165, Mar. 2006.
15. A. Komijani, A. Natarajan, and A. Hajimiri, "A 24-GHz, +14.5-dBm Fully-Integrated Power Amplifier in 0.18 μ m CMOS," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 9, pp. 1901-08, Sept. 2005.
16. HFSS, High frequency structure simulator [Online]. Available: <http://www.ansoft.com>
17. IE3D, MoM-based electromagnetic simulator [Online]. Available: <http://www.zeland.com>
18. A. Komijani and A. Hajimiri, "A 24 GHz, +14.5 dBm fully-integrated power amplifier in 0.18 μ m CMOS," *Proc. IEEE Custom Integrated Circuits Conf.*, Oct. 2004, pp. 561-564.
19. Advanced Design System (ADS), TLINP: 2-Terminal Physical Transmission Line Model [Online], Available: <http://eesof.tn.agilent.com/products/adsoview.html>
20. W. H. Haydl, "On the use of Vias in Conductor-Backed Coplanar Circuits," *IEEE Trans. on Microwave Theory and Techniques*, vol. 50, no. 6, pp. 1571-1577, June 2002.
21. A. Komijani and A. Hajimiri, "A Wideband 77GHz, 17.5dBm Power Amplifier in Silicon," *IEEE Custom Integrated Circuits Conference*, pp. 571-575, Sept. 2005.
22. R. Aparicio and A. Hajimiri, "Capacity Limits and Matching Properties of Integrated Capacitors," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 3, pp. 384-393, March 2002.
23. D. M. Pozar, *Microwave Engineering*, Wiley, 2005.

24. T. Sowlati and D. M. W. Leenaerts, "A 2.4GHz 0.18 μ m CMOS Self-Biased Cascode Power Amplifier," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 8, pp. 1318-1324, Aug. 2003.
25. E. O. Johnson, "Physical limitations on frequency and power parameters of transistors," *IRE Int. Convention Record*, vol. 13, pp. 27-34, Mar. 1965.
26. K. K. Ng, M. R. Frei, and C. A. King, "Reevaluation of the f_T BV_{ceo} Limit on Si Bipolar Transistors," *IEEE Trans. on Electron Devices*, vol. 45, no. 8, pp. 1854-1855, Aug. 1998.
27. S. D. Kee, The class E/F family of harmonic-tuned switching power amplifiers. Ph.D. Thesis, Caltech, 2002.
28. I. Aoki, *et al.*, "Distributed Active Transformer: A New Power Combining and Impedance Transformation Techniques," *IEEE MTT*, pp. 316-332, Jan. 2002.
29. 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 Journal of Solid-State Circuits*, vol. 40, no. 10, pp. 2008-2018, Oct. 2005.
30. B. A. Floyd, *et al.*, "SiGe Bipolar Transceiver Circuits Operating at 60GHz," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 1, pp. 156-167, Jan. 2005.
31. E. Afshari, H. Bhat, X. Li, and A. Hajimiri, "Electrical Funnel: A Broadband Signal Combining Method," *IEEE International Solid-State Circuits Conference*, pp. 206-208, Feb. 2006.
32. H. O. Granberg, "Broadband Transformers and Power Combining Techniques for RF," Motorola Applications Note AN749, Motorola Semiconductor Products, Inc.
33. D. B. Rutledge, Nai-Shuo Cheng, R. A. York, R. M. Weikle, M. P. De Lisio, "Failures in Power-Combining Arrays," *IEEE Trans. Microwave Theory and Techniques*, vol. 47, no. 7, part 1, pp. 1077-1082, July 1999.
34. U. R. Pfeiffer and D. Gordon, "A 23-dBm 60-GHz Distributed Active Transformer in a Silicon Process Technology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, pp. 857-865, May 2007.
35. H. Kaufman, "Bibliography of Nonuniform Transmission Lines," *IRE Transactions—Antennas and Propagation*, vol. 3, pp. 218-220, 1955.
36. A. C. Scott, *Active and Nonlinear Wave Propagation in Electronics*, Wiley, New York, NY, 1970.
37. M. J. W. Rodwell, M. Kamegawa, R. Yu, M. Case, E. Carman, and K. Giboney, "GaAs Nonlinear Transmission Lines for Picosecond Pulse Generation and Millimeter-Wave Sampling," *IEEE Transactions on Microwave Theory and Techniques*, vol. 39, no. 7, pp. 1194-1204, July 1991.
38. W.-S. Duan, "Nonlinear Waves Propagating in the Electrical Transmission Line," *Europhysics Letters*, vol. 66, pp. 192-197, 2004.
39. E. Afshari and A. Hajimiri, "Nonlinear Transmission Line for Signal Shaping on Silicon," *IEEE Journal of Solid-State Circuits*, vol. 40, pp. 744-752, 2005.
40. E. Afshari, H. S. Bhat, A. Hajimiri, and J. E. Marsden, "Extremely Wideband Signal Shaping Using One-and Two-Dimensional Nonuniform Nonlinear Transmission Lines," *Journal of Applied Physics*, vol. 99, 2006.
41. E. Afshari, H. S. Bhat, and A. Hajimiri, "Electrical Lens: a Novel Analog Fourier Transformation Technique," *IEEE Transactions on Circuits and Systems I*, submitted.
42. K. C. Gupta and M. D. Abouzahra, *Analysis and Design of Planar Microwave Components* IEEE Press, 1994.
43. B. Jagannathan, *et al.*, "Self-aligned SiGe NPN transistors with 285 GHz f_{max} and 207 GHz f_T in a manufacturable technology," *IEEE Electron Device Lett.*, vol. 23, no. 5, pp. 258-260, 2002.
44. S. C. Cripps, *RF Power Amplifiers for Wireless Communications*, Boston, MA: Artech House, 1999.
45. H. Li, H. M. Rein, T. Suttrop, and J. Böck, "Fully Integrated SiGe VCOs with Powerful Output Buffer for 77-GHz Automotive Radar Systems and Applications Around 100-GHz," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 10, pp. 1650-1658, Oct. 2004.
46. E. Afshari, H. Bhat, and A. Hajimiri, "Electrical Funnel: A Broadband Signal Combining Method," *IEEE Journal of Solid-State Circuits*, submitted.

47. E. Worner, C. Wild, W. Muller-Sebert, R. Locher, and P. Koidl, "Thermal conductivity of CVD diamond films: High-precision, temperature-resolved measurements," *Diamond and Related Materials*, Vol. 5, No. 6, pp. 688-692, 1996.
48. U. R. Pfeiffer, D. Goren, B. A. Floyd, and S. K. Reynolds, "SiGe Transformer Matched Power Amplifier for Operation at millimeter-wave Frequencies," *Eur. Solid-State Circuits Conf.*, Sep. 2005, pp. 141-144.
49. A. Valdes-Garcia, S. Reynolds, and U. R. Pfeiffer, "A 60 GHz Class-E Power Amplifier in SiGe," *Proc. Asian Solid-State Circuits Conf.*, 2006, pp. 199-202.
50. C. Wang, Y. Cho, C. Lin, H. Wang, C. Chen, D. Niu, J. Yeh, C. Lee, and J. Chern, "A 60 GHz Transmitter with Integrated Antenna in 0.18 μ m SiGe BiCMOS Technology," *IEEE International Solid-State Circuits Conference*, 2006, pp. 186-187.
51. U. R. Pfeiffer and D. Gordon, "A 20 dBm Fully-Integrated 60 GHz SiGe Power Amplifier With Automatic Level Control," *IEEE Journal of Solid-State Circuits*, vol. 42, pp. 1455-1463, July 2007.

Chapter 7

Integrated Beamforming Arrays

Harish Krishnaswamy and Hossein Hashemi

7.1 Introduction

The advantages of using millimeter waves for communication and radar systems have been known for many years. The major impediments for the wide deployment of commercial millimeter-wave systems have been high cost and poor performance. With the advancement in silicon based processing technologies and their unprecedented device performance, it is likely that the cost of such systems will be lowered in a manner similar to other wireless communication devices such as cell phones and wireless local area network cards. Antenna arrays are used to enhance the performance of wireless communication and radar systems under various names such as phased arrays, beam-forming arrays, spatial diversity and MIMO transceivers.

As is the case with most fundamental discoveries, the origin of the antenna array is shrouded in some doubt. Some point to Nobel Prize winning scientist Guglielmo Marconi and his landmark transatlantic wireless communication experiment in December 1901 [1]. The original antennas conceived for the experiment consisted of an array of twenty aerials. These were unfortunately destroyed by devastating storms, and hence mere two-element arrays were used, through which a repeated Morse Code signal representing the letter 'S' was successfully transmitted from Poldhu in Cornwall, United Kingdom to St. Johns in Newfoundland, Canada.

The credit for the invention of the electronically-steered, linear, phased array is universally given to another Nobel Prize winning scientist Luis Alvarez. His invention was primarily motivated by the U.S. war effort in WW II and formed the basis for *Eagle*, the first radar-based bombing system. Even today, the use of phased arrays is most widespread in the realm of defense, with most warships and fighter planes

Harish Krishnaswamy
University of Southern California, Los Angeles, CA 90089-0271, e-mail: hkrishna@usc.edu
Hossein Hashemi
University of Southern California, Los Angeles, CA 90089-0271, e-mail: hossein@usc.edu

featuring phased-array radars. An example is the AN/SPY1 phased array radar, which is a part of the U.S. Navy’s *Aegis Combat System*.

The current trend of integration of phased arrays on silicon for commercial applications is characterized by a completely different set of challenges when compared to the archetypal phased array for military applications (Table 7.1). Military phased arrays are traditionally used in high-performance radar systems and hence employ thousands of elements to achieve a fine spatial resolution. Beam-scanning is used to locate and track multiple targets simultaneously. As silicon-based technologies are only recently coming into millimeter-wave capability, these phased arrays have traditionally been built using discrete components based on compound semiconductors. Unit cost is usually not a concern, and in the case of warships and, to an extent, fighter planes, size is of limited importance as well. As an example, the AN/SPY1 radar employs 68 receiving sub-arrays, each of which utilizes 2 array modules. Each array module in turn is composed of 32 radiating elements for a total of 4352 elements. The phase-shifters used are non-reciprocal, toroidal ferrite phase shifters [2].

Table 7.1 Comparison of conventional antenna arrays suitable for military applications versus those suitable for the emerging commercial applications.

	Conventional (> 50 years)	Emerging (≈5 years)
Applications	Military Radar	Wireless Communications Automotive Radar
Typical Range	Long Range (> 1km)	Short Range (< 100m)
Array Size	Large (100-10000)	Small (4-64)
Why an Array?	* Focussed, High-Power Beam * Multiple, Simultaneous Beams * Spatial Interference Cancellation * SNR Improvement in RX	* SNR Improvement in RX * Relaxed PA Requirement * Link Reliability (Comm.) * Spatial Selectivity (Radar)
Driver (in order)	* Performance * Size * Cost	* Cost * Size * Power Consumption
Realization Technology	Module-based III-V	Single Chip Silicon

On the other hand, the burgeoning commercial applications, such as high data rate wireless communications for Wireless Personal Networks (WPANs) at 24GHz and 60GHz and vehicular radar at 22-29GHz and 77GHz, are relatively low-performance systems when compared with military systems. The link distances involved are of the order of a few meters and a fine spatial resolution is not required. As a result, these commercial phased arrays will likely employ tens of radiating elements, rather than thousands. The unit cost is a critical issue for market success, and hence, integration onto silicon-based technologies, particularly CMOS, is critical. This is rendered feasible by the lower required performance, in terms of Effective Isotropic Radiated Power (EIRP) and array sensitivity for example, and the ability of the latest generation of silicon-based technologies to handle millimeter-wave frequencies.

Further, silicon-based technologies boast the advantage of being able to integrate millions of devices onto a single chip with near-zero incremental device cost and high reliability. This can be harnessed to increase system functionality and implement calibration circuitry to fine-tune system performance at virtually no extra cost. The latter is particularly important as the packaging of single-chip antenna arrays at millimeter-wave frequencies is challenging. Specifically, the interface between the single-chip and the widely spaced off-chip antennas introduces channel mismatches that deteriorate array performance. Calibration circuitry allow for the correction of packaging mismatches, and hence can greatly reduce packaging effort and cost.

Much in the way that advancement in silicon integration and mixed-signal IC design led to the dramatic cost reduction and performance enhancement of commercial wireless communication systems, silicon based antenna arrays will play a key role toward the wide deployment of future millimeter wave communication and sensing systems.

7.2 What is a Phased Array?

A phased array is a multiple-antenna system that *electronically* modifies the direction of transmission/reception of the electromagnetic beam. **This is done by introducing a variable time delay in each antenna's signal path to compensate for the path differences in free space.** Fig. 7.1 depicts the block diagram of an N-channel phased array receiver. The uniform antenna spacing is assumed to be d , and each antenna's signal path contains a variable delay block, following which the different signal paths are combined. A plane-wave beam is assumed to be incident on the array at an angle of θ_{in} to the normal direction. Because of the spacing between the antennas, the beam will experience a time delay equal to $\frac{d \sin \theta_{in}}{c}$, where c is the speed of light in free-space, in reaching successive antennas. Hence, if the incident beam is a sinusoid at frequency ω with an amplitude of A , the signals received by each of the antennas can be written as

$$S_i(t) = A \cos \left(\omega \left(t - (i-1) \frac{d \sin \theta_{in}}{c} \right) \right). \quad (7.1)$$

The incident plane wave experiences a *linear delay progression* in arriving at the successive antennas. Therefore, to compensate for this, the variable delay blocks must be set to a similar but reverse delay progression. Fig. 7.1 shows the i^{th} delay block set to $(N-i+1)\Delta\tau$. This would perfectly compensate for the incident progression if $\Delta\tau = \frac{d \sin \theta_{in}}{c}$. The signal in each channel at the output of the variable delay block can be written as

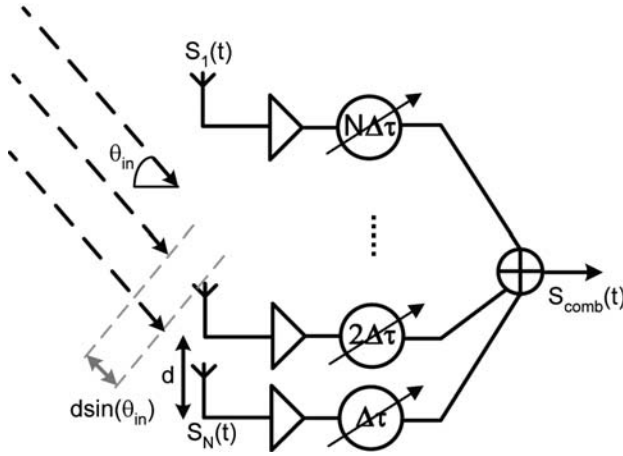


Fig. 7.1 Basic phased array receiver block diagram

$$S_{i,delayed}(t) = G \times A \cos \left(\omega(t - (i-1) \frac{d \sin \theta_{in}}{c} - (N-i+1) \Delta \tau) \right), \quad (7.2)$$

where G is the gain of each channel's front end. We are now in a position to compute the *array factor* (AF) of this phased array receiver. AF is defined as the additional power gain achieved by the phased array receiver over the power gain of a single channel. After summing $S_{i,delayed}(t)$ across the different channels and determining the ratio of the power of the summed signal to that of the signal in each channel, AF can be found as

$$AF(\Delta \tau, \theta_{in}) = \left(\frac{\sin \frac{N(\omega \Delta \tau - \frac{\omega d}{c} \sin \theta_{in})}{2}}{\sin \frac{\omega \Delta \tau - \frac{\omega d}{c} \sin \theta_{in}}{2}} \right)^2. \quad (7.3)$$

When $\Delta \tau = \frac{d \sin \theta_{in}}{c}$, an additional power gain of N^2 is achieved by the phased array as the path differences in free-space are perfectly compensated for and the received sinusoids are added coherently. AF is lower for other angles of incidence, and hence the angle of incidence of peak gain, called the beam-pointing angle, can be written in terms of $\Delta \tau$ as

$$\theta_m = \sin^{-1} \left(\frac{c \Delta \tau}{d} \right). \quad (7.4)$$

The fact that AF is lower for other angles of incidence indicates that the phased array receiver possesses spatial selectivity. This can be extremely critical in wireless

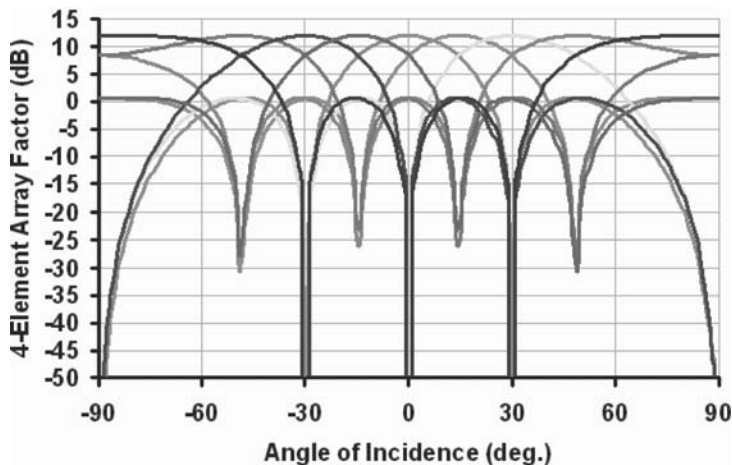


Fig. 7.2 4-channel receiver array factor for different values of $\Delta\tau$, namely $\frac{\pi}{4\omega}, \frac{2\pi}{4\omega}, \dots, \frac{8\pi}{4\omega}$. The inter-antenna spacing is assumed to be $\frac{\lambda}{2}$, where λ is the free-space wavelength at the frequency of operation ω .

systems that are interference limited - a strong interferer that is located in a direction that is different from the desired transmitter can be rejected by a phased array receiver. Fig. 7.2 shows AF versus the angle of incidence for different values of $\Delta\tau$, namely $\frac{\pi}{4\omega}, \frac{2\pi}{4\omega}, \dots, \frac{8\pi}{4\omega}$. A 4-element phased array receiver is considered with the inter-antenna spacing set to $\frac{\lambda}{2}$, where λ is the free-space wavelength at the frequency of operation ω . Each setting of $\Delta\tau$ results in a different beam-pointing angle. Hence, the changing of the value of the variable delay elements in each signal path allows us to electronically steer the beam¹.

From antenna theory, the beamwidth (in radians) of a radiating aperture is approximately equal to $\frac{\lambda}{D}$, where D is the width of the aperture. In the case of a linear phased array with a half-wavelength inter-antenna spacing, the total width is $(N-1)\frac{\lambda}{2}$, and hence the beamwidth is given by

$$\text{Beamwidth} \approx \frac{2}{N-1}. \quad (7.5)$$

An increase in the number of elements results in a narrower beam.

¹ $\frac{\lambda}{2}$ is the most commonly used inter-antenna spacing. A smaller spacing reduces the array's spatial selectivity. A larger spacing results in multiple main lobes.

As can be seen from the array patterns of Fig. 7.2, there are incidence angles where the received signal completely vanishes. These are called *nulls* and occur because the signals from the different channels cancel each other. In the formulation of this section, all channels are assumed to have the same gain. However, by modifying the gain of the individual channels, it is possible to arbitrarily set the location of the nulls, which is useful for interference cancellation. In addition, there are local maxima in the array pattern away from the main lobe. These are called *sidelobes* or *grating lobes*. The number of sidelobes increases with an increase in the number array elements. For the 4-element example in Fig. 7.2, there are two sidelobes, one on either side of the main lobe.

In addition to spatial selectivity, phased array receivers have the ability to improve the receiver sensitivity. In the direction of peak gain, a phased array achieves an addition power gain of N^2 over that of a single channel. However, if one assumes that each antenna picks up *uncorrelated* noise from the ambient surroundings, then the total output noise after the combining of the different channels is only N times larger than that of a single channel. Therefore, the signal-to-noise (SNR) ratio improves by a factor of N in an N -channel phased array. This SNR-improvement is also seen when the noise of the channel front-end dominates over the input noise (high noise figure receivers), as the front-end noise is uncorrelated across the different channels.

The array factor may also be viewed as the enhancement achieved over the antenna gain of a single element. In this context, phased arrays are also called *active antennas*, as the directionality of this enhancement can be electronically controlled. In the context of transmitters, the implementation of an N -channel phased array transmitter implies that each channel must generate $\frac{1}{N^2}$ times the power of a single-element transmitter to maintain the same received power level in the direction of maximum radiation. Thus, the total transmit power that needs to be generated is N times lower than the single-element case. This is specially significant in silicon-based technologies, specifically CMOS, as the low breakdown voltages render power generation challenging.

7.2.1 Case Study: A 60GHz WPAN Link Budget

As was mentioned in the introduction, there is a significant industrial and academic effort towards the deployment of 60GHz WPAN systems. One of the significant challenges for 60GHz systems is the low achievable link budget as free space path loss increases quadratically with carrier frequency. Additionally, as discussed, in the case of CMOS, the output power levels that are achievable in single transmitting elements are severely limited by the low breakdown voltages. In this case study, we examine the link budget at 60GHz and demonstrate how phased array transceivers alleviate the link budget requirement.

For indoor, dense, multipath environments, OFDM is the preferred modulation scheme due to its high spectral efficiency in these scenarios. Let us consider an OFDM signal composed of 192 data sub-carriers and 16 pilot sub-carriers with

a bandwidth of roughly 325MHz. Assuming a reasonable receiver noise figure of 10dB, the resulting sensitivity can be computed as

$$P_{noise} = -174 + 10\log(325) + NF = -79dBm. \quad (7.6)$$

On the transmitter side, typical 60GHz CMOS PAs achieve output-referred 1-dB compression points of around 6dBm only[3]. Furthermore, the substantial peak-to-average power ratio (PAPR) in OFDM requires significant backoff in the PA. Assuming a backoff of 6dB, the average transmitted power is around 0dBm. The free-space path loss for a 10m link is 88dB at 60GHz. Assuming that the transmitting and receiving antennas have a gain of 5dB and an implementation loss of another 5dB, the received power can be computed to be

$$P_{RX} = 0 - 88 + 5 + 5 - 5 = -93dBm. \quad (7.7)$$

Thus, this CMOS-based WPAN link budget is deep in the negative SNR regime by 14dB. However, phased array techniques can come to the rescue. If a 8-channel phased array is employed on the transmitter side, the effective transmitter antenna gain is enhanced by 18dB. A 4-channel receiver phased array enhances SNR by 6dB. If both are employed in the link, the SNR is boosted to a positive 10dB. In addition, phased arrays offer the advantage of being able to harness alternate reflection paths, such as reflections of walls, when the direct line-of-sight (LOS) path is broken.

7.3 Phased Arrays versus Timed Arrays

As was described in section 7.2, a phased array modifies the direction of transmission/reception of the electromagnetic beam by introducing a variable time delay in each antenna's signal path to compensate for the path differences in free space. Integrated variable time delay blocks are difficult to implement in practice, particularly on silicon. Therefore, in narrowband systems, the required variable time delay is often approximated with a variable phase shift, and the variable delay elements are replaced with phase shifters. The term *phased array* is actually a misnomer when variable delay elements are used and applies to arrays that use the narrowband approximation (Fig. 7.3(b)). A more appropriate term for an array that employs variable delay elements is *timed array* (Fig. 7.3(a)).

The validity of the delay-phase approximation in phased arrays naturally depends on the instantaneous bandwidth of the system. The approximation begins to fail when the instantaneous bandwidth of the system becomes large. Since timed arrays do not employ the delay-phase approximation, their functionality is theoretically unaffected by the signal bandwidth. In this chapter, we examine the phenomenon of

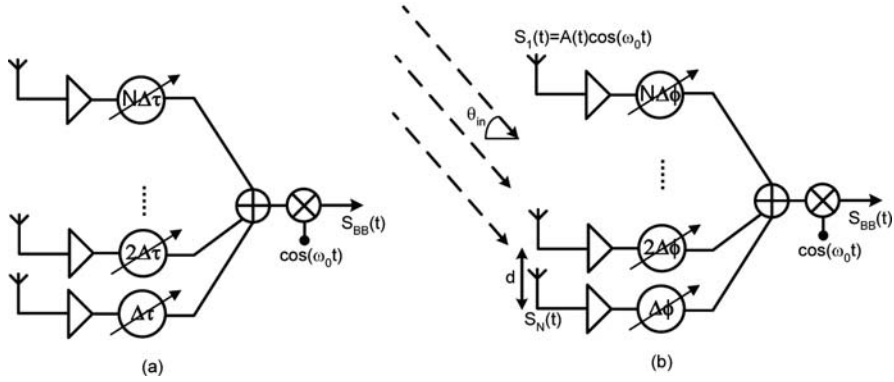


Fig. 7.3 (a) Timed array (b) Phased array.

Array-induced Inter-Symbol Interference in phased arrays, which is a direct result of this approximation.

Let us assume that an electromagnetic beam is incident on the phased array of Fig. 7.3(b) at an angle θ_{in} . If one assumes that the signal received by the first antenna is of the form $S_1(t) = A(t)\cos(\omega_0 t + \alpha(t))$, then the signal received by the i^{th} antenna can be written as

$$S_i(t) = A \left(t - (i-1) \frac{d \sin \theta_{in}}{c} \right) \cos \left(\omega_0 \left(t - (i-1) \frac{d \sin \theta_{in}}{c} \right) + \alpha \left(t - (i-1) \frac{d \sin \theta_{in}}{c} \right) \right). \quad (7.8)$$

Assuming that the antenna spacing, d , is equal to one half of the free-space wavelength at the frequency of operation, $S_i(t)$ reduces to

$$S_i(t) = A \left(t - (i-1) \frac{d \sin \theta_{in}}{c} \right) \cos \left(\omega_0 t - (i-1) \Delta\phi + \alpha \left(t - (i-1) \frac{d \sin \theta_{in}}{c} \right) \right), \quad (7.9)$$

where $\Delta\phi = \pi \sin \theta_{in}$. To receive this incident beam with maximum sensitivity, the phase shifters must be set to compensate for the free-space path difference. Specifically, $\Delta\phi_i$ must be set to $(i-1)\Delta\phi$. Assuming that the second harmonic generated by downconversion is filtered out, the resultant downconverted baseband signal then becomes

$$S_{BB}(t) = \sum_{i=1}^N \frac{A \left(t - (i-1) \frac{d \sin \theta_{in}}{c} \right) \cos \left(\alpha \left(t - (i-1) \frac{d \sin \theta_{in}}{c} \right) \right)}{2}. \quad (7.10)$$

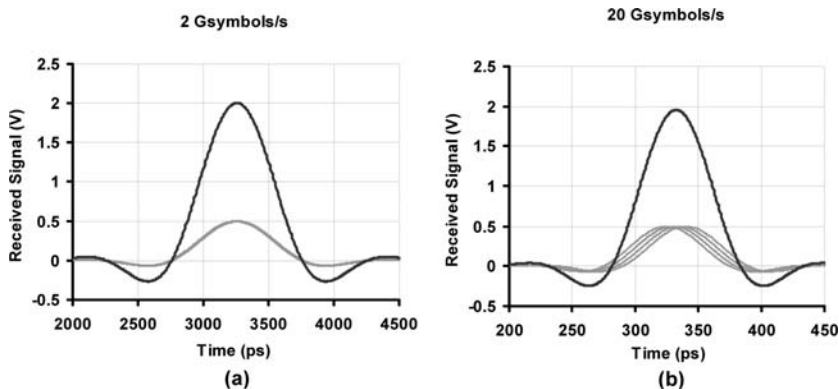


Fig. 7.4 Downconverted baseband pulses from each channel for different symbol rates. The electromagnetic beam is assumed to be incident at 36° to a 4-element 60GHz phased array.

Equation (7.10) reveals that the delay experienced by the modulation signal in reaching the different antennas is not compensated for, resulting in an Array-induced Inter-Symbol Interference effect. This is illustrated in Fig. 7.4, which depicts simulation results from a 4-element 60GHz linear phased array. An electromagnetic beam is assumed to be incident at 36° and the phase shifters are set appropriately. The incident beam is amplitude modulated with raised-cosine-filtered baseband pulses with a roll-off factor of 0.5. For simplicity, phase modulation is assumed to be absent. Different symbol rates (and hence signal bandwidths) are considered. Fig. 7.4 depicts the downconverted baseband contribution of each channel, and the resultant combined baseband pulse. When the symbol rate is 2G/s, the uncompensated delays in the baseband pulses are negligible when compared to the pulse-width. Hence, the baseband pulses from the different channels lie virtually on top of each other and no appreciable ISI is seen. On the other hand, when the symbol rate is 20G/s, the uncompensated delays now become a significant fraction of the pulse-width. The ISI effect now becomes noticeable and the combined baseband pulse is reduced in amplitude. Since the baseband pulse-width is dependent on the signal bandwidth, and the uncompensated delays are a function of the operating frequency, it follows that the severity of the ISI-effect is dependent on the *fractional bandwidth* of the signal, which is the ratio of signal bandwidth to carrier frequency. This underscores an important motivating factor behind the push to millimeter-wave frequencies : 5GHz of ISI-free bandwidth can easily be achieved in a 60GHz phased array, but would require a true-time-delay implementation if deployed in the 3-10GHz frequency range.

The reduction in the amplitude of the combined baseband pulse can also be viewed as a decrease in the *SNR*. This is because the total output noise power remains unchanged with the approximation of time delays with phase shifts, if one assumes that each antenna picks up uncorrelated noise from the ambient surroundings. To quantify this *SNR* decrease, different symbol rates are considered and the reduction in baseband pulse amplitude and hence *SNR* are computed for different angles of

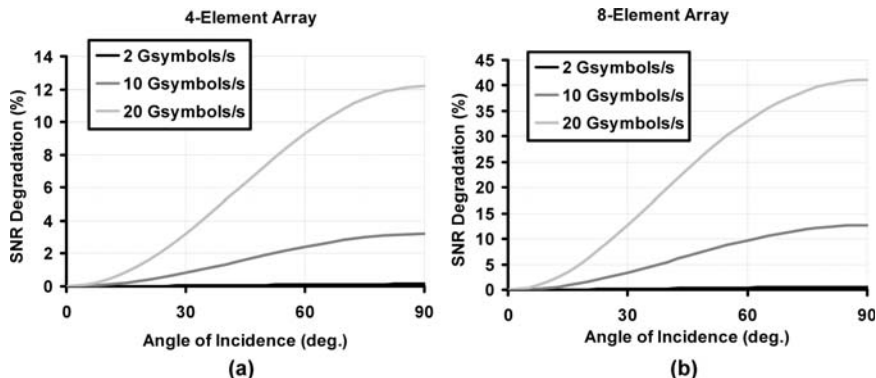


Fig. 7.5 *SNR degradation due to the phase-shift approximation for different angles of incidence and symbol rates in (a) a 4-element array (b) an 8-element array.*

incidence. The results are shown in Fig. 7.5. As is evident from the figure, the *SNR* degradation increases with the angle of incidence, as there is a larger inter-antenna delay that is uncompensated for in the information signal. The *SNR* degradation also increases with the symbol rate (and hence signal bandwidth) as was discussed in the previous paragraph. Finally, the *SNR* degradation due to the phase-shift approximation is larger for larger array sizes due to the larger delay between the first and the last antenna.

It is clear that signal bandwidths as large as 5GHz show insignificant *SNR*-deterioration due to the phase-shift approximation at 60GHz for the anticipated array sizes, thus eliminating the need for true-time-delay implementations. However, the 22-29GHz frequency band, allocated for short range automotive UWB radar systems, is an example of an application where a large signal bandwidth may necessitate the use of timed arrays. The radar range resolution is inversely proportional to the signal bandwidth; a signal with 5GHz bandwidth can achieve a theoretical range resolution of 3cm . A common radar signal waveform is a pulsed sinusoid. Consider a pulsed sinusoid with a center frequency of $f_0 = 25.5\text{GHz}$ and 200ps pulse width for an automotive radar system. The bandwidth of this signal is roughly $1/200\text{ps}$ or 5GHz. The beam width of an array with $\lambda/2$ spacing was given in (7.5). In order to have a beam width (spatial resolution) of 7.5° , a 16-element array is needed. The time delay between successive array elements, given by $\sin(\theta_{in})/2f_0$, is 16.9ps for $\theta_{in} = 60^\circ$. As can be seen from Fig. 7.6, there is no overlap between the signals of the last five elements (A12-A16) and the first element (A1) for the 60° incidence angle. A broadband phase shifter will align the sinusoids so that they add up coherently, only if they have some overlap. Delay elements are needed to shift the signals in time domain and align them. In this example, the 16-element array for the 60° incident angle will behave like a 12-element one *at best*! It should be reminded again that the need for variable true time delay element versus variable phase shifter is a function of signal fractional bandwidth, array size, and the maximum scanning angle. For a

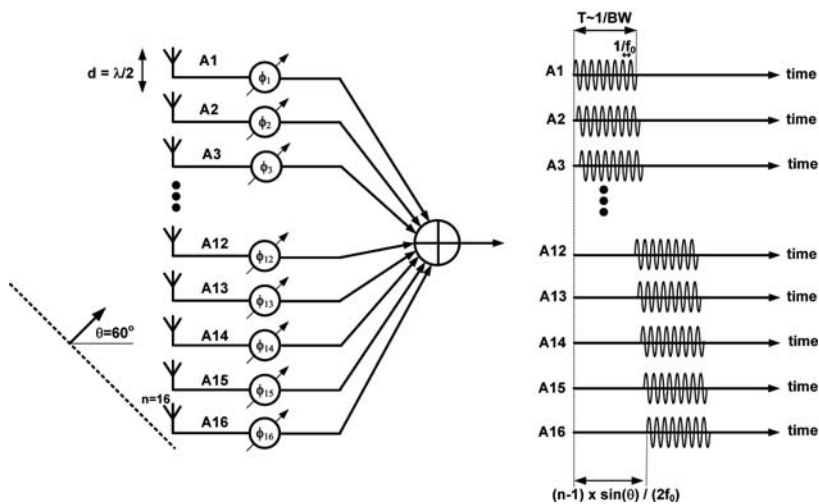


Fig. 7.6 The need for *timed arrays* in vehicular radar for 22-29GHz operation.

given fractional bandwidth, the larger the array size or the steeper the scanning angle, the more a variable true time delay element will be needed.

7.4 Conventional Phased Array Architectures

The phase-shifters required to achieve phased-array functionality can be incorporated in different parts of the transmitter/receiver chain. This results in three distinct phased array architectures - *RF Phase-shifting*, *LO Phase-shifting* and *Digital Arrays*. This chapter examines the trade-offs involved in these three architectures in detail². In addition, common phase-shifter circuits for the implementation of each of these architectures in a silicon process are presented from the existing literature.

² Although all schematics represent receiver arrays, the basic principles apply to transmitter arrays as well.

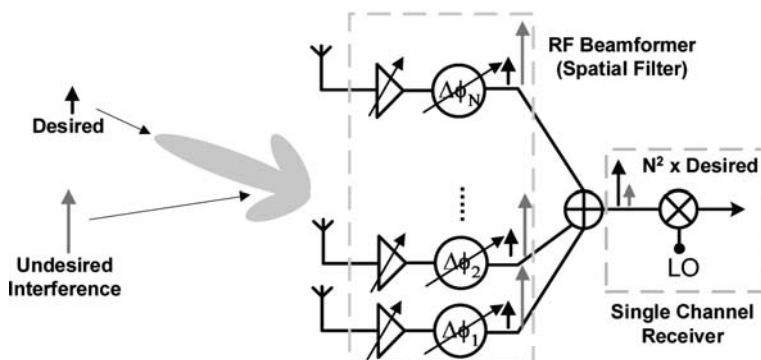


Fig. 7.7 The RF phase-shifting architecture

7.4.1 RF Phase-shifting

Fig. 7.7 illustrates the RF phase-shifting architecture for phased arrays. In this architecture, the signals in the various channel are phase-shifted and combined in the RF domain. The combined signal is then downconverted to baseband using any generic receiver such as heterodyne, homodyne or other image rejection architectures. RF phase-shifting has traditionally been the most widespread phased-array architecture because of its ability to insulate a larger portion of the receiver chain from strong, in-band interferers, as is shown in Fig. 7.7. A weak, desired signal and a strong, in-band interferer are assumed to be incident on the RF Phase-shifting array, with the interferer incident along a null direction. Since the combining point occurs prior to downconversion in this architecture, the interferer is cancelled prior to the down-conversion mixer. Therefore, the dynamic range requirements on the mixer and the blocks that follow it are alleviated.

The main challenge in this architecture is the implementation of RF phase-shifters in silicon. Passive implementations tend to be lossy while active phase shifters must be designed with sufficient linearity to accommodate strong interferers. An active phase shifter with insufficient linearity essentially negates the advantage that the RF phase-shifting architecture enjoys in interference rejection. In addition, the noise performance of the phase shifter is critical as the phase shifter lies in the RF signal path and hence can potentially degrade the system noise figure. Another design choice is the implementation of variable true-time delay elements for wideband timed arrays versus phase shifters for narrowband phased arrays.

As was mentioned earlier, control of the individual channel amplitudes is desirable as it allows for the modification of the null locations in the array pattern. Since the spatial filtering is done completely in the RF domain in this architecture, the variable-gain amplifiers required for individual amplitude control must be placed in the RF domain. This is challenging as RF blocks are usually parasitic-sensitive, and hence a change in the gain usually is accompanied by a change in the phase response.

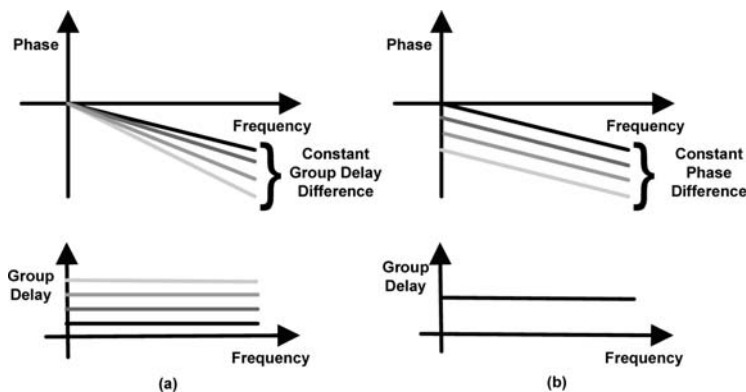


Fig. 7.8 (a) Phase and group-delay profile of a variable true-time delay element (b) Phase and group-delay profile of a broadband phase shifter.

Therefore, the RF variable-gain amplifiers must work in conjunction with the RF phase-shifters to set the null location while not altering the primary beam-pointing direction.

7.4.1.1 RF Variable True-Time Delay Elements

Variable true-time delay elements exhibit a constant group delay difference across different settings over frequency. Their use in timed arrays allows array functionality to be preserved independent of signal bandwidth. Fig. 7.8(a) shows the phase response of a variable true-time delay element. The response of a broadband phase shifter is shown in Fig. 7.8(b) for contrast. Broadband phase shifters maintain a constant phase difference between their different settings across the designed frequency range. As a result, phased arrays that employ these broadband phase shifters can only operate *in a narrow frequency range* around any center frequency in the designed frequency range.

Variable delay elements typically employ transmission-line structures. Since the delay of an electromagnetic wave in a transmission line depends on the wave velocity and the distance travelled, either one or both of these may be varied to control the delay of the RF signal. Fig. 7.9 summarizes mechanical techniques that are used to accomplish this. The *trombone line* varies the delay by mechanically changing the distance travelled by the RF signal. The name arises because its operation is similar to a trombone, where the position of a tuning slide is varied to change the tube length and hence the pitch. The wave velocity may be varied by controlling the separation between the signal and ground conductors or by modifying the nature of the dielectric between them.

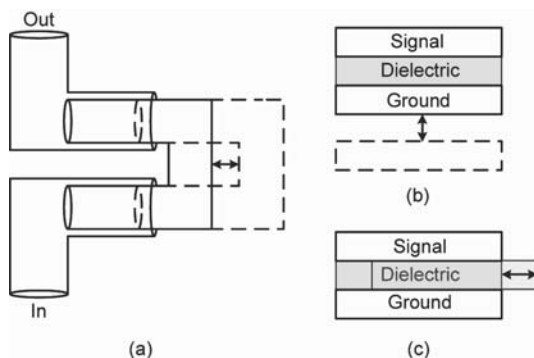


Fig. 7.9 (a) The mechanical trombone line manipulates the distance travelled by the signal (b) Manipulation of the wave velocity through control of the signal-ground separation (c) Manipulation of the velocity through control of the dielectric material between the signal and ground conductors.

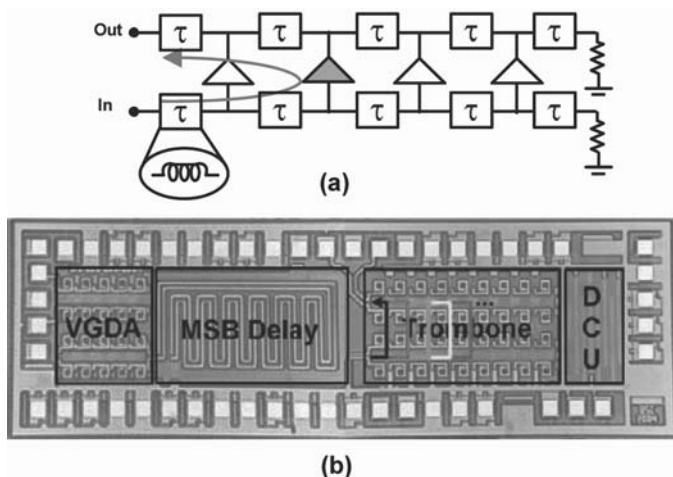


Fig. 7.10 (a) The trombone line delay element (b) A 1-13GHz UWB beamformer in a $0.18\mu\text{m}$ SiGe BiCMOS technology [5]. The beamformer employs a trombone line delay element (©IEEE 2006).

Fig. 7.10(a) depicts the electrical equivalent of the trombone line. The delay element consists of an input and an output transmission line. The transmission lines are typically implemented as lumped, quasi-distributed structures, and at each node, buffer amplifiers are employed to transfer the signal from the input to the output transmission line. At a time, only one buffer amplifier is activated and by switching between buffer amplifiers, the distance travelled by the electromagnetic signal and hence its delay are varied³. Fig. 7.10(b) is a chip microphotograph of a 1-13GHz

³ The circuit diagram of the trombone line bares a strong resemblance to that of a distributed amplifier with two key differences - the output terminal is on the other end of the output transmission line and only one buffer amplifier is on at a time.

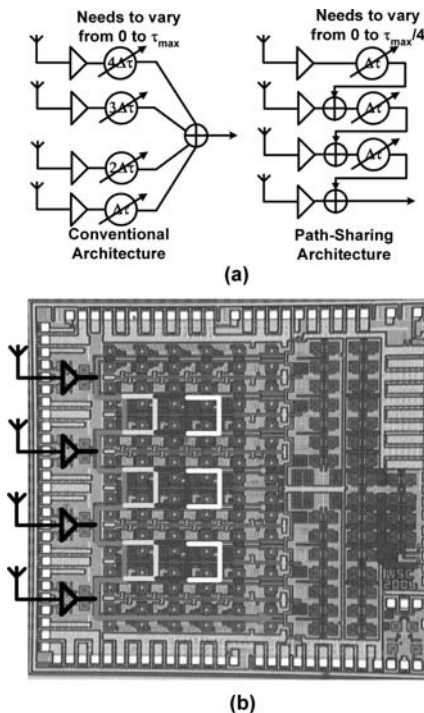


Fig. 7.11 (a) Path-sharing timed array architecture (b) A 4-channel, 0.13 μm CMOS UWB beamformer employing the path-sharing architecture (©IEEE 2007).

UWB beamformer that employs a trombone line delay element and is implemented in a 0.18 μm SiGe BiCMOS technology [5].

The component values for the trombone line are set by the desired characteristic impedance and delay/phase resolution. In Fig. 7.10(a), the sections of the quasi-distributed transmission line are low-pass in nature, with series inductors and shunt capacitors. The capacitors of each section are usually completely derived from the parasitic capacitances of the inductors and the buffer amplifiers. If the inductors have an inductance of L and the parasitic capacitance at each node is denoted by C_{par} , then the characteristic impedance of each line is $\sqrt{\frac{L}{C_{\text{par}}}}$. The delay resolution, or the delay difference between two successive buffer amplifier settings, is $\sqrt{LC_{\text{par}}}$. Therefore, if the desired delay resolution and characteristic impedance are known, the inductance value and tolerable parasitic capacitance at each node are fixed. The allowable parasitic capacitance at each node, coupled with the unity gain frequency (f_T) of the technology used, determines the amount of gain that can be achieved in the buffer amplifiers. For additional design guidelines and considerations, the reader is referred to [5]. In the UWB beamformer reported in [5], the lines are designed for

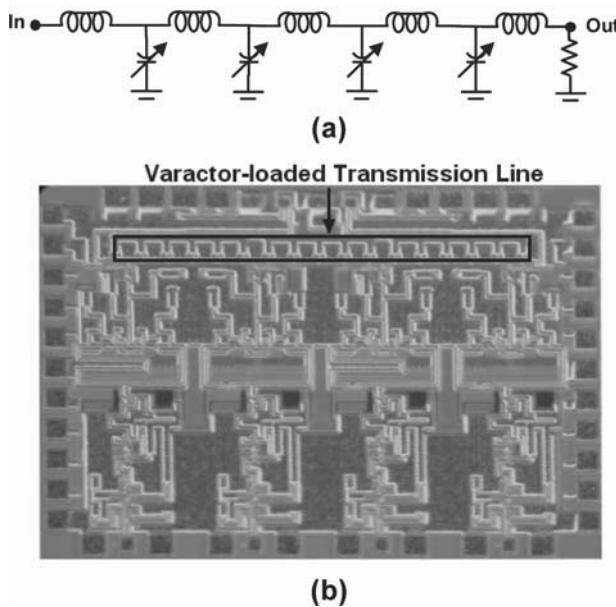


Fig. 7.12 (a) The varactor-loaded transmission line delay element. (b) A 4-element 60GHz phased array receiver in a $0.13\mu\text{m}$ SiGe BiCMOS technology employing varactor-loaded transmission line delay elements in a path sharing architecture [8] (©IEEE 2007).

a characteristic impedance of 100Ω (differential) and the delay resolution obtained is 4ps.

In order to reduce the area required by the trombone line delay elements, a *path-sharing* architecture has proposed for timed arrays in [6]. The architecture is depicted in Fig. 7.11(a) and takes advantage of the fact that a timed array requires a linear delay progression across the receiver channels. Each channel is delayed using a trombone line and is then combined with the adjacent channel, thus introducing the required delay progression. When compared with the conventional architecture, the maximum delay requirement from the trombone lines is reduced by a factor of N due to the series reuse of the lines. This results in significant area savings. Fig. 7.11(b) shows the chip microphotograph of the 4-channel, $0.13\mu\text{m}$ CMOS UWB beamformer reported in [6] that employs the path-sharing architecture. The architecture can be used with other delay elements or phase shifters as well to reduce the delay/phase requirements.

The *varactor-loaded transmission line* delay element as shown in Fig. 7.12(a). The capacitance of each section is realized in part through varactors. By varying the capacitance of the varactors, the wave velocity and hence the delay of the RF signal is varied. An unfortunate consequence of this is that characteristic impedance of the line also changes with a change in capacitance, which deteriorates the matching at the input and output terminals. Therefore, the extent of mismatch that is tolerable at

the terminals sets the allowable delay variation⁴. Another challenge that is specific to millimeter-wave implementations is the Quality Factor (Q) of varactors in silicon-based processes. Varactors, typically realized through MOS capacitors, exhibit poor Q values (<10 at 60GHz) at millimeter-wave frequencies[7], which results in high loss levels in the varactor-loaded transmission line. For instance, [8] reports a 4-element 60GHz phased array receiver in $0.13\mu\text{m}$ SiGe that employs the varactor-loaded transmission line delay element in the path-sharing architecture mentioned above. The path-sharing architecture reduces the requirement on each delay element to 45° of phase variation in the implementation reported in [8]. Over this phase variation range, the insertion loss varies from 2 to 4.5dB due to the loss of the varactors employed. The chip microphotograph of this phased array receiver is shown in Fig. 7.12(b).

The trombone line delay element is not a true passive delay element due to the presence of the buffer amplifiers, which limit the linearity performance. The varactor-loaded transmission line, on the other hand, is truly passive and hence highly linear⁵.

7.4.1.2 RF Phase Shifters

The first RF phase shifter considered in this subsection is the *switched transmission line* phase shifter is shown in Fig. 7.13(a). The phase shifter consists of multiple, quasi-distributed, transmission-line Π -sections, combined with MOSFET switches. When the digital control bit of the first section, b_1 , is high, switch M1 is open and M2 is closed. The inductor L_1 and capacitors of value C_1 then form a Π -section. Their values are chosen to yield a characteristic impedance of 50 ohms ($\sqrt{\frac{L_1}{2C_1}} = 50\Omega$) and an insertion phase of 180° , the highest phase-shift bit ($\omega\sqrt{2L_1C_1} = 180^\circ$). When b_1 is low, L_1 is shorted by switch M_1 . Furthermore, switch M_2 is open and its capacitive parasitics are resonated out by inductor L_{s1} . As a result, the capacitances of value C_1 are rendered ineffective and the Π -section as a whole is bypassed. A cascade of multiple, quasi-distributed, Π -sections with bypass-capability corresponding to the different phase-shift bits completes the phase shifter design. It should be noted that while this phase shifter appears to be a true-time delay element, the narrowband match between L_{s1} and the parasitics of M_2 limits its performance to that of a phase shifter.

Since the MOSFETs are used only as passive switches, this phase shifter exhibits good linearity performance. The main challenge in the design of the switched transmission line phase shifter is the loss associated with the MOSFET switches. In [9], the authors report a 4-bit switched transmission line phase shifter operating from 30-

⁴ Alternately, the load impedance can be varied to maintain matching. However, this involves incorporating input impedance tunability in the following stage, and it would have to be accomplished without modifying the phase response of that stage.

⁵ The varactors in the transmission line have a voltage-dependent capacitance that is a source of nonlinearity. However, for this delay element, compression of the output amplitude and variations in the output phase tend to occur only at extremely high power levels and are seldom sources of concern.

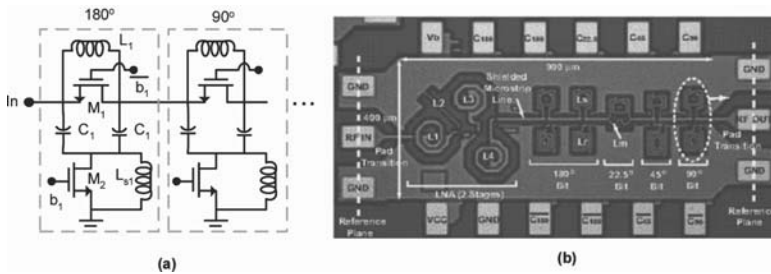


Fig. 7.13 (a) Switched transmission line phase shifter concept (b) Chip microphotograph of a 30-38GHz LNA and switched transmission line phase shifter implemented in a 0.13 μm SiGe BiCMOS process [9] (©IEEE 2007).

38GHz and implemented in a 0.13 μm SiGe BiCMOS technology. The phase shifter is preceded by an LNA that exhibits 15dB of gain at 34GHz. The insertion loss of the phase shifter is reported to be around 13dB from 30-38GHz and the overall insertion loss of the LNA-phase shifter combo is reported to be $1 \pm 1.5\text{dB}$ across the different phase-shift states at 34GHz, including the input and output pads. Fig. 7.13(b) shows the chip microphotograph of the implemented LNA and phase shifter.

The *high-pass/low-pass* phase shifter functions by taking the difference between the phase response of a high-pass filter path and a low-pass filter path. Fig. 7.14(a) shows the schematic of the first bit of such a phase shifter. Each individual bit has a high-pass and low-pass paths, with Single-Pole-Double-Throw (SPDT) switches on both sides. Depending on the switch states, either the high-pass or the low-pass path is inserted. Therefore, across the switch states, a phase difference is created that is equal to the difference between the insertion phases of the high- and low-pass paths. Each individual bit is designed for different phase differences.

The filter elements may be designed for flat gain in the frequency band of operation and a linear phase response. The challenges include the losses in the passive elements and the switches. Additionally, the switch may increase the phase variation associated with the phase shifter. For design guidelines, the reader is referred to [10]. Fig. 7.14(b) shows the chip microphotograph of a 12GHz phase shifter implemented in a SiGe process that employs the high-pass/low-pass phase shifter for the highest (180°) bit [11].

The high-pass/low-pass phase shifter is fundamentally passive and hence exhibits good linearity performance. However, if the SPDT switches are implemented in an active manner to minimize insertion loss, a large amount of current may need to be burnt in the switches to maintain linearity performance.

Another common RF phase shifter is the *reflection-type phase shifter* (RTPS), which employs a 4-port directional coupler and purely reflective (i.e. imaginary) loads. Fig. 7.15(a) depicts an example, where the directional coupler is implemented as a branchline 3dB 90° coupler. The *through* and *coupled* ports are terminated with the reflective terminations and the isolated port is used as the output. The incident signal at the input reaches the output only through reflections at the imaginary

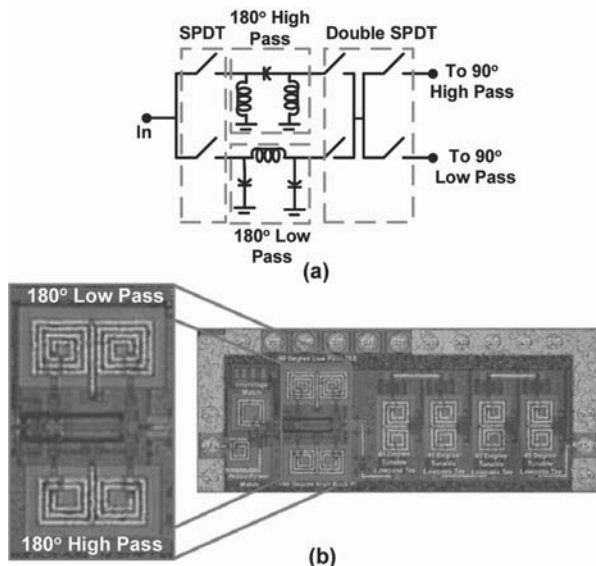


Fig. 7.14 (a) Schematic of the first bit of a High-Pass/Low-Pass phase shifter (b) A 12GHz SiGe phase shifter employing the High-Pass/Low-Pass phase shifter for the highest (180°) bit of phase shift (©IEEE 2005).

terminations. Therefore, by varying the value of the imaginary terminations, the phase of the reflection coefficient at the through and coupled ports and hence the phase of the output are varied.

If the imaginary impedances at the *through* and *coupled* ports are denoted by jX , then the total phase shift from the input to the output is given by $-\frac{\pi}{2} - 2 \tan^{-1}(\frac{Z_o}{X})$, where Z_o is the impedance to which the ports are matched. Fig. 7.15(b) plots this phase shift as a function of capacitance when the reflective terminations are implemented through varactors. The operating frequency and port impedance are assumed to be 60GHz and 50Ω respectively. To achieve a total phase variation of 180°, the termination capacitance must vary from 0 to ∞ , which obviously cannot be realized through any physical varactor. Therefore, to maximize the achievable phase shift range, higher order terminations are employed, such as series LC networks. A detailed discussion on reflective termination design may be found in [12].

The RTPS, being fundamentally passive, shows excellent linearity characteristics. The main challenge in integrated, silicon-based RTPS design is loss. The two main sources of loss in an RTPS are the transmission-line losses in the directional coupler and the loss in the reflective terminations. The latter is specially significant at millimeter-wave frequencies due to the limited Q of varactors. To reduce phase shifter loss, active negative resistance circuits have been used in RTPS designs [13]. This, however, limits the linearity and noise performance of the RTPS.

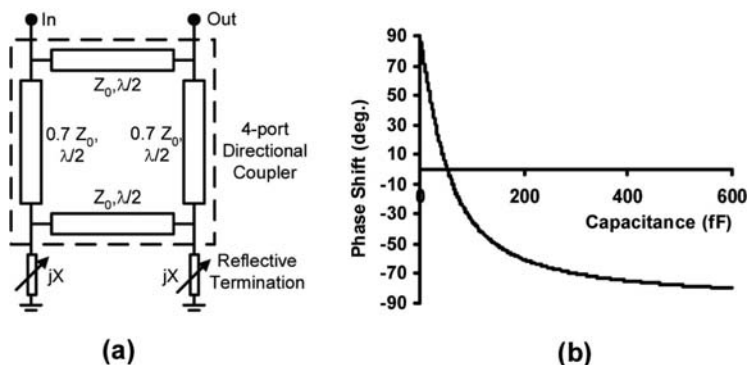


Fig. 7.15 (a) Reflection-type phase shifter (RTPS). The directional coupler in this case is implemented as a branchline 3dB 90° coupler. (b) RTPS phase shift as a function of termination capacitance.

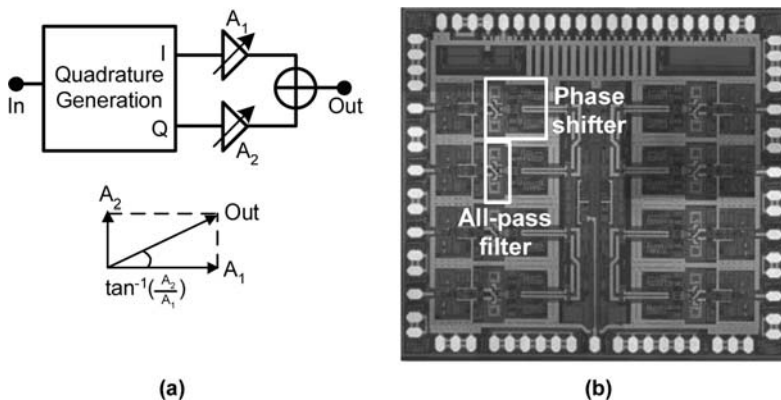


Fig. 7.16 (a) The vector-summing phase shifter (b) Chip microphotograph of an 8-element, 6-18GHz, phased-array receiver implemented in a 0.18 μm SiGe process (©IEEE 2007).

The final phase shifter discussed in this subsection is the *vector-summing phase shifter*, illustrated in Fig. 7.16(a). The input RF signal is split into two components of equal power and 90° phase difference using a quadrature generation block. The two components are then amplified with variable weights using variable-gain amplifiers and combined. If A_1 and A_2 are the weights imparted to the 0° and 90° components, the phase of the output is $\tan^{-1} \frac{A_2}{A_1}$. Through appropriate choice of A_1 and A_2 , it is possible to achieve an arbitrary phase at the output while maintaining a constant small signal gain. This phase shifter is also called a *phase interpolator*, because, in effect, an interpolation is performed between the 0° and 90° components. Other names include *Polar Modulator* and *Cartesian Combiner*.

The quadrature generation block can be implemented in a number of ways. Examples include the 90° hybrid coupler and quadrature all-pass filters (QAF). In [14], the authors report an 8-element phased-array receiver front-end that employs QAF-based phase interpolators as broadband phase shifters. A second-order all-pass network is used to generate the quadrature signals and 4 bits of variation is achieved in the phase shifter. The chip microphotograph of the 6-18GHz phased-array receiver is shown in Fig. 7.16(b). Other published works that employ this phase shifter include a 1GHz implementation in 1998 [15], a 2.4GHz system [16] and a 60GHz phased array [17].

Being an active phase shifter, the linearity performance is poor and a large power consumption is usually required to achieve a high dynamic range.

7.4.2 LO Phase-shifting

Fig. 7.17 displays the LO phase-shifting architecture for phased arrays. Since the mixing of the RF signal with the LO essentially results in a subtraction of their phases, phase-shifting of the LO of each signal path is equivalent to phase-shifting the RF signal. The advantage of this architecture over the RF phase-shifting approach is that the phase-shifters are removed from the RF signal path. As a result, the nonlinearity, loss and the noise performance of the phase-shifters no longer have a direct impact on the system performance. However, as is depicted in Fig. 7.17, strong, in-band interferers are cancelled only after the combining point, which occurs after the downconversion mixers. As a result, the mixers must have sufficient dynamic range to withstand the interferers, which usually requires a large power dissipation.

Any of the RF phase shifters presented in the previous subsection may be used in the LO path to phase shift the LO signal for each channel. In general, the performance requirements on LO-path phase shifters are more relaxed when compared to RF-path phase shifters, and hence they can be expected to consume less area/power. Other techniques, including tuned ring oscillators and coupled oscillator arrays, are also available for this architecture.

The delay-phase approximation is inherent in the LO phase-shifting architecture as the LO is a single tone and cannot compensate for the delay progression of the received signals over a wide bandwidth.

7.4.2.1 Tuned Ring Oscillators

Tuned ring oscillators can be used to generate the multiple LO phases required for this architecture [18],[19]. The tuned ring oscillator consists of a number of amplification stages, with each cell comprised of a transconducting gain cell driving a resonant load. Fig. 7.18(a) illustrates an N -element CMOS tuned ring oscillator where the transconducting gain cell is realized as a differential pair. Prior to the closing of the ring, a phase inversion is introduced. To ensure that the total phase-shift in the ring equals 0, each element must now sustain a phase-shift of $\frac{180^\circ}{N}$ across itself. Thus, the

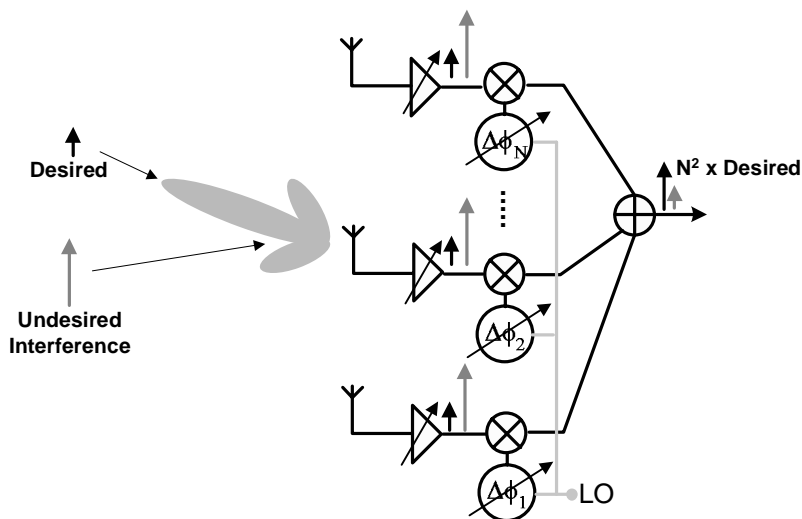


Fig. 7.17 The LO phase-shifting Architecture

different element outputs represent different phases of the oscillation frequency in steady-state. Fig. 7.18(b) depicts the different element outputs of an 8-element tuned ring, with the inter-element phase shift being $\frac{180^\circ}{8} = 22.5^\circ$. It should be noted that in an N -element tuned ring, not N but $2N$ LO phases are available, spanning the entire $0^\circ - 360^\circ$ range. This is because, due to the differential nature of the tuned ring, each element inherently offers a phase *and* its inverse.

The oscillation frequency of the tuned ring does not coincide with the center frequency of the resonant loads. In fact, the tuned ring must operate off the center frequency of the resonant loads so that each element may provide the requisite phase shift. This leads to a deterioration in oscillation amplitude and hence, phase noise, as the impedance presented by a resonant load is maximum at its center frequency. The extent of this deterioration is dependent on the inter-element phase shift, which is inversely proportional to N , the number of elements. For a detailed discussion on the phase noise of the tuned ring oscillator and design guidelines, the reader is referred to [20].

A practical challenge in the use of tuned ring oscillators is the routing of the element outputs to the different antenna signal paths. Fig. 7.18(c) depicts the chip microphotographs of a 4-channel 24GHz transmitter in $0.13\mu\text{m}$ CMOS [19] and an 8-channel 24GHz receiver in $0.18\mu\text{m}$ SiGe [18]. The phase distribution network tends to occupy significant chip area, and can be substantially lossy. This increases the current requirements on the oscillator buffers to maintain the desired LO amplitude. Furthermore, it is critical that the distribution network be perfectly symmetric so that symmetry of the phases is not disturbed. Design considerations for the phase distribution network may be found in [21].

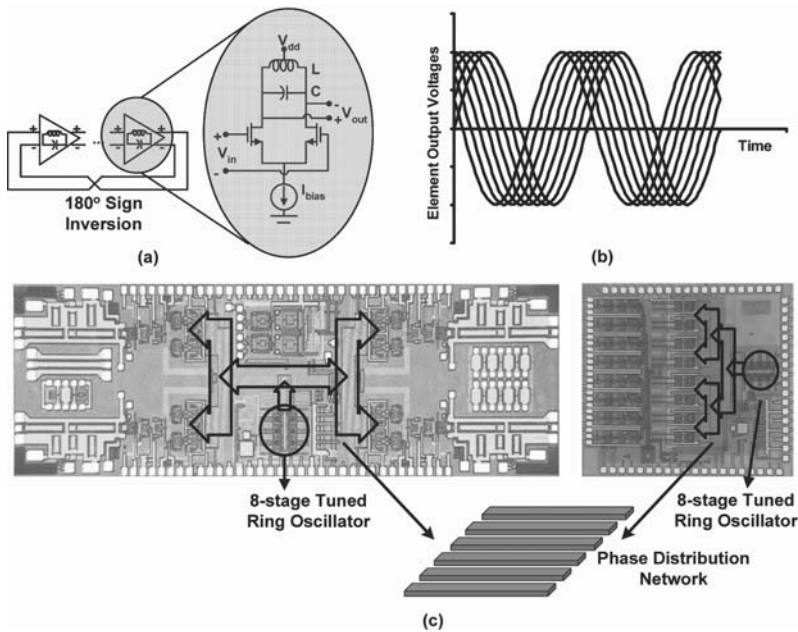


Fig. 7.18 (a) The tuned ring oscillator (b) Element voltages in an 8-element tuned ring oscillator (c) Chip microphotographs of a 4-element 24GHz phased array transmitter in a $0.18\mu\text{m}$ CMOS process ([19]) (©IEEE 2005) and an 8-element 24GHz phased array receiver in a $0.18\mu\text{m}$ SiGe process ([18]) (©IEEE 2004). Both employ an 8-element tuned ring oscillator in an LO phase-shifting architecture.

7.4.2.2 Coupled Oscillator Arrays

Another means of generating phase-shifted LO signals involves the coupling of multiple free-running oscillators together to form a *coupled oscillator array* (COA). The dynamics of COAs have been extensively studied in the past ([22], [23]), but the implementations have traditionally employed discrete, compound-semiconductor transistors. The usage of COAs for integrated, silicon-based phased arrays involves several design challenges and nuances.

The principle of operation of COAs is based on the concept of injection-locking, which was first dealt with in the electrical domain by Robert Adler [24]. When an electrical oscillator is injected with an external signal whose frequency is close to the free-running frequency of the oscillator, the oscillator becomes synchronized in frequency to the external signal. Fig. 7.19 depicts a traditional nMOS, cross-coupled, LC oscillator with an external signal injected into the LC tank in the current domain. If the free-running frequency of the LC oscillator is $\omega_0 = \frac{1}{\sqrt{LC}}$, and the frequency of the external signal is ω_{inj} , then the oscillator locks to ω_{inj} if the offset frequency $\Delta\omega_{inj} = |\omega_{inj} - \omega_0|$ lies within a locking range $\Delta\omega_{lock}$ given by

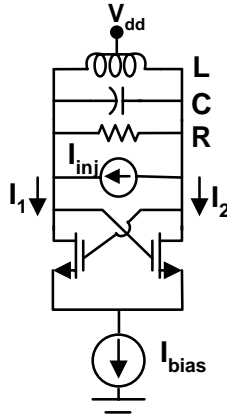


Fig. 7.19 Cross-coupled nMOS LC oscillator with an external signal injected into the LC tank in the current domain.

$$\Delta\omega_{lock} = \frac{\omega_0 \epsilon}{2Q}, \quad (7.11)$$

where Q , the Quality Factor of the LC tank, is given by $\frac{R}{\omega_0 L}$. ϵ is called the injection gain and is defined as the ratio of the amplitude of the injected current to the amplitude of the fundamental component of the oscillator's differential current $\frac{I_1 - I_2}{2}$. A larger injected current leads to a larger locking range. It should be noted, however, that this result assumes *weak* injection ($\epsilon \ll 1$). The phase difference $\Delta\phi$ between the locked oscillator and the injected signal is given by

$$\Delta\phi = \sin^{-1} \left(\frac{\omega_{inj} - \omega_0}{\Delta\omega_{lock}} \right). \quad (7.12)$$

When injection occurs at the free-running oscillation frequency, there is no phase difference. When injection is performed at the locking range boundaries, a 90° phase difference is seen.

The concept of injection locking can be extended to a linear array of coupled oscillators with nearest-neighbour coupling, as depicted in Fig. 7.20(a). N free-running oscillators are arranged in a linear array and each oscillator is injected with the outputs of its nearest neighbours. In the circuit diagram of Fig. 7.20(a), the individual oscillators are implemented as LC oscillators with cross-coupled, nMOS negative- g_m cells and the outputs of the nearest neighbours are injected in the current domain using differential pairs. If all oscillators are identical with a center-frequency of ω_0 and an LC tank quality factor of Q , and the edge elements are detuned in frequency by $\Delta\omega$ as shown in Fig. 7.20(a), then the oscillator outputs become synchronized to a common oscillation frequency of ω_0 and exhibit a linear phase progression. The

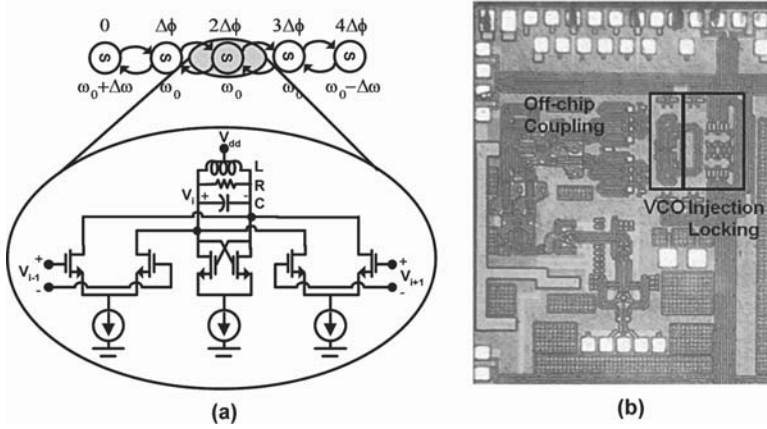


Fig. 7.20 (a) A linear array of coupled oscillators with nearest-neighbour coupling (b) Chip microphotograph of each transmitter element in the scalable, 60GHz phased array transmitter reported in [25] (©IEEE 2006). Injection locking is utilized only for the synchronization of multiple chips to a common reference. The phase shifts required for beam-steering are obtained from RF phase shifters.

successive phase difference $\Delta\phi$ is given by

$$\Delta\phi = \sin^{-1} \left(\frac{2Q\Delta\omega}{\omega_0\epsilon} \right), \quad (7.13)$$

where ϵ , called the coupling gain, is the ratio of the injected current amplitudes to the amplitude of fundamental current in the free-running oscillators. The presence of a *controllable*, linear phase progression at the outputs of the oscillators implies that each oscillator output needs to be routed to a single phased array channel for beam-steering. In other words, no phase distribution network is necessary. The mathematical formulation and basis for this linear phase progression may be found in [23]. It should be mentioned that, as was the case with injection locking, this behavior is predicated on the assumption of *weak* injection ($\epsilon \ll 1$).

The main challenge in the implementation of silicon-based, integrated phased arrays employing COAs is the sensitivity of COAs to element mismatches and process variations. If the center-frequencies of the oscillators deviate from the nominal value of ω_0 due to process variations and mismatches, the linear phase progression is disturbed. If the deviations are large, the oscillators may not even lock to each other. The case of a single oscillator locked to an external injected signal may be taken to illustrate the level of sensitivity. For a nominal center frequency of 60GHz, an injection gain of 0.1 and a Quality Factor of 15, which is typical for silicon processes at these frequencies, the injection locking range may be computed to be 200 MHz from (7.11). Therefore, a shift in the center frequency of the LC tank by 200

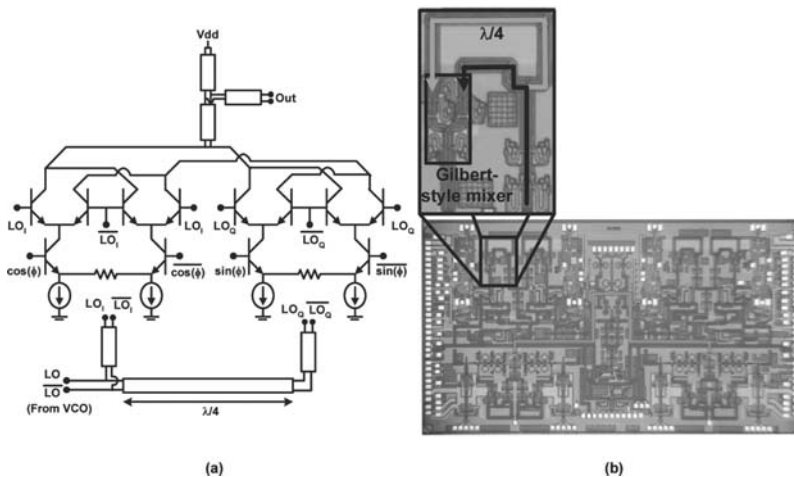


Fig. 7.21 (a) Circuit diagram of the local LO path Phase Interpolators used in [26] (b) Chip microphotograph of the 77GHz, 4-element phased array transceiver reported in [26] (©IEEE 2006).

MHz can prevent locking from occurring. Even if the center-frequency variations are largely below 200 MHz, the phase shift between the locked oscillator and injection reference would vary significantly. In [25], the authors report a scalable, 60GHz phased-array transmitter that employs injection locking only for the synchronization of multiple chips to a common reference. The phase shifts required for beam-steering are obtained from RF phase shifters. The chip microphotograph of this transmitter is depicted in Fig. 7.20(b). The design and use of calibration circuitry that can detect and correct on-chip variations is an interesting line of research that could render COAs a viable option.

7.4.2.3 LO Path Phase Interpolators

In [26], the authors report a 77GHz 4-element phased array transceiver employing an LO phase-shifting architecture, and the phase-shifting of the LO is accomplished through *local* phase interpolators. The circuit diagram of the local phase interpolators is shown in Fig. 7.21(a), and the chip microphotograph of the transceiver is depicted in Fig. 7.21(b). The quadrature LO signals are generated using a quarter-wavelength transmission line and the weights for the quadrature components are applied through doubly-balanced, Gilbert-type mixers. The weighted quadrature signals are then combined in current domain.

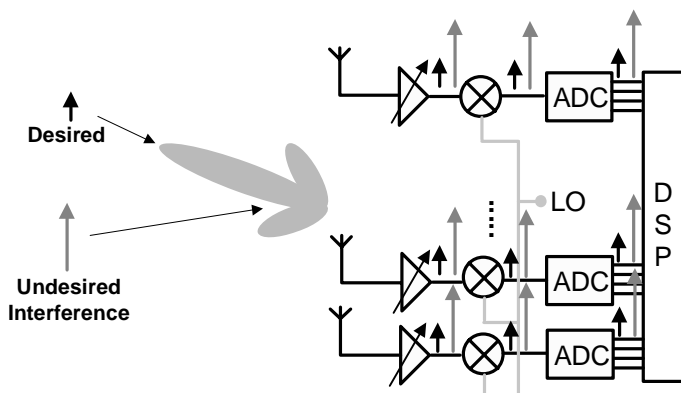


Fig. 7.22 The Digital Array Architecture

7.4.3 Digital Arrays

In the digital array architecture, which is depicted in Fig. 7.22, each phased-array channel is digitized using an Analog-to-Digital Converter (ADC) and the bits of all channels are then processed using a Digital Signal Processing unit (DSP), where the spatial filtering is performed. Therefore, strong interferers get spatially cancelled only after digital signal processing, and hence the RF mixer and ADC of each channel and the DSP unit must have sufficient dynamic range to handle the interferers. Furthermore, virtually the entire RF chain is replicated for each channel. These two factors result in a rather power-hungry design.

The main advantage of the digital array is its versatility. A wide variety of complex, signal-processing algorithms can be implemented using DSP units. Such phased arrays are also called *smart antennas* and are extensively used in the cellular-phone industry. These algorithms allow the smart antenna to distinguish among desired signals, multipath and interfering signals, as well as to calculate their directions of arrival. In addition, smart antennas can adaptively update their beam patterns, so as to track the desired signal with the beam's main lobe and track the interferers with nulls. Multi-beam and multiple-input-multiple-output (MIMO) functionality, which will be discussed in greater detail later in this chapter, can also be incorporated into smart antennas. For a tutorial on smart antennas and common signal processing algorithms that they employ, the reader is directed to [27].

7.4.4 Comparative View of the Conventional Architectures

Table 7.2 presents a summary and comparison of the different phased array architectures detailed in this section. The true-time delay elements allow timed arrays to

Table 7.2 Comparison of conventional phased array architectures.

Arch.	Phase Shifter	Bandwidth/ Data-rate	Interference Cancellation	Area	Power
RF Path	Trombone Line	High (Time Delay)	Moderate	High	Moderate
	Varactor-loaded	High (Time Delay)	Good	High	Low
	Trans. Line				
	Switched Trans. Line	Moderate (Phase shift)	Good	High	Low
	High-pass/low-pass	Moderate (Phase shift)	Good	High	Low
YJ	RTPS	Moderate (Phase shift)	Good	High	Low
	Phase Interpolator	Moderate (Phase shift)	Moderate	Moderate	Moderate
LO Path	Multiphase LC Oscillator	Moderate (Phase shift)	Moderate	Moderate	Moderate
	Coupled Oscillator Array	Moderate (Phase shift)	Moderate	Moderate	Moderate
	LO-path Phase Interpolation	Moderate (Phase shift)	Moderate	Moderate	Moderate
BB Path	Digital Arrays	High	Poor	High	High

function over wide signal bandwidths, which implies high data rates for wireless communication applications and fine range resolution in radar. Phase-shifter based architectures on the other hand can operate only over moderate bandwidths. The all-passive RF phase shifters and delay elements exhibit the greatest interference cancellation ability. The active RF phase shifters, delay elements and LO phase-shifting architectures demonstrate only moderate interference rejection. A larger amount of power would need to be consumed in the building blocks for these architectures to withstand strong interferers. Digital arrays are the poorest in this regard, and an extremely large amount of power would need to be spent in boosting the dynamic range of the mixers and ADCs of each channel.

In terms of area, the RF phase shifters and delay elements tend to be the most demanding. An exception is the RF phase interpolator, which can be somewhat compact. It should be noted that the path-sharing architecture may be employed to alleviate the area requirement of an RF phase-shifting architecture. The LO phase-shifting architectures have moderate area requirements - the phase shifters in the LO path can often be implemented in a small area as their linearity and noise figure performances are not critical. However, the architecture does require a mixer for each channel, which increases the area requirement somewhat. The digital array architecture requires a large area despite the lack of phase shifters due to the need for a mixer and an ADC in each channel.

In terms of power consumption, the RF phase-shifting architectures are the most efficient, particularly when the phase shifters/delay elements are passive. This is

because the architecture only requires one mixer. However, it should be noted that if the phase shifter/delay element loss is high, RF amplification stages may be required which may increase the power consumption. The LO phase-shifting architectures have moderate power consumption requirements primarily because of the need for a mixer in each signal path. The power consumption of the mixers is often not negligible as they need to withstand strong, in-band interferers. The buffers needed to distribute the LO to the different channels also contribute to the power consumption. The digital Array architecture is the most power hungry of the three due to the need for high-dynamic-range mixers and ADCs in each channel.

7.5 The VPRO-PLL Phased Array Architecture

A more recently developed phased array architecture is the Variable-Phase Ring Oscillator (VPRO) and Phase-Locked Loop (PLL) architecture proposed by the authors of this chapter. The architecture was first introduced in [28] and is unlike the conventional architectures delineated in the previous section. The architecture eliminates key building blocks such as mixers, phase shifters and power splitters/combiners, allowing for compact and low-power implementations. This section gives a broad description of the principle of operation of the architecture⁶.

7.5.1 VPRO Concept

At the core of the architecture lies the Variable-Phase Ring Oscillator or VPRO. The VPRO (Fig. 7.23(a)) consists of a number of elements connected in a ring configuration, with each element comprising a nonlinear gain block driving a tuned load. Fig. 7.23(b) shows a MOS implementation of each element, where the gain block is realized through a differential pair. The VPRO is similar to the conventional tuned ring oscillator discussed earlier in this chapter save for the fact that an electrically tunable phase-shifter is introduced in the ring. The phase boundary condition dictates that the total phase shift in the ring must be an integral multiple of 2π . Therefore, the phase shift across each element (from V_{i-1} to V_i) is given by $\Delta\phi = \frac{2k\pi - \Phi_{ext}}{N}$, $k \in \mathbb{Z}$. Hence, an electrically tunable linear phase progression is established, which is the requirement for beam-steering. It must be noted that to sustain this phase shift across each element, the oscillator functions at a frequency that is off the LC center frequency. This depicted qualitatively in Fig. 7.23(c), which shows the oscillation frequencies for different inter-element phase shifts and an LC center frequency of 24GHz.

⁶ Portions of this text are taken from [30], ((c) IEEE 2008).

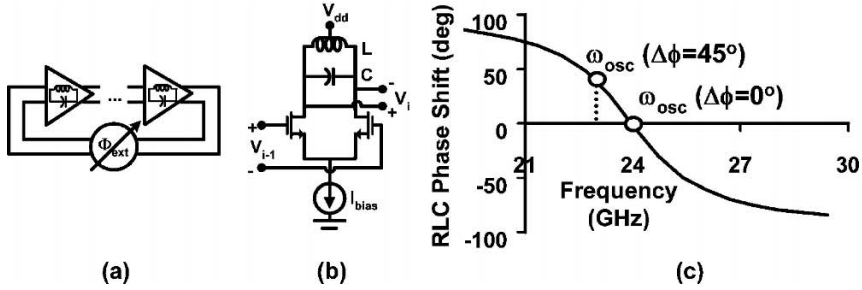


Fig. 7.23 (a) Block diagram of the Variable Phase Ring Oscillator (VPRO) (b) MOSFET-based differential pair implementation of a VPRO element (c) Dependence of VPRO oscillation frequency (ω_{osc}) on the inter-element phase shift ($\Delta\phi$).

7.5.2 Transmit Mode

The presence of a linear phase progression across the elements of the VPRO implies that, in the transmit mode, the output of each element can simply be connected to an antenna to accomplish beam-steering. No phase-distribution network is required. Fig. 7.24 shows the block diagram of the VPRO-PLL architecture in TX mode, with each element's output connected to a power amplifier and antenna.

As was mentioned before, when the external phase-shifter's setting is changed to steer the beam, the oscillation frequency of the VPRO changes as the phase shift across each tuned load changes. The incorporation of a PLL around the VPRO, shown in Fig. 7.24 with frequency dividers, phase-frequency detector (PFD), charge pump (CP) and a loop filter, ensures that the operating frequency remains constant while maintaining the desired phase progression.

The modulation of information onto the carrier may also be accomplished through the PLL. A signal injected in the current domain (I_{in}) into the loop filter in parallel with the charge pump's output current gets phase modulated (PM) onto the carrier as shown in Fig. 7.24. Fig. 7.24 also depicts the small signal model of the PLL, with K_{vco} , N_{div} , K_{PD} and $F(s)$ representing the VPRO tuning gain, frequency division ratio, PFD-CP gain and loop filter transfer function respectively. θ_{in} and I_{in} are the reference phase and injected modulation current respectively. The VPRO phase θ_{vpro} may be computed to be

$$\theta_{vpro}(s) = \frac{K_{PD}K_{vco}N_{div}F(s)}{sN_{div} + K_{PD}K_{vco}F(s)}\theta_{in}(s) + \frac{K_{vco}N_{div}F(s)}{sN_{div} + K_{PD}K_{vco}F(s)}I_{in}(s). \quad (7.14)$$

From the expression above, it is clear that the injected current gets directly translated to the phase of the VPRO through a transfer function. The bandwidth of the VPRO-PLL architecture in transmit-mode is limited by this transfer function.

This PLL-phase modulation approach is similar to the *translational loop* used extensively in polar transmitters for the GSM standard. The nonlinearity of the oscillator's K_{vco} profile can cause variations in the modulation gain and bandwidth

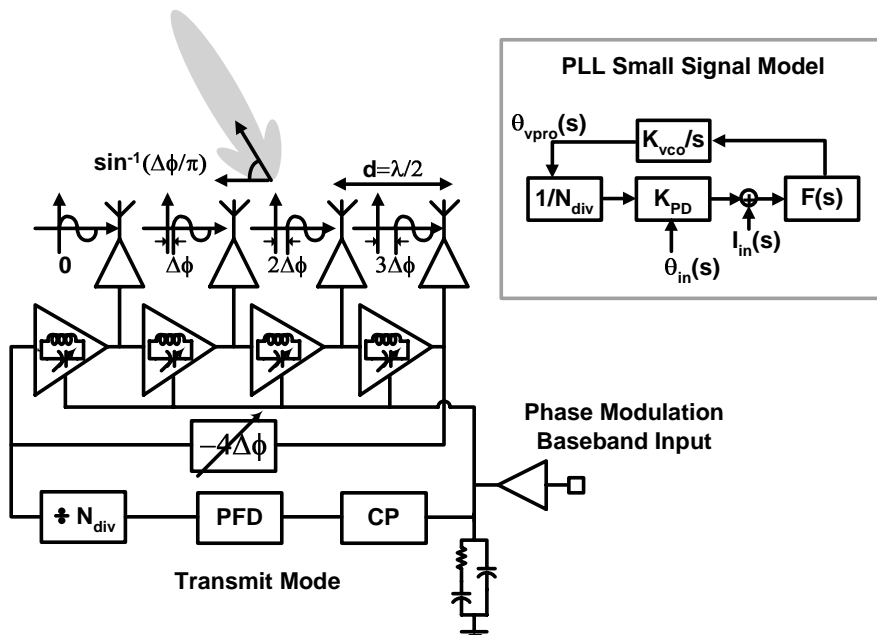


Fig. 7.24 Principle of operation of the VPRO-PLL architecture in transmit (TX) mode.

and self-calibration techniques have been developed for GSM polar transmitters to tackle this problem. Such techniques may be applied to the VPRO-PLL architecture as well. Support for amplitude modulation schemes, such as QAM, can also be introduced in a manner similar to GSM polar transmitters through the incorporation of RF, variable-gain amplifiers in each channel after the VPRO to provide envelope information to the signal.

7.5.3 Receive Mode

The block diagram of the VPRO-PLL architecture in receive mode is depicted in Fig. 7.25(a). The signals received by the antennas are amplified by LNAs and then are injected into each element of the VPRO in the current domain. The VPRO and PLL phase-shift and power-combine the received signals to accomplish phased-array spatial selectivity and down-convert the combined signal at the control voltage, thus fulfilling all the requirements of a phased-array receiver without employing explicit phase shifters, power combiners or mixers.

Consider an N -element generalization of the VPRO depicted in Fig. 7.25(a), with currents injected in shunt into every LC tank (Fig. 7.25(b)). The injected signals are assumed to be sinusoids with frequency ω_{inj} (close to the free-running frequency of the VPRO ω_{osc}) and a constant phase progression $\Delta\theta$ so that $I_{inj,i} = I_{inj} \cos(\omega_{inj}t +$

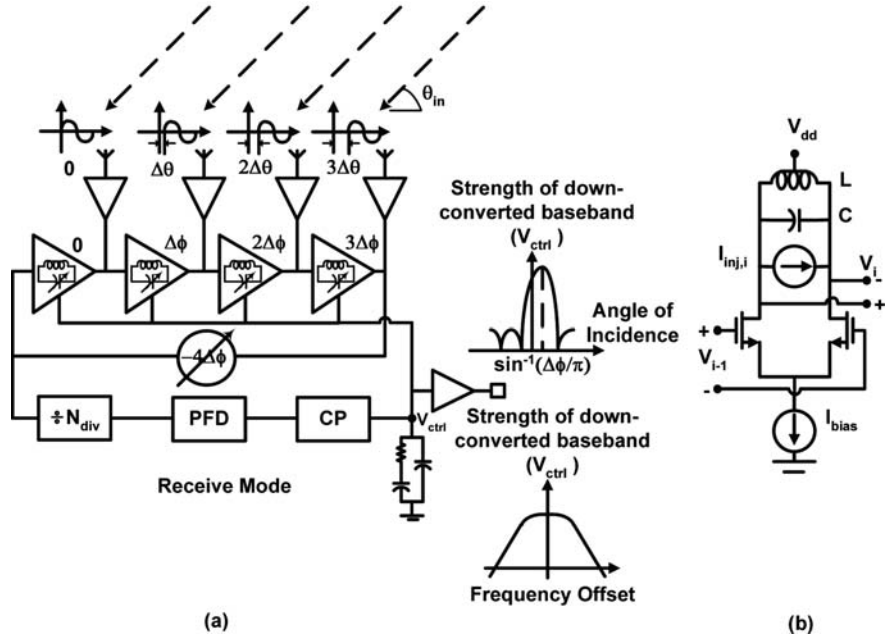


Fig. 7.25 (a) Principle of operation of the VPRO-PLL architecture in receive (RX) mode (b) Currents being injected into the LC tank of each element of the VPRO.

$(i-1)\Delta\theta$). As was noted in Section 7.2, $\Delta\theta$ is related to the angle of incidence θ_{in} (shown in Fig. 7.25(a)) as $\Delta\theta = \pi \sin \theta_{in}$. Like all electrical oscillators, as was discussed in the injection-locking section earlier, the VPRO may lock to the frequency of the injected signals if that frequency lies within a *locking range*. In the case of the VPRO, the locking range may be determined to be

$$\Delta\omega_{lock} = \left(\frac{\omega_{osc}\epsilon(1 + \tan^2 \Delta\phi)}{2Q_{inj} + \tan \Delta\phi} \right) \left(\frac{\sin \frac{N}{2}(\Delta\theta - \Delta\phi)}{N \sin \frac{\Delta\theta - \Delta\phi}{2}} \right). \quad (7.15)$$

$Q_{inj} = \frac{R}{\omega_{inj}L}$ is the Quality Factor of the tuned loads at the injection frequency and $\epsilon = \frac{I_{inj}}{I}$ is the injection gain and is equal to the ratio of the injection amplitude (I_{inj}) to the amplitude of the fundamental current of each VPRO element (I). The second term in the expression for locking range ($\frac{\sin \frac{N}{2}(\Delta\theta - \Delta\phi)}{N \sin \frac{\Delta\theta - \Delta\phi}{2}}$) is identical to the array factor of a phased array of N elements, where $\Delta\phi$ is the progressive phase-shift introduced in the transmitter/receiver and $\Delta\theta$ is the progressive phase-shift suffered by the transmitted/received beam due to the path difference in air in the direction of interest. Hence, the VPRO exhibits intrinsic beam-forming tendencies which are

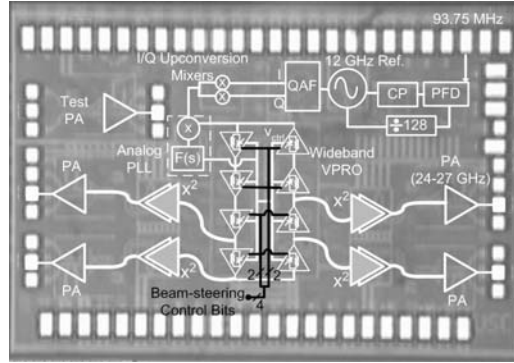


Fig. 7.26 Chip microphotograph of 24-27GHz, 4-channel, 0.13 μ m CMOS, UWB VPRO-PLL phased array transmitter [29] (©IEEE 2007).

manifested in its locking range. The constant multiplicative factor reduces to Adler's locking range result when all elements are in phase ($\Delta\phi = 0$).

When a VPRO stabilized by a PLL is subjected to external injection, the response is even more interesting. An oscillation is seen in the control voltage at the frequency difference between the injection frequency and the lock frequency of the PLL. In other words, the VPRO-PLL receiver downconverts the injected signals at the control voltage. This oscillation is a direct result of the fact that the injected signals attempt to pull the VPRO away from the PLL lock frequency and the PLL counteracts and attempts to restore it. The amplitude of the V_{ctrl} oscillation can be derived to be

$$\frac{|V_{ctrl}|}{I_{inj}} = \frac{\omega_{PLL}(1 + \tan^2 \Delta\phi)}{I(2Q_{inj} + \tan \Delta\phi)} \times \frac{\sin \frac{N}{2}(\Delta\theta - \Delta\phi)}{N \sin \frac{\Delta\theta - \Delta\phi}{2}} \times \left| \frac{K_{pd}F(j\Delta\omega_{inj})}{j\Delta\omega_{inj}N_{div} + K_{pd}K_{vco}F(j\Delta\omega_{inj})} \right|, \quad (7.16)$$

where ω_{PLL} is the lock frequency of the PLL and hence the operating frequency of the system. As mentioned before, I_{inj} is the amplitude of the current injected into each VPRO element and I is the amplitude of the fundamental current of each VPRO element. $\Delta\omega_{inj} = \omega_{inj} - \omega_{PLL}$ is the frequency difference between the injection frequency and the lock frequency of the PLL. N_{div} , K_{vco} , K_{pd} and $F(s)$ are the division ratio, VPRO tuning gain, phase-frequency detector gain and loop filter response respectively. From (7.16), the amplitude of the oscillation at V_{ctrl} shows phased-array spatial selectivity. Hence, full phased-array receiver functionality (downconversion with spatial power combining) is achieved by using V_{ctrl} as the output node. The received signal at V_{ctrl} is also directly proportional to the injected signal strength I_{inj} . This implies that the VPRO-PLL receiver has linear, small-signal gain, which is essential for amplitude-modulation schemes such as QAM.

From (7.16), the down-converted and spatially-combined signal at the control voltage has a frequency response that is governed by the PLL design parameters, such as N_{div} , K_{pd} , K_{vco} and $F(s)$. In general, the frequency response is low-pass for traditional loop-filters.

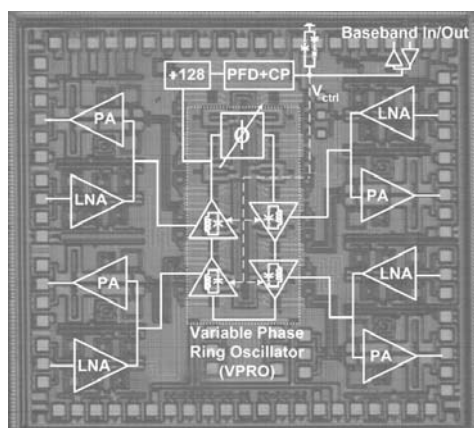


Fig. 7.27 Chip microphotograph of 24GHz, 4-channel, 0.13 μ m CMOS, VPRO-PLL phased array transceiver [28] (©IEEE 2007).

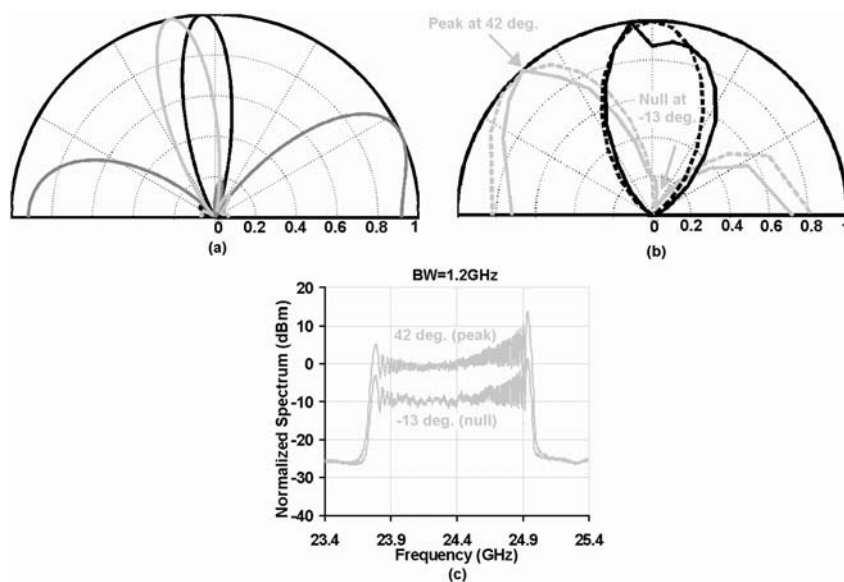


Fig. 7.28 (a) Measured 4-element, narrowband array patterns for the 4-channel, 24-27GHz, UWB phased array transmitter for selected array settings in polar coordinates (b) Measured 2-element, narrowband array patterns. The dashed lines represent the theoretical patterns (©IEEE 2007) (c) Combined UWB spectrum along the peak and null angles of transmission of a selected array setting when two elements are active (©IEEE 2007). The array setting is the same as the one used for the grey curve in the narrowband 2-element patterns.

Fig. 7.27 shows the chip microphotograph of the 24GHz, 4-channel, $0.13\mu\text{m}$ CMOS, VPRO-PLL phased array transceiver reported by the authors of this chapter in [28]. The VPRO is locked to a low-frequency reference in a divide-by-128 PLL. Due to the high division ratio, the frequency selectivity of the transmit and receive transfer functions is high, which makes the transceiver appropriate for narrowband operation. In [29], the authors report a 4-channel, 24-27GHz, UWB phased-array transmitter in $0.13\mu\text{m}$ CMOS based on the VPRO-PLL architecture (Fig. 7.26)⁷. In this UWB implementation, the VPRO is locked to an on-chip reference in a high-speed, dividerless, analog PLL which allows the architecture to support wideband phase and frequency modulation. The measured, *narrowband* array patterns of the UWB phased array transmitter are shown in Figs. 7.28(a) and 7.28(b) when four and two channels are active respectively. As is expected from (7.5), the beamwidth is reduced when four elements are active. It should be noted that these patterns are measured through on-wafer probing to eliminate the effects of packaging mismatches. Fig. 7.28(c) shows the combined spectrum along two different angles of transmission for two active channels when the transmitter emits a UWB signal with 1.2GHz of bandwidth. The two angles considered correspond to the expected peak and null locations for the selected array setting, which is the same setting as the one used for the light grey curve in Fig. 7.28(b). The combined spectrum is seen to be suppressed by at least 10 dB in the null direction when compared to the peak. This verifies the UWB spatial selectivity of the array. For additional details on the VPRO-PLL architecture, including design guidelines and more measured results from the prototypes, the reader is directed to [30] and [31].

7.6 The Effect of Mismatch in Phased Arrays

The performance in the presence of channel mismatches is a critical parameter of any phased array. These variations result in amplitude and phase mismatches in the radiated/received signals and hence adversely affect the beam pattern [32],[33]. As an example, Fig. 7.29 shows the simulated array factor of a 4-element array in the presence of the deterministic amplitude and phase errors mentioned in the figure. The nominal beam-pointing direction is normal to the array, and the errors cause a deviation from this direction. The array also shows reduced peak gain and one of the sidelobe levels is higher. For reference, 15° of phase variation corresponds to approximately $100\mu\text{m}$ of length at 60GHz, assuming an effective dielectric constant of 4.

In this section, the effect of phase mismatches between the channels of a conventional RF phase shifting array is analyzed. Simple equations are presented in this chapter to quantify the effect of mismatches on various facets of array performance, including beam-pointing angle and sidelobe rejection ratio (SLRR). These mismatches arise from two possible sources. The first source is the intra-chip variation

⁷ For 3GHz bandwidth and four array channels around a center frequency of 25.5GHz, a timed array implementation is not required.

that is inherent to any process technology. The second source, which is particularly pronounced at millimeter-wave frequencies, arises from packaging mismatches in the interface between the integrated phased array and the off-chip antennas.

7.6.1 Beam-pointing Error

For phased arrays that employ the delay-phase approximation and utilize half-wavelength-spaced antennas, the array factor of (7.3) can be rewritten as

$$AF(\Delta\phi, \theta_{in}) = \left(\frac{\sin \frac{N(\Delta\phi - \pi \sin \theta_{in})}{2}}{\sin \frac{\Delta\phi - \pi \sin \theta_{in}}{2}} \right)^2, \quad (7.17)$$

where $\Delta\phi$ is the phase progression introduced in the array. The beam-pointing angle can then be written as $\theta_m = \sin^{-1} \frac{\Delta\phi}{\pi}$.

In [32] and [33], the effect of channel mismatches on the beam-pointing angle of a conventional RF phase-shifting array is analyzed. The authors demonstrate that to first order, amplitude mismatches between phased array channels do not affect the beam-pointing angle, and only phase deviations need to be considered. The effect of phase errors on the beam-pointing angle can be summarized as

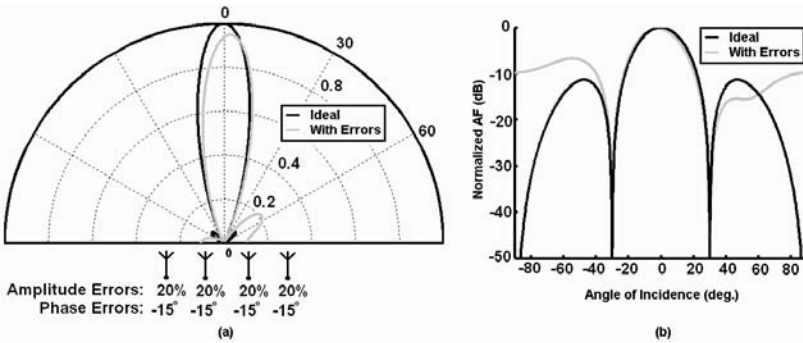


Fig. 7.29 (a) Sample 4-element array factor (normalized to the ideal peak array gain of 16) in the presence of amplitude and phase errors in polar coordinates (b) The same array factor plotted in the dB scale in Cartesian coordinates.

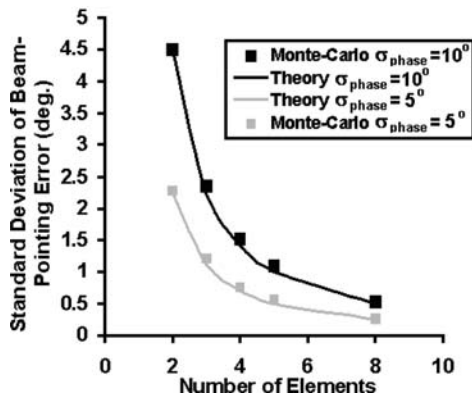


Fig. 7.30 Standard deviation of the beam-pointing error as obtained from (7.19) and 300-iteration Monte-Carlo simulations of a conventional RF-phase-shifting array with $\sigma_{\text{phase}}=5^\circ$ and 10° . The nominal phase shift of each channel is 0° , resulting in a nominal pointing angle that is normal to the array.

$$\Delta\theta_m \approx \frac{-\sum_{m=1}^N \sum_{n=1}^N (\delta_m - \delta_n)(m-n)}{\pi \cos \theta_m \frac{(N-1)N^2(N+1)}{6}}, \quad (7.18)$$

where $\Delta\theta_m$ is the beam-pointing error, θ_m is the beam-pointing angle in the absence of errors and δ_m is the phase error in the m^{th} phased array channel. This formula is derived for antennas that are half-wavelength apart, and considers only the linear terms of a Taylor Series expansion. For additional details on the formulation and derivation, the reader is directed to [33].

If each channel sustains phase errors that are small, independent and identically distributed with a distribution of $N(0, \sigma_{\text{phase}}^2)$, the variance of the beam-pointing error can be computed from (7.18) to be

$$\sigma_{\text{beam}}^2 = \frac{12\sigma_{\text{phase}}^2}{\pi^2 \cos^2 \theta_m (N-1)N(N+1)}. \quad (7.19)$$

Fig. 7.30 depicts σ_{beam} as obtained from (7.19) and 300-iteration Monte-Carlo simulations of a conventional RF phase-shifting array with $\sigma_{\text{phase}}=5^\circ$ and 10° . The nominal phase shift of each channel is 0° , resulting in a nominal pointing angle that

is normal to the array. It is interesting to note that from (7.19) and Fig. 7.30, the variance of the beam-pointing error is seen to fall at the rate of $\frac{1}{N^3}$ as N is increased.

7.6.2 Sidelobe Rejection Ratio

Sidelobes are a feature of all phased arrays. In the absence of errors, the locations of the main lobe, sidelobes and nulls in the array factor can be determined by differentiating (7.17) with respect to θ_{in} and setting the derivative equal to zero. The locations of the main lobe and the sidelobes are the solutions to

$$\tan \frac{N(\Delta\phi - \pi \sin \theta_{in})}{2} = N \tan \frac{\Delta\phi - \pi \sin \theta_{in}}{2}. \quad (7.20)$$

The trivial solution of $\Delta\phi = \pi \sin \theta_{in}$ corresponds to the main lobe. Solving this transcendental equation for the first or any other sidelobe is difficult. However, it is found that

$$\pi \sin \theta_{lobe} \approx \Delta\phi \pm \phi_{N,lobe}, \text{ where } \phi_{N,lobe} = \frac{2.929\pi}{N} \quad (7.21)$$

proves to be an excellent approximation for the location of the first sidelobes on either side of the main lobe. The sidelobe rejection ratio (*SLRR*), defined as the ratio of the power of the main lobe to the power of these first sidelobes, can now be computed as ⁸

$$SLRR_0 = \frac{N^2}{AF(\Delta\phi, \theta_{lobe})} = \frac{N^2 \sin^2(\frac{\phi_{N,lobe}}{2})}{\sin^2(\frac{N\phi_{N,lobe}}{2})}. \quad (7.22)$$

In order to determine the *SLRR* in the presence of channel mismatches, the array factor in the presence of mismatches is written as

$$AF_{err}(\Delta\phi, \theta_{in}, A_1..A_N, \delta_1..\delta_N) = \left| \sum_{i=1}^N \left(1 + \frac{\Delta A_i}{A}\right) e^{j((i-1)(\Delta\phi - \pi \sin \theta_{in}) + \delta_i)} \right|^2, \quad (7.23)$$

where $\frac{\Delta A_i}{A} = \frac{A_i - A}{A}$ is the normalized amplitude error of each element, δ_i is the phase error of each element and $\Delta\phi$ is the array's phase progression. The deterioration in the *SLRR* due to small amplitude and phase errors can be determined using a Taylor Series as

$$AF_{err}(\Delta\phi, \theta_{lobe}, A_1..A_N, \delta_1..\delta_N) = AF(\Delta\phi, \theta_{lobe}) + \sum_{k=1}^N \frac{dAF_{err}}{d\delta_k} \Big|_{\text{all } \delta_i, \Delta A_i = 0} \delta_k + \sum_{k=1}^N \frac{dAF_{err}}{d\Delta A_k} \Big|_{\text{all } \delta_i, \Delta A_i = 0} \Delta A_k. \quad (7.24)$$

⁸ The subscript of 0 signifies that mismatches and process variations are absent.

The derivatives in the equation shown above may be computed from (7.23), and are found to be

$$\frac{dAF_{err}}{d\delta k} \Big|_{\text{all } \delta_i, \Delta A_i = 0} = \frac{2 \sin \frac{N\phi_{N,lobe}}{2} \sin \left(\left(k - \frac{N+1}{2} \right) \phi_{N,lobe} \right)}{\sin \frac{\phi_{N,lobe}}{2}}, \quad (7.25)$$

$$\frac{dAF_{err}}{d\Delta A_k} \Big|_{\text{all } \delta_i, \Delta A_i = 0} = \frac{2 \sin \frac{N\phi_{N,lobe}}{2} \cos \left(\left(k - \frac{N+1}{2} \right) \phi_{N,lobe} \right)}{A \sin \frac{\phi_{N,lobe}}{2}}. \quad (7.26)$$

If amplitude errors are absent and the channel phase errors are small, independent and identically distributed with a distribution of $N(0, \sigma_{phase}^2)$, the variance of $AF_{err}(\Delta\phi, \theta_{lobe})$ can be computed from (7.24) to be

$$\sigma_{AF_{err}(\Delta\phi, \theta_{lobe})}^2 = 2\sigma_{phase}^2 \frac{\sin^2 \frac{N\phi_{N,lobe}}{2}}{\sin^2 \frac{\phi_{N,lobe}}{2}} \left(N - \frac{\sin N\phi_{N,lobe}}{\sin \phi_{N,lobe}} \right). \quad (7.27)$$

The standard deviation of the $SLRR$ ($\sigma_{SLRR} = \frac{N^2}{AF_{err}(\Delta\phi, \theta_{lobe})}$) can then be computed to be

$$\sigma_{SLRR}^2 = 2N^4 \sigma_{phase}^2 \frac{\sin^6 \frac{\phi_{N,lobe}}{2}}{\sin^6 \frac{N\phi_{N,lobe}}{2}} \left(N - \frac{\sin N\phi_{N,lobe}}{\sin \phi_{N,lobe}} \right). \quad (7.28)$$

Fig. 7.31(a) depicts σ_{SLRR} as obtained from (7.28) and 300-iteration Monte-Carlo simulations of a conventional RF phase-shifting array with $\sigma_{phase}=2.5^\circ$ and 5° . The nominal phase shift of each channel is 0° . For large array sizes, from (7.28), $\sigma_{SLRR} \approx \frac{138.6\sigma_{phase}}{\sqrt{N}}$ and hence reduces with an increase in array size. Fig. 7.31(b) presents another visualization, depicting the theoretical 1- σ confidence interval for the $SLRR$ in dB-scale.

7.6.3 Implications on Array Packaging

As was discussed earlier in this chapter, there are two fundamental reasons that motivate the use of phased arrays in commercial wireless applications and radar. The first includes the increase in the transmitted power and the improvement in SNR at the receiver, which alleviate link budget requirements. The second is the improved directionality of the wireless link, which minimizes spatial interference and multipath effects. For applications that wish to harness the former, the accuracy of the beam-pointing angle is of importance as an error in the direction of transmission/reception would deteriorate the link budget. On the other hand, in environments where spatial interference and multipath effects dominate, the sidelobe rejection ratio is a critical parameter. The analyses and simulations in this section indicate that channel mismatches have a greater impact on sidelobe rejection than beam-pointing angle for

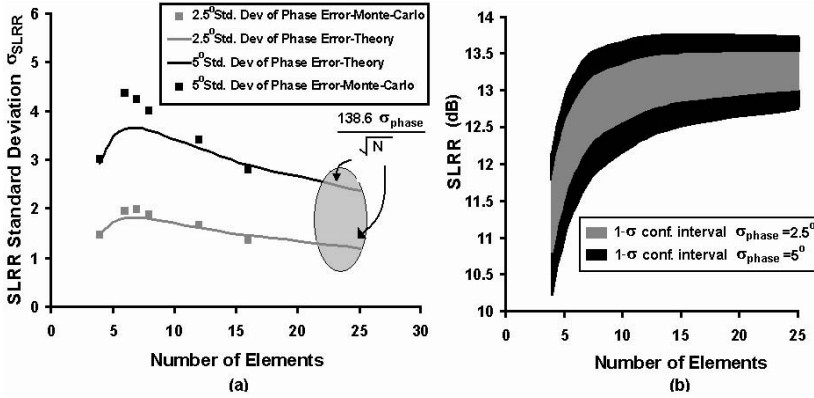


Fig. 7.31 (a) Standard deviation of the $SLRR$ in a conventional RF phase-shifting array as obtained from (7.28) and Monte-Carlo simulations. The nominal beam-pointing angle is normal to the array. (b) Theoretical 1- σ confidence interval for the $SLRR$.

typical array sizes and mismatch values. This implies that packaging is a greater concern for applications that are dominated by interference and multipath effects. Accurate, robust and repeatable packaging techniques, especially for multiple antenna systems, have yet to be developed at millimeter-wave frequencies. Techniques such as flip-chip antenna bonding [34] and on-chip antennas [35] show promise in combating this problem.

7.6.4 Array Calibration

The mismatch between multiple channels, induced by process, packaging, or connection to the antennas, as well as channel-to-channel signal coupling deteriorates the antenna array performance and is often calibrated in high performance military-type phased arrays. Process mismatches aside, a phase error of $\pm 15^\circ$ at millimeter-wave frequencies corresponds to a few tens of microns length variation of a wirebond, printed circuit board traces, chip-antenna connections, or antennas themselves if implemented off-chip. Variations of this order, even if calibrated at the time of manufacturing, might occur during the lifetime of the system. In order to realize robust silicon-based single-chip antenna arrays for commercial communication and sensing applications, on-chip testing and calibration techniques may be developed that measure the deterioration of array performance and appropriately correct the amplitudes and phases. In a communication system, measurement of array performance can be done at the system level where the gains and phases are adjusted to maximize the SNR. In a sensing array such as the automotive radar application, calibration through

sending a training sequence between the transmitter and receiver, e.g., smart antenna concepts, is not possible. Calibration techniques similar to [42],[43] that rely on the mutual coupling between the array antenna elements may be utilized, although the effect of on-chip coupling though the substrate must be considered as well.

7.7 Quantization Error in Phased Arrays

Practical phase shifters or delay elements have a finite number of settings and hence yield a discrete set of phases/delays. As a result, the steering angles achievable in practical phased arrays are also quantized. This section analyzes the implications of this quantization.

Fig. 7.32 depicts an N -element, RF phase-shifting, homodyne, phased array receiver with quadrature downconversion and half-wavelength-spaced antennas. The variable phase-shifters for each signal path are assumed to be controlled by n digital control bits, giving rise to 2^n phase-shift settings. For example a 3-bit phase-shifter can sustain shifts of -180° , -135° , $90^\circ \dots 135^\circ$, corresponding to a phase-shifting resolution of $\Delta\phi_{res} = \frac{360^\circ}{2^n} = 45^\circ$ and a beam-steering resolution of $\Delta\theta_{res} = \sin^{-1} \frac{\Delta\phi_{res}}{\pi} = \sin^{-1} \frac{1}{2^{n-1}} = 14.48^\circ$

An incoming QAM signal of the form $i(t)\sin(\omega t) + q(t)\cos(\omega t)$ is assumed, where $i(t)$ and $q(t)$ are the in-phase (I) and quadrature (Q) information signals. The I and Q LO signals are assumed to be $\sin(\omega t)$ and $\cos(\omega t)$ respectively, and the RF phase-shifter Φ_k is set to $(k-1)\Delta\phi$ (with $\Delta\phi$ set to the discrete phase difference closest to the incoming phase difference of $\pi \sin \theta_{in}$). The final combined signals in the I and Q channels can be determined to be

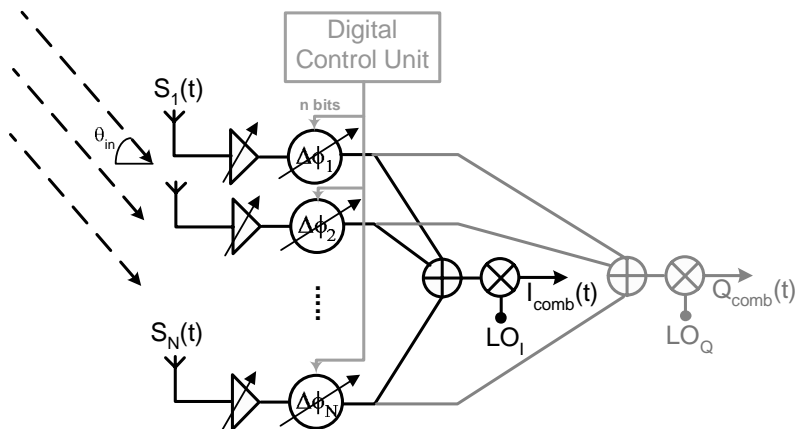


Fig. 7.32 N -element RF phase-shifting homodyne phased array receiver with quadrature downconversion.

$$I_{comb}(t) = \frac{i(t)}{2} \sqrt{AF(\Delta\phi, \theta_{in})}, \quad (7.29)$$

$$Q_{comb}(t) = \frac{q(t)}{2} \sqrt{AF(\Delta\phi, \theta_{in})}. \quad (7.30)$$

where AF is the array factor defined in (7.17). A number of assumptions are implicit in the above equations. The second harmonics produced by the mixers are assumed to be filtered out. Secondly, the bandwidth of the information signal is assumed to be small compared to the carrier frequency to eliminate the Array-induced ISI effect. Finally, if the discrete phase-shifter settings are not able to perfectly compensate for the delay in free space ($\Delta\phi \neq \pi \sin \theta_{in}$), the recovered constellation is rotated due to leakage of I into the Q channel and Q into the I channel. This systematic rotation is assumed to be undone in the receiver.

When the discrete phase-shifter settings are not able to perfectly compensate for the delay in free space, $AF(\Delta\phi, \theta_{in}) < N^2$, the peak array power gain, causing the receiver to show reduced gain and hence, reduced SNR . This is a direct result of the fact that the direction of arrival of the incoming signal and the beam-pointing angle are mismatched. The resultant SNR degradation can be quantified as

$$SNR = SNR_o \times \frac{AF(\Delta\phi, \theta_{in})}{N^2}, \quad (7.31)$$

where SNR_o is the SNR in the absence of the effect of discrete phases, and includes the input SNR , receiver front-end noise figure (NF) and the $10\log(N)$ phased array improvement⁹.

The SNR reduction can be related to a degradation in the Error Vector Magnitude (EVM) as

$$EVM = \frac{1}{\sqrt{SNR}} = \frac{1}{\sqrt{SNR_o}} \times \frac{N}{\sqrt{AF(\Delta\phi, \theta_{in})}}. \quad (7.32)$$

Fig. 7.33(a) shows the theoretical EVM degradation factor, i.e., the $\frac{N}{\sqrt{AF}}$ -term in (7.32), as a function of the angle of incidence for 4- and 8-element arrays, with 3- and 4-bit phase-shifting in each case. A Simulink simulation of an 8-element quadrature RF phase-shifting homodyne receiver with 3-bit phase-shifters receiving a 16-QAM input is also included. SNR_o is set to be 33.2dB (corresponding to, for instance, an input SNR of 30.2dB and a front-end NF of 6dB, along with the $10\log(8)$ array SNR improvement), resulting in $EVM_o = 1.5\%$.

As can be seen in Fig. 7.33(a), the maximum degradation in EVM occurs when the progressive phase-shift of the incoming wave in space lies right between the phase-shifting resolution of the receiver. This maximum degradation can be computed to be

$$\left(\frac{EVM}{EVM_o} \right)_{max} = \frac{N \sin \frac{\Delta\phi_{res}}{4}}{\sin \frac{N \Delta\phi_{res}}{4}}. \quad (7.33)$$

⁹ The $10\log(N)$ SNR improvement assumes that each antenna picks up uncorrelated noise from the surroundings. It is also true when the receiver front-end NF dominates over the input noise, as the front-end noise would be uncorrelated between different signal paths.

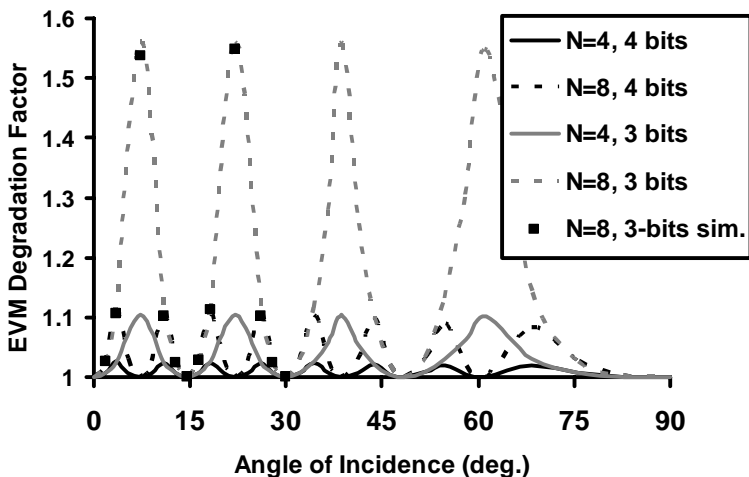


Fig. 7.33 *EVM* degradation due to discrete phases as a function of the incidence angle, phase-shifter bits and number of elements N . A Simulink simulation of a 16-QAM input incident on a 8-element quadrature RF phase-shifting homodyne receiver with $SNR_o = 33.2\text{dB}$ and 3-bit phase-shifters is also shown.

As N , the number of elements, is increased, for a fixed phase-shifting resolution, the maximum *EVM* degradation increases as the beam becomes narrower, resulting in a higher gain and *SNR* penalty when the incoming direction and the beam-pointing angle are mismatched. This effect is reflected in Fig. 7.33(a). Therefore, for a practical phased array, the choice of the number of phase-shifter bits is governed by N and the desired upper bound on *EVM* degradation, with a larger number of bits required for larger arrays.

7.8 Multi-Beam Antenna Arrays

Military radars have been the major application of antenna arrays. In these applications, it is highly desirable to locate and track multiple targets simultaneously. Mechanical steering of multiple directional antennas, each for one target, is not possible in many applications such as airborne radars. Antenna arrays with enough degrees of freedom in the amplitude and delay (phase) of each RF path can form and scan multiple simultaneous beams. A linear combination of narrowband RF signals with different phase shifts and amplitudes can form beams at the desired directions. Given a narrowband signal, the most general case is to incorporate independent amplitude and phase control in each RF path¹⁰. The phase shifted and amplified

¹⁰ It should be reminded again that in a narrowband system, phase shifting the RF signal can be done in the local oscillator path as well.

(or attenuated) signals are not all combined, since multiple outputs corresponding to different beam angles are needed. Rather, these signals are all down converted to baseband independently. The baseband signals are then converted to a digital data stream and processed appropriately to provide the information coming from multiple spatial angles. This is often referred to as digital beam-forming. Access to all RF signals independently and processing them in the digital domain not only creates multiple simultaneous beams, but also is used in advanced Adaptive Space Time Array Processors to increase the signal-to-clutter ratio. When the array size is large, converting all the phase shifted and amplified RF signals to baseband independently requires much area (I/O signals) and power consumption not only in the analog to digital data conversion, but also in the digital signal processor (digital beam former). Independent access to all elements' amplitudes and phases is usually not required to achieve the desired number of independent beams. Therefore, it is common to divide a larger array to smaller sub-arrays, say with a size of 4×4 in a 2D array. Within a sub-array, each RF path has an independent phase and amplitude control; but, all signals are finally combined. In other words, only the combined signals within the sub-arrays are down-converted to baseband for further processing. The appropriate size of the sub-array depends on the system requirements such as the total number of desired beams, array size, and also the signal bandwidth.

Multiple beams can be formed in the RF domain as well. As discussed before, a linear delay (or phase) progression between different RF paths creates a beam at a particular angle. In a standard phased array, this delay (phase) progression is adjusted electronically. If multiple delays or phases are created simultaneously, multiple beams can be formed in the RF domain simultaneously. This is simply illustrated in Figure 7.34. An example of a modular multi-beam array based on this straight forward implementation is given in [36]. In order to form k independent beams in an n -element array, $k \times n$ phase shifters and variable gain amplifiers are needed.

The main advantage of multiple beam formation in the RF domain is that interference signals are cancelled at the front end prior to down-conversion and conversion to digital bits due to the spatial selectivity of the RF beam-former. This reduces the dynamic range and hence the power consumption of RF mixers and data converters. In principle, each output of a multi-beam RF beam-former can be connected to a separate transceiver allowing for processing of signals from all spatial segments simultaneously (spatial filter bank). However, in many applications, this capability is not required, especially, as it comes at the expense of chip area and power consumption of several transceivers. Oftentimes, only a few simultaneous beams are sufficient at any time to improve the system robustness and functionality. If not all the created angles in the RF beam-former are needed at a given time, fewer number of transceivers (up/down-converters, ADCs/DACs) can be turned on. Alternatively, an RF beam selector can choose the appropriate outputs of the multi-beam RF array processor and feed them to a few transceivers in this Switched-Beam Multi-Beam array. The disadvantage of an RF multi-beam approach, if implemented in the straight forward manner as shown in Figure 7.34, is the larger area of multiple RF variable true time delay elements (phase shifters) and power combiners. Figure 7.35 com-

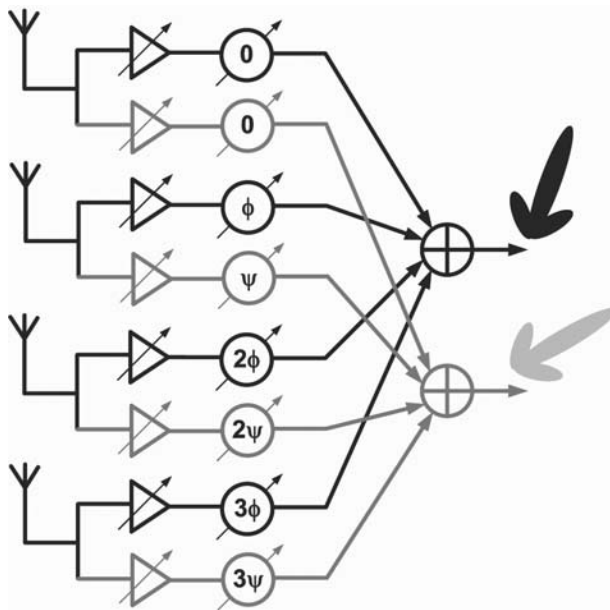


Fig. 7.34 A phased array that generates multiple beams through the usage of additional delay elements/phase shifters.

compares the RF Beam-Forming versus the Digital Beam-Forming architectures in a multi-beam array.

The fact that for a given incident angle, the time delay (phase shift) difference between adjacent elements is constant can be used to create multi-beam architectures with fewer number of area-hungry variable time delay elements. The most famous two have been proposed by Butler [37] (Fig. 7.37) and Blass [38] (Fig. 7.37) for narrowband and wideband arrays, respectively. Both of these solutions were developed assuming discrete component implementations and are not necessarily optimum for integration in a small chip size.

Recently, Chu has proposed an architecture where multiple beams are formed in the RF domain in a small area [39]. Chu's implementation relies on true time delay elements and hence can be used for both wideband and narrowband arrays (phased arrays and timed arrays). The basic principle behind the architecture is shown in Figure 7.38 where two antenna elements are connected to the two ends of a transmission line¹¹. At any point along the transmission line, signals that are received by either of the antennas will face a delay proportional to the length of line from their corresponding antennas to the measured point. The delay difference between these signals dictates the incident angle. For instance, both received signals

¹¹ The effect of loading and reflections are ignored in the discussions for simplicity. These are taken care of in the actual implementation [39].

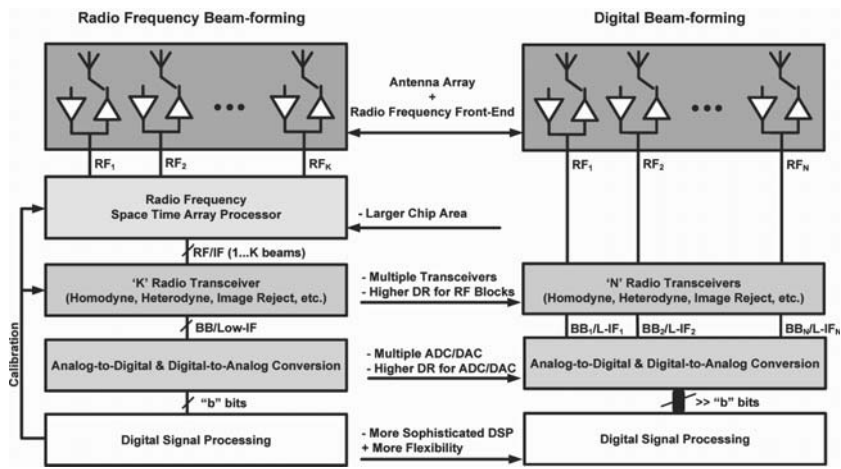


Fig. 7.35 A comparison of RF beamforming and Digital beamforming architectures for multi-beam arrays.

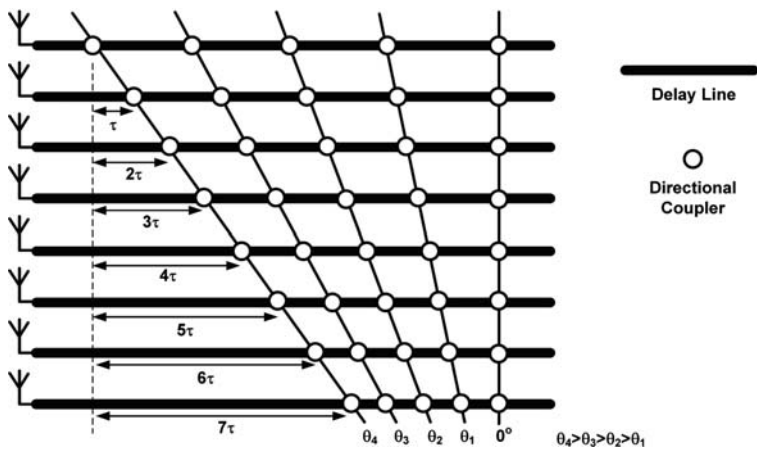


Fig. 7.36 The Blass matrix for wideband multi-beam arrays.

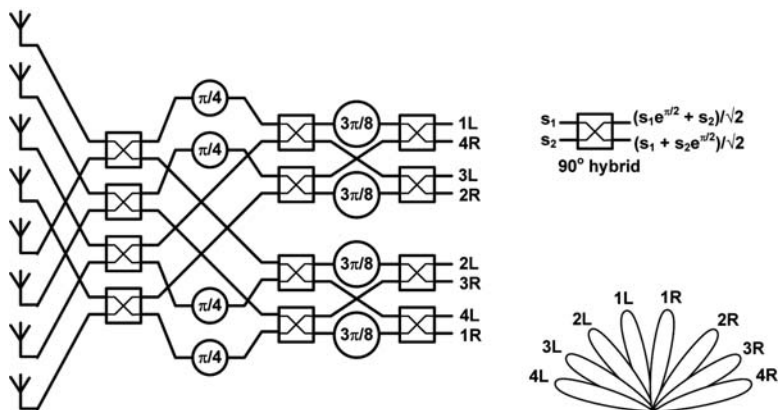


Fig. 7.37 The Butler matrix for narrowband multi-beam arrays.

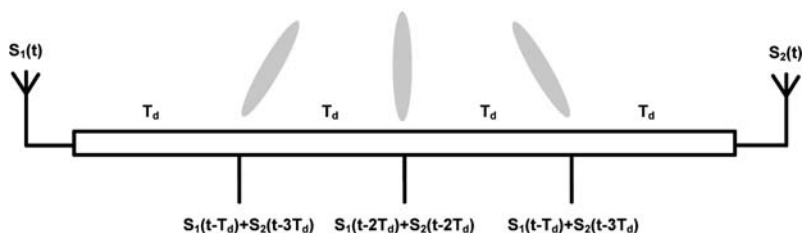


Fig. 7.38 Chu's architecture for RF multi-beam arrays.

reach the middle point with equal delay. Hence, the middle point corresponds to the beam that is normal to the plane of antennas. In this simplified example, it is clear that using a single delay line with several tap points, each corresponding to a different incident angle, leads to a much more compact realization compared with both Blass and Butler architectures. In [39], the authors report a 2D version of the array where 7×7 simultaneous beams are formed in an array with 2×2 elements. The architecture can be extended to $2N \times 2N$ antenna elements and $K \times K$ simultaneous beams.

It is also worth commenting on the possible advantages of a multi-beam array for wireless communication. Transmitting multiple data streams for different angles at the base station of a cellular system increases the communication capacity. At millimeter waves, creating multiple beams at the receiver has certain advantages as well. One of the major impediments in realizing millimeter wave wireless communica-

tions has been the poor link robustness due to the ease in obstructing the line of sight. Phased arrays can somewhat help by steering the beam away in an attempt to find an alternative communication path through reflections. Alternatively, a multi-beam array at the transmitter and receiver can be used to create several parallel communication channels, some perhaps through reflections, to improve the link robustness. The decision to design a multi-beam array versus a single-beam scanning array in millimeter waves mostly depends on the application and involves the usual tradeoff between reliability, performance, cost, and power consumption.

7.9 Antenna Arrays and Multiple Input Multiple Output (MIMO) Transceivers

Spatial diversity utilizing multiple spaced antennas is an attractive way to increase the signal-to-noise ratio and hence enhance the Shannon channel capacity. In a statistically fading environment, the received signal varies rapidly as the distance between the transmitter and receiver is altered by a fraction to a few wavelengths¹² [44]. Spatially separated antennas can be used to extract more information and effectively increase the higher data rate. If multiple receiving antennas are sufficiently spaced apart, their received signal levels are almost independent in a statistically fading channel. A decision circuit can simply pick the antenna with the highest received SNR (selection diversity) or more optimally it can combine the complex weighted received signals¹³ from all the antennas to achieve the maximum possible SNR (maximal ratio combining). The receiver architecture for such diversity systems will be similar to those already discussed for beam-forming arrays; although, the algorithm to set the complex weights (amplitudes and phases) will be different. In the previously discussed beam forming schemes, the goal has been to create a line-of-sight between the transmitter and receiver in order to maximize the transmit power efficiency, increase the receive SNR, and spatially filter the interference signals. In diversity systems, the goal is to maximize the received SNR in a highly scattering environment with little or no line-of-sight between the transmitter and receiver.

Similarly, in a statistically fading environment, transmitted signals from spatially separated antennas experience independent channels as they reach the receiver. If channel properties from each transmitting antenna to the receiver are known, the transmitter can shape the signal that is fed into each antenna in such a way that the received signals add coherently. Even in the case where the communication channel is not known to the transmitter, by transmitting different signals from various antennas over time, a larger SNR at the receiver can be achieved in the so-called Space-Time Coding schemes [45]. As a more general case, we can imagine multiple-input multiple-output, MIMO, systems with 'M' transmitting antenna elements and 'N'

¹² The minimum required distance between the antennas that results in independent (uncorrelated) communication channels is a function of size, shape, and number of scattering objects and their distance to the antennas. This typically varies from a few tenths to a few wavelengths.

¹³ Complex weight assumes narrowband signals.

receiving antenna elements. Assuming independent communication channels from each transmitting antenna to a receiving antenna, it can be shown that besides the achieved diversity gain, the capacity of a MIMO system increases with the number of antenna elements. Although some space time codes require only delay-and-sum processing or a linear combination of complex weighted signals at the receiver, but many advanced codes require other forms of processing on the received signals. In this case, each antenna must be accompanied by an independent transceiver and all the processing is done at baseband. Whether space-time coding is attractive for the wireless communication applications envisioned at millimeter waves or not is still unclear. In many scenarios, the distance between the transmitter and receiver is short and LOS can easily be created though beam-forming. Moreover, at millimeter waves, the overall SNR will be dominated by the receiver's NF and therefore a mere combining of signals from multiple parallel receivers will enhance the SNR in most cases. Given the fact that the most dominant rationale behind millimeter wave communication is the large user channel bandwidth ($> 1\text{GHz}$), digital processing of multiple received signals does not appear to be power efficient. Therefore, the RF beam-forming architectures described in this chapter will likely be more suitable.

7.10 Concluding Remarks

Over the past few years, silicon based integrated millimeter wave systems have generated great research interest due to their advantages in offering high data rate in wireless communications and high resolution in radar and imaging systems at a lower cost. Various forms of multi-antenna systems provide a plethora of solutions for wireless communications, radar, and imaging systems. Millimeter waves offer more bandwidth for ultra high data rate wireless communications and better resolution for radar and imaging systems, while reducing the required size of integrated systems in a multi-antenna configuration. Integration of a complete multi-antenna system in silicon results in substantial improvements in cost, size, and reliability and provides numerous opportunities to perform on-chip signal processing and conditioning.

There have already been several successful university demonstrations of integrated phased array receivers, transmitters, and transceivers at 24GHz and 77GHz over a relatively short span of five years [18],[19],[26],[35],[28],[29]. In addition, an integrated beam-former at 60GHz has been reported [8]. This beam-former chip can be added to existing millimeter wave receiver or transmitter chipsets to provide spatial selectivity and improve sensitivity.

Silicon based integrated wideband timed array transceivers are more recent efforts [5],[6] where most of the focus has been on frequencies below 15GHz. Integrated broadband beam-formers at microwaves and millimeter waves have been reported. However, they provide a constant phase shift, and not a constant group delay, between adjacent elements and are not appropriate for signals with instantaneously wide bandwidth. An impulse-based ultra-wideband automotive radar array operating at

22GHz - 29GHz is one example where timed array transceivers may be needed instead of conventional phased arrays.

Packaging and antenna design, already key elements and research topics in a low cost realization of millimeter wave systems, are even more important in the context of transceiver arrays. There is an ongoing debate on the issue of off-chip versus on-chip antennas [35]. In the context of arrays, the advantages of on-chip antennas are lower chip-antenna interconnect loss, more robustness to variations, and ease of system integration. However, on-chip antennas are not efficient due to the loss of silicon substrate or the small distance between the metal layers in on-chip patch antennas when the silicon substrate is shielded. Plus, the cost of silicon area used for on-chip antennas is not negligible. Off-chip antenna arrays necessitate a low cost, robust, and preferably symmetric chip-antenna interconnect scheme. Low cost integration of the silicon chips with efficient antenna arrays is one of the areas that deserves more research. Low cost testing and calibration of integrated transceiver arrays for commercial applications are other important areas that need to be addressed.

Overall, with applications in a variety of sectors including commercial, space, military, healthcare, environmental, and pure scientific discoveries, millimeter wave and broadband silicon based multiple-antenna integrated systems constitute a rich area of research for some more years to come.

References

1. P. K. Bondyopadhyay, "The first application of array antenna", in *2000 IEEE International Conference on Phased Array Systems and Technology*, May 2000, pp. 29-32.
2. D. Parker and D. C. Zimmerman, "Phased arrays - part II: implementations, applications and future trends," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 3, pp. 688-698, March 2002.
3. T. Yao, M. Gordon, K. Yau, M. T. Yang, S. P. Voinigescu, "60-GHz PA and LNA in 90-nm RF-CMOS," in *2006 IEEE RFIC Symposium Digest of Technical Papers*, June 2006.
4. E. Grass, F. Herzel, M. Piz, Y. Sun and R. Kraemer, "Implementation aspects of Gbit/s communication system for 60 GHz band," in *Proceedings of the 14th Wireless World Research Forum (WWRF 14)*, July 2005.
5. J. Roderick, H. Krishnaswamy, K. Newton and H. Hashemi, "Silicon-Based Ultra-Wideband Beam-Forming," *IEEE Journal of Solid State Circuits*, vol. 41, no. 8, pp. 1726-1739, Aug. 2006.
6. Ta-Shun Chu, J. Roderick, H. Hashemi, "A 4-Channel UWB Beam-Former in 0.13 μ m CMOS using a Path-Sharing True-Time-Delay Architecture," in *ISSCC Digest of Technical Papers*, vol. 50, Feb. 2007, pp. 426-613.
7. C. Cao, K. K. O, "Millimeter-wave voltage-controlled oscillators in 0.13 μ m CMOS technology," *IEEE Journal of Solid State Circuits*, vol. 41, no. 6, pp. 1297-1304, June 2006.
8. A. Natarajan, B. Floyd and A. Hajimiri, "A Bidirectional RF-Combining 60GHz Phased-Array Front-End," in *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, vol. 50, Feb. 2007, pp. 202-597.
9. Byung-Wook Min and Gabriel M. Rebeiz, "Ka-Band BiCMOS 4-Bit Phase Shifter with Integrated LNA for Phased Array T/R Modules," in *2007 IEEE/MTT-S International Microwave Symposium Digest*, pp. 479-482, June 2007.
10. M. Morton, J. P. Comeau, J. D. Cressler, M. Mitchell and J. Papapolymerou, "Sources of Phase Error and Design Considerations for Silicon-Based Monolithic High-Pass/Low-Pass Microwave Phase Shifters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 12, pp. 4032-4040, Dec. 2006.
11. T. M. Hancock, G. M. Rebeiz, "A 12-GHz SiGe phase shifter with integrated LNA," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 3, part 1, pp. 977-983, March 2005.
12. C. T. Charles and D. J. Allstot, "A 2-GHz Integrated CMOS Reflective-type Phase Shifter with 675° Control Range," in *Proceedings of the 2006 IEEE International Symposium on Circuits and Systems*, May 2006, pp. 381-384.
13. H. Zarei and D. J. Allstot, "A low-loss phase shifter in 180 nm CMOS for multiple-antenna receivers," in *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, Feb. 2004, pp. 392-534.
14. Kwang-Jin Koh and Gabriel M. Rebeiz, "An X- and Ku-Band 8-Element Linear Phased Array Receiver," in *Proceedings of the 2007 IEEE Custom Integrated Circuits Conference*, Sep. 2007, pp. 761-764.
15. M. Chua and K. Martin, "1 GHz Programmable Analog Phase Shifter for Adaptive Antennas," in *Proceedings of the 1998 IEEE Custom Integrated Circuits Conference*, May 1998, pp. 11-14.
16. A. Afsahi, A. Behzad, S. Au, R. Roufoogaran and J. Rael, "An Area and Power Efficient Phase Shifter + Mixer Circuit Applied to WLAN System," in *2007 IEEE Radio Frequency Integrated Circuits Symposium Digest of Technical Papers*, June 2007, pp. 357-360.
17. S. Alalusi and R. Brodersen, "A 60GHz Phased Array in CMOS," in *Proceedings of the 2006 IEEE Custom Integrated Circuits Conference*, Sep. 2006, pp. 393-396.
18. Xiang Guan, H. Hashemi and A. Hajimiri, "A fully integrated 24-GHz eight-element phased-array receiver in silicon," *IEEE Journal of Solid State Circuits*, vol. 39, no. 12, pp. 2311-2320, Dec. 2004.
19. A. Natarajan, A. Komijani and A. Hajimiri, "A fully integrated 24-GHz phased-array transmitter in CMOS," *IEEE Journal of Solid State Circuits*, vol. 40, no. 12, pp. 2502-2514, Dec. 2005.

20. H. Krishnaswamy and H. Hashemi, "A rigorous phase noise analysis of tuned ring oscillators," in *2007 IEEE Radio and Wireless Symposium Digest of Technical Papers*, Jan. 2007, pp. 43-46.
21. H. Hashemi, Xiang Guan, A. Komijani, A. Hajimiri, "A 24-GHz SiGe phased-array receiver - LO phase-shifting approach," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 2, pp. 614-626, Feb 2005.
22. K. D. Stephan, "Inter-injection locked oscillators for power combining and phased arrays," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-34, pp. 1017-1025, Oct. 1986.
23. P. Liao, R. A. York, "A new phase-shifterless beam-scanning technique using arrays of coupled oscillators," *IEEE Transactions on Microwave Theory and Techniques*, vol. 41, no. 10, pp. 1810-1815, Oct. 1993.
24. R. Adler, "A study of locking phenomena in oscillators," *Proceedings of the IRE*, vol. 34, pp. 351-357, June 1946.
25. J.F. Buckwalter, A. Babakhani, A. Komijani and A. Hajimiri, "An integrated subharmonic coupled-oscillator scheme for a 60-GHz phased-array transmitter," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 12, part 2, pp. 4271-4280, Dec. 2006.
26. A. Natarajan, A. Komijani, X. Guan, A. Babakhani and A. Hajimiri, "A 77-GHz Phased-Array Transceiver With On-Chip Antennas in Silicon: Transmitter and Local LO-Path Phase Shifting," *IEEE Journal of Solid State Circuits*, vol. 41, no. 12, pp. 2807-2819, Dec. 2006.
27. M. Chryssomallis, "Smart antennas," *IEEE Antennas and Propagation Magazine*, vol. 42, no. 3, June 2000.
28. H. Krishnaswamy and H. Hashemi, "A Fully Integrated 24GHz 4-Channel Phased-Array Transceiver in 0.13 μ m CMOS Based on a Variable-Phase Ring Oscillator and PLL Architecture," in *ISSCC Digest of Technical Papers*, vol. 50, Feb. 2007, pp. 124-591.
29. H. Krishnaswamy and H. Hashemi, "A 4-Channel 24-27 GHz UWB Phased Array Transmitter in 0.13 μ m CMOS for Vehicular Radar," in *Proceedings of 2007 IEEE Custom Integrated Circuits Conference*, Sep. 2007, pp. 753-756.
30. H. Krishnaswamy and H. Hashemi, "A Variable-Phase Ring Oscillator and PLL Architecture for Integrated Phased Array Transceivers," submitted to the *IEEE Journal of Solid State Circuits*.
31. H. Krishnaswamy and H. Hashemi, "A 4-Channel 24-27 GHz UWB Phased Array Transmitter in 0.13 μ m CMOS," submitted to the *IEEE Journal of Solid State Circuits*.
32. K. R. Carver, W. K. Cooper and W. L. Stutzman, "Beam-pointing errors of planar-phased arrays," *IEEE Transactions on Antennas and Propagation*, vol. 21, no. 2, pp. 199-202, March 1973.
33. J. Shen and W. Pearson, "The phase error and beam-pointing error in coupled oscillator beam-steering arrays," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 1, pp. 386-393, Jan. 2005.
34. U. R. Pfeiffer, J. Grzyb, D. Liu, B. Gaucher, T. Beukema, B. A. Floyd, and S. K. Reynolds, "A chip-scale packaging technology for 60-GHz wireless chipsets," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 8, pp. 3387-3397, Aug. 2006.
35. A. Babakhani, X. Guan, A. Komijani, A. Natarajan and A. Hajimiri, "A 77-GHz Phased-Array Transceiver With On-Chip Antennas in Silicon: Receiver and Antennas," *IEEE Journal of Solid State Circuits*, vol. 41, no. 12, pp. 2795-2806, Dec. 2006.
36. G. Estep, R. Gupta, T. Hampsch, M. Zaharovits, L. Pryor, C. Chen, A. Zaghloul and F. Assal, "A C-Band Beam-forming Matrix for Phased-array Antenna Applications," in *IEEE MTT-S International Microwave Symposium Digest*, vol. 3, May 1995, pp. 1225-1228.
37. J. Butler and R. Lowe, "Beam-Forming Matrix Simplifies Design of Electronically Scanned Antennas," *Electronic Design*, pp. 170-173, April 1961.
38. J. Blass, "Multidirectional Antenna: A New Approach to Stacked Beams," in *IRE International Conference Record*, vol. 8, part 1, 1960.
39. T. Chu and H. Hashemi, "A CMOS UWB Camera with 7x7 Simultaneous Active Pixels," submitted to *ISSCC 2008*.
40. R. Jee-Youl B. Kim, and I. Sylla, "A new low-cost RF built-in self-test measurement for system-on-chip transceivers," *IEEE Transactions on Instrumentation and Measurement*, vol. 55, no. 2, pp. 381-388, April 2006.

41. R. Staszewski, I. Bashir, and O. Eliezer, "RF built-in self test of a wireless transmitter," *IEEE Transactions on Circuits and Systems II*, vol. 54, no. 2, pp. 186-190, February 2007.
42. A. Fenn, D. Temme, W. Delaney, and W. Courtney, "The development of phased-array radar technology," *Lincoln Laboratory Journal*, vol. 12, no. 2, pp. 321-340, 2000.
43. H. Aumann, A. Fenn, and F. Willwerth, "Phased array antenna calibration and pattern prediction using mutual coupling measurements," *IEEE Transactions on Antennas and Propagation*, vol. 37, no. 7, pp. 844-850, July 1989.
44. T. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall, 1996.
45. S. M. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451-1458, October 1998.

Index

- 1/ f noise, 33, 43, 99
- F_{min} , 97–98, 113, 116
- G_{MAG} , 109
- MAG , 89, 132, 149
- MSG , 85, 87, 89, 110, 131, 132, 135
- NF_{min} , 25, 39–42, 44, 45, 47–49
- Z_0 , *see* characteristic impedance
- α , *see* propagation loss
- β , *see* propagation constant
- γ , *see* propagation constant
- f_T , 33, 36, 43, 44, 109, 113, 118, 121, 148
- f_{max} , 33, 38, 39, 43, 44, 80–85, 87, 89, 110, 113, 118, 121, 129, 148
 - current density, 84
- g_{ds} , 43
- 60 GHz, 2, 12

- active antennas, 250
- active device, 79–105
 - design, 80–88
- activity, 81
- air bridge, 65
- amplifier
 - design challenges, 109–111
 - differential, 129
 - low-noise, *see* LNA
 - multi-stage, 110, 129–138
- array factor, 248–250, 276, 279–282, 286
- array-induced ISI, 251, 252
- attenuation characteristics, 18
- automotive radar, 18

- Back-End of Line, *see* BEOL
- balanced line, *see* differential line
- balanced mode, *see* odd mode
- balun, 68, 76, 129, 146
 - insertion loss, 146
- BC, *see* body contacted
- beam-pointing angle, 248, 280, 284, 286, 287
- beam-pointing error, 280, 281
- beamformer, 258–260
- beamwidth, 249, 279
- BEOL, 39, 49–51, 55, 111
- BiCMOS, 33
- Bloss matrix, 289–291
- body contacted, 42
- branch-line hybrid, 141
- BSIM3, 89, 91, 93, 96
- BSIM4, 99
- bulk CMOS, 35, 41, 44
- Butler matrix, 289, 291

- calibration of phased arrays, 247, 269, 275, 285, 294
- capacitance, 61–64, 67–70, 72–76, 78–81, 83, 85, 88, 89, 92, 102
 - parasitic, 88, 92, 111
- capacitive degenerated VCOs, 168–170
- capacitors, 70–76, 80, 81, 89, 106
 - bypass, 72, 74, 119, 124, 125, 131
 - coupling, 72, 73
 - MIM, 71, 73, 78, 119, 131
 - model, 73
 - MOM, 72–74, 78, 124, 131, 136
 - MOS, 69, 74, 75
 - varactor, 69, 74, 75, 77
- Cartesian Combiner, 264
- cascode, 121, 124, 129–132, 142, 149
 - model, 132
 - noise, 124, 130
 - shared-junction, 131–132, 136
 - stability, 131
- characteristic impedance, 59–62, 66–69, 77
- Clapp oscillators, 174

- CMOS, 8, 34
 - technology scaling, 9
- Colpitts oscillators, 171–175
- comb capacitor, *see* MOM capacitor
- common base, 115
 - LNA, 118
- common emitter, 115
 - LNA, 121
- common gate
 - amplifier, 128–129
 - noise, 128
- common mode, *see* even mode
- common mode rejection, 129
- common source, 89
 - amplifier, 124–128
- compact models, 79
- compression point, 127
- contact, 87
- coplanar waveguide, *see* CPW
- copper, 30
- coupled oscillator arrays, 267, 269
- coupler, 146
 - rat-race, 146
- CPW, 64, 65, 77, 93, 119, 124, 136, 142, 149, 204, 300
 - conductor-backed, 209
 - substrate-shielded, 204
- cross-coupled oscillators, 164–170
- current re-use, 132
- cutoff, *see* f_T
- de-coupling capacitors, *see* capacitors, bypass
- design rules, 25, 52–54
- dielectric loss, *see* substrate loss
- differential amplifier, 129
- differential lines, 67, 78
- differential mode, *see* odd mode
- digital array, 255, 270, 272
- distributed
 - circuit, 59, 67, 70, 99
 - effects, 80, 82, 89
 - elements, 97, 99, 109–111, 118, 146
- distributed oscillators, 179–180
- dividers
 - direct-injection, 191
 - see* frequency dividers, 188
- down-conversion mixers, *see* mixers
- DRC, *see* design rules
- driving safety, 19
- dummy fill, *see* dummy metal
- dummy metal, 53, 66
- eddy currents, 65, 69, 71
- efficiency, 127
- EM simulation, *see* field solver
- emitter degeneration, 115
- even mode, 67–69
- extrinsic parasitics, 39
- FB, *see* floating body
- FCC, 20
- field solver, 54, 64–66, 85, 111
- finger capacitor, *see* MOM capacitor
- flicker noise, 99
- floating body, 42
- F_{min} , *see* F_{min}
- frequency cutoff, *see* f_T
- frequency dividers, 180–192
 - injection-locked, 189–192
 - operation range, 180–181
 - regenerative, 185–188
 - static, 182–184
- frequency doubler, 149
- Friis formula, 112
- gain, 111
- gate contact, 83, 87
- gate resistance, *see* resistance, gate
- General Purpose (GP), 34
- GP, *see* General Purpose
- grounding, 59, 64–70, 74, 76, 84, 111, 122, 129, 131, 145
- high-K, 29, 32
- high-pass/low-pass phase shifter, 262, 263
- hybrid, 140
- IIP3, *see* third-order intercept point
- image rejection, 121, 122
- IMD, 49, 50, 66
- impedance match, 111, 112
- inductance, 59, 61–64, 67–70, 72, 74, 79, 81, 83, 89, 92
 - lead, 70, 71, 73, 89, 92
 - parasitic, 89, 111
- inductive degeneration, 115, 116, 121
- inductors, 69–71, 81, 153
 - ring, 70, 71, 75, 77
 - spiral, 70, 77, 141, 152, 153
 - structure, 167
- injection-locked dividers, 189–192
- input match, 124
- integration, 27, 34, 48–50, 55, 129
- Inter Metal Dielectric, *see* IMD
- isolation, 117, 129, 130
- kink effect, 43
- large-signal model, 90–97

- measurement, 93–97
 - table-based, 90
- LC tank, *see* resonator
- LDD, 36, 38
- linearity, 111, 112, 129, 137
- LNA, 111–138
 - design, 116
 - noise figure, 119, 122
- LO phase-shifting, 265, 266, 270–272
- LO power, 138
- load-pull, 94
- loop inductor, *see* ring inductor
- loss tangent, *see* substrate loss
- low doped drain, *see* LDD
- low noise amplifier, *see* LNA
- Low Power (LP), 34
- low-K, 30, 66
- LP, *see* Low Power
- MAG, *see* G_{MAG}
- Mason's unilateral gain, *see* unilateral gain
- matching, *see* impedance matching
- maximum available gain, *see* G_{MAG}
- maximum oscillation frequency, *see* f_{max}
- maximum stable gain, *see* G_{MSG}
- medical applications, 22
- metal, 27–30, 39, 40, 49–51, 53, 111
 - density, 53–55
 - gate, 29, 32
 - loss, 60, 61, 63–66, 69, 77, 78, 88, 89, 208
- microstrip, 64, 65, 67
- Miller capacitance, 36, 38, 68, 141, 149
- Miller dividers, 185–188
- MIMO, 23, 245, 271, 292, 293
- minimum achievable noise, *see* F_{min}
- mixer, 138–147
 - active, 138
 - conversion gain, 142
 - dual-gate, 142–145
 - gilbert cell, 145–147
 - passive, 138
 - single-balanced, 146
 - single-transistor, 138–142
- mm-wave oscillators, 176–177
- mobility, 30, 38
- modeling
 - capacitors, 73–74
 - extraction, 93
 - large-signal, 90–97
 - noise, 97–105
 - parameters, 79
 - transistors, 79–90
 - transmission line, 65–66
- MOS11, 99
- MSG, *see* G_{MSG}
- multi-beam arrays, 271, 288, 289
- NF, *see* noise figure
- noise
 - BSIM, 99
 - cascode, 102
 - conductance, 113, 114
 - correlation, 99, 114
 - excess drain noise, 99, 101
 - factor, *see* noise figure
 - figure, 46, 97, 111, 112, 136
 - match, 113, 116
 - measurements, 98, 101–105, 137
 - model, 97–105
 - models, 98–101
 - parameters, 113, 115, 116
 - Pospiezsalski, 97, 101
 - PRC model, 100–101
 - resistance, 113, 114
 - sensitivity parameter, 113
 - Van Der Ziel, 99
- noise figure, minimum, *see* NF_{min}
- non-linearity, 96
- non-quasi-static resistance, *see* resistance, non-quasi-static
- notch, 121, 122
- NQS, *see* resistance, non-quasi-static
- nulls, 249, 271, 281
- odd mode, 67–70
- pads, 119, 124, 211
- parasitics, 80, 87
- partially depleted (PD), 42
- passive devices, 59–78
- passive loss, 203
- path-sharing architecture, 259, 261, 271
- patterned ground shields, 50
- PD, *see* partially depleted
- permittivity, 50
- phase
 - delay, 110
 - interpolator, 264, 270, 271
 - noise, 162–164
 - shifting, 255, 256, 265, 271, 272, 280, 283–285, 287
 - velocity, 62, 67–69
- phased array, 21, 245–253, 255, 260, 261, 265, 266, 269–273, 276–280, 285–289, 294
- Pierce oscillators, 175
- polar modulator, 264
- poly gate resistor, 82–83
- Pospiezsalski, *see* noise, Pospiezsalski

- power combining, 140, 217
 - distributed active transformer, 218
 - electrical funnel, 221
 - electrical lens, 223
- power match, 113, 116
- power transistor, 211
 - breakdown voltage, 215
 - cascode, 214
 - cut-off frequency, 215
 - gain, 211
 - layout, 233
 - stability, 211
- process variation, 125
- propagation
 - constant, 59, 62, 70
 - loss, 61, 68, 136
- push-push oscillators, 178–179
- Q factor, *see* quality factor
- quadrature balanced mixer, 141
- quality factor, 59, 61, 62, 64, 69–75, 77, 78, 106, 115
- quantization error in phased arrays, 285
- quarter-wave line, 122, 146
- quasi-static assumption, 79
- radar, 18–21
- receiver
 - CMOS, 148–150
 - SiGe, 151–154
- reflection-type phase shifter, 263, 264, 272
- regenerative dividers, 185–188
- resistance
 - gate, 82, 87, 92, 97
 - non-quasi-static, 101
 - spreading, 92
- resistive losses, 89
- resonance, 124, 132, 136, 146
- resonance frequency, 61, 71, 72, 74, 77, 78
- resonator, 59, 61, 69, 71, 72, 77–78, 108
- return current, 59, 67–70
- reverse gain, 130
- RF pads, *see* pads
- ring inductor, *see* inductor, ring
- round table device, 86–88, 124, 132
 - noise, 102
- s-parameters, 119
 - bias-dependent, 93
- scaling, 27
- self-resonance frequency, *see* resonant frequency
- series loss, *see* metal loss
- Shannon's Theorem, 3
- shared-junction cascode, *see* cascode
 - shared-junction
- shield, 64, 69, 71
- shunt loss, *see* substrate loss
- shunt-peaking, 192
- sidelobe, 249, 250, 279, 281, 282
- Sidelobe Rejection Ratio, 280–284
- SiGe HBT, 25–27, 33, 34, 47, 48
- SiGe transceiver, 151–155
- silicon cost, 8, 27
- Silicon Heterojunction Bipolar, *see* SiGe HBT
- skin-effect, 64–66, 89, 111, 203, 208
- slow wave, 61, 69, 70, 77
- small-signal model, 89–90
- smart antennas, 270, 271
- SNR, 112
- SOI, 35, 41, 42, 44
- spiral inductor, *see* inductor, spiral
- stability, 109, 110, 119, 124, 125, 136, 141, 142, 145, 149, 157, 174, 175, 184, 195, 211–215, 233
 - circles, 127
 - factor, 85, 124, 130, 132, 135
- static dividers, 182–184
- strain, *see* mobility
- stub, 118, 122, 141, 142, 196, 233
 - matching, 119
- substrate
 - loss, 60, 61, 63–66, 76–78, 83, 84, 204
 - network, 80, 82
 - resistivity, 50, 51, 119
- switched transmission line phase shifter, 261, 262
- third-order intercept point, 112
- THz, 23
- timed array, 251, 254–257, 259, 271, 277, 291, 294
- transformers, 75–76, 129, 147
- transistor
 - extrinsic, 80, 89
 - intrinsic, 80
 - layout, 38, 39, 80–89
 - modeling, *see* modeling, transistors
 - multi-finger, 81
 - noise parameters, 113
 - optimal layout, 84
 - power, *see* power transistor
 - sizing, 116
 - small-signal model, *see* small-signal model
 - wiring, 85–88
- transmission line, 49–54, 59–72, 77–79, 106, 107, 111, 124, 141, 142, 153

- varactor loaded, *see* varactor,loaded
- transmission line
- trombone line, 257, 258, 261
- tuned ring oscillator, 266, 267, 273
- tuning range, 162, 166
- unilateral, 117
- unilateral gain, 81–82, 87, 128
- unity power-gain frequency, *see* f_{max}
- UWB, 1
- Van Der Ziel, 99
- varactors, 74–75
 - loaded transmission line, 260, 261
- variable-phase ring oscillator, 273–277
- VCO, 149
- vector-summing phase shifter, 264
- velocity, *see* phase velocity
- vertical parallel-plate capacitor, 211
- via, 72, 83, 111
- voltage-controlled oscillators (VCOs),
 - 161–180
 - capacitive degenerated, 168–170
 - Clapp, 174
 - Colpitts, 171–175
 - cross-coupled, 164–170
 - distributed, 179–180
 - inductor structure, 167
 - mm-wave, 176–177
 - phase noise, 162–164
 - Pierce, 175
 - push-push, 178–179
 - tuning range, 162, 166
- VPRO-PLL Phased Array Architecture,
 - 273–279
- wavelength, 110
- WiFi, 1
- wirebond, 211
- X-ray, 22

Continued from page ii

A Practical Introduction to PSL

Cindy Eisner and Dana Fisman

ISBN 978-0-387-35313-5, 2006

Thermal and Power Management of Integrated Systems

Arman Vassighi and Manoj Sachdev

ISBN 978-0-387-25762-4, 2006

Leakage in Nanometer CMOS Technologies

Siva G. Narendra and Anantha Chandrakasan

ISBN 978-0-387-25737-2, 2005

Statistical Analysis and Optimization for VLSI: Timing and Power

Ashish Srivastava, Dennis Sylvester, and David Blaauw

ISBN 978-0-387-26049-9, 2005