# HIGH-EFFICIENCY MILLIMETER-WAVE SIGNAL GENERATION FOR COMMUNICATION, IMAGING, AND RADAR APPLICATIONS

by

Amirahmad Tarkeshdouz

M.Sc., Cornell University, 2016

## A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

## THE REQUIREMENTS FOR THE DEGREE OF

## DOCTOR OF PHILOSOPHY

in

## THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES (ELECTRICAL AND COMPUTER ENGINEERING)

## THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

May 2019

© Amirahmad Tarkeshdouz, 2019

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

High-Efficiency mm-Wave Signal Generation for Communication, Imaging, and Radar Applications

submitted by	Amirahmad Tarkeshdouz	in partial fulfillment of the requirements for		
the degree of	Doctoral of Philosophy			
in	Electrical and Computer Engineering			
Examining Con	mmittee:			
Co-supervisor				
Shahriar Mirat	obasi			
Co-supervisor				
John Madden				
Supervisory C	ommittee Member			
Nicolas Jaeger				
University Exa	aminer			
Roman Krems				
University Exa	aminer			

## Additional Supervisory Committee Members:

Supervisory Committee Member

Supervisory Committee Member

### Abstract

Over the last decade, integrated circuit technology has witnessed a surge in interest toward millimeter-wave (mm-wave) frequencies (30 to 300 GHz), mainly due to the increasing number of promising applications enabled by and implemented in this frequency band. Imaging and biomolecular spectroscopy are among the main applications of mm-wave frequencies. The ability to characterize nanostructures could have a dramatic impact on basic research in materials design and biosensing. Many magnetic-resonance (MRI) based imaging techniques used to explore biological structures benefit partly from being carried out at very low temperatures. Lack of transistor models at such low temperatures is a major challenge for circuit design. In the first part of this thesis, we address the fundamentals of signal generation at low temperatures and develop a custom-designed, application-specific signal source for an imaging experiment implemented in a 0.13-µm bipolar/complementary metal-oxide semiconductor (BiCMOS) process. The basics of signal generation outlined in this part can be expanded to include other needs for cryogenic applications. Radio technology has evidenced a rapid evolution with the launch of the analogue cellular systems in 1980s. As technology evolves, so does our expectation on how we could use it. Growing demand for high throughput and capacity on one end, and the ever-increasing desire towards ubiquitous connectivity on the other end, have strained current schemes and standards, calling for alternative novel high capacity systems. To address capacity demands, mm-wave spectrum provides a unique opportunity due to the availability of wider and unpopulated frequency bands which better facilitate high-speed communications. In this context, the emerging 5<sup>th</sup> Generation Mobile Network (5G) is also expected to use mm-wave bands. However, energy efficiency

remains a critical issue for all of these applications and has become an important challenge to deal with in face of higher performance requirements. In the second part of this thesis, we propose circuit solutions and strategies for efficient signal generation at very high-frequency bands and demonstrate state-of-the-art mm-wave signal sources implemented in 65-nm and 0.13-µm CMOS with DC-to-RF efficiencies of around 15%, the highest efficiencies reported to date for CMOS oscillators operating in the vicinity of and beyond 100 GHz.

## Lay Summary

This thesis summarizes the efforts toward approaching the limits of silicon-based technologies for efficient signal generation at millimeter-wave frequencies. We push the limits of electronic design in various silicon-based integrated circuit technologies and exploit transistor capabilities to achieve high speed, low power consumption, an improved performance overall. To validate the concepts developed in this research, proof-of-concept prototypes are implemented in different technologies and are successfully tested. The proposed design approaches can be employed to develop circuits and systems that further enable applications such as imaging, radar, sensing and communications for medical, security and Internet of Things (IoT) systems.

## Preface

I, Amirahmad Tarkeshdouz, am the **principal contributor** of the entire chapters in this dissertation. Professor Shahriar Mirabbasi and Professor Ehsan Afshari have supervised the research by providing technical discussions and editing the manuscript. During my PhD, I have collaborated with a number of scholars and research groups as reflected in publication list below. I would like to thank them all, especially my great friend and collaborator, Milad Haghi Kashani. He and I equally contributed to the material of the Chapter 5, where I primarily assisted with the development of the theory for this chapter as well as providing technical assistance with the design, fabrication, and measurement of the proposed voltage-controlled oscillator. I also co-prepared and wrote the manuscript of the publication. Please find below the list of the published/accepted works which partially form the main chapters of this thesis:

- A. Tarkeshdouz, A. Mostajeran, S. Mirabbasi, E. Afshari, "A 91 GHz Fundamental VCO with 6.1% DC-to-RF Efficiency and 4.5 dBm Output Power in 0.13-μm CMOS," published in *IEEE Solid-State Circuits letter (SSC-L)*, vol. 1, no. 4, pp.102-105, April 2018. (Chapter 3)
- A. Tarkeshdouz, M. Haghi Kashani, E. Hadizadeh Hafshejani, S. Mirabbasi, E. Afshari, "An 82.2-to-89.3 GHz CMOS VCO with DC-to-RF Efficiency of 14.8%," accepted to be presented at IEEE Radio Frequency Integrated Circuits (RFIC) Symposium, June 2019, (Chapter 4)
- M. Haghi Kashani, A. Tarkeshdouz, R. Molavi, E. Afshari, and S. Mirabbasi, "A Wide Tuning Range Low-Phase-Noise mm-Wave CMOS VCO with Switchable Transformer-Based Tank," *IEEE Solid-State Circuits letter (SSC-L)*, vol. 1, no. 4, pp. 82-85, April 2018. (Chapter 5)

## **Table of Contents**

Abstract.	iii
Lay Sum	naryv
Preface	vi
Table of <b>(</b>	Contents vii
List of Ta	blesx
List of Fig	gures xi
List of Ab	breviations xvii
Acknowle	dgementsxviii
Dedicatio	n xviiii
Chapter 1	: Introduction1
1.1	mm-Wave Technology Overview
1.2	mm-Wave Signal Generation
1.3	mm-Wave Signal Generation at Cryogenic Temperatures5
<b>1.4</b>	mm-Wave High-Power Signal Generation7
1.5	High-Efficiency Broad-Band mm-Wave Signal Generation8
1.6	Outline9
Chapter 2	2: A SiGe HBT mm-Wave Oscillator for MRFM Imaging11
2.1	Introduction11
2.2	Low-Temperature Electronics
2.2.1	Magnetic Resoance and Sample Considerations
2.2.2	Cryogenic Device Physics Considerations
2.2.3	Circuit Design Given the "Hypothetical Ongoing Scaling"
2.2.4	An Efficient High-Power Fundamental Oscillator
2.3	Implementation and Circuit Design
	vii

2.4	Prototype Measurement and Results	
2.5	Conclusions	
Chapte	r 3: Efficient Signal Generation Close-to-f <sub>max</sub>	40
3.1	Introduction	
3.2	A High-Efficiency Class-E-Like Oscillator Topology	
3.2	2.1 Drain Efficiency in a Signle-Transistor Power Amplifier	
3.2	2.1. The Start-up/Loop Gain Condition for Oscillation Build-up	
3.3	Circuit Implementation Details	
3.4	Measurement Setup and Experimental Results	
3.5	Conclusion	
Chapte	r 4: Efficient mm-Wave Signal Generation on CMOS	64
41	Introduction	64
4.2	A High-efficiency Single-Active-Device-Based Oscillator	
4.3	Circuit Implementation Details	
4.4	Measurement Setup and Experimental Results	
4.5	Conclusion	
Chapte	r 5: A mm-Wave VCO with Switchable Triple-Coupled Transformer	78
5 1	Introduction	78
5.2	The Coarse Tuning Approach	
5.2	The Proposed Triple-Coupled Transformer Model	
5.2	<ul> <li>Verification of the Proposed Model</li> </ul>	
5.3	VCO Implementation and Design	
5.4	Measurement Results	
5.5	Conclusion	
Chapte	r 6: Conclusion	
<u> </u>		111
6.1	Achievements	
6.1	2 A SIGE HB1 mm-wave Source for MRFM Imaging	
6.1	.3 Close-to- <i>f<sub>max</sub></i> Efficient Signal Generation	
6.1	.4 Super-Efficient mm-Wave Signal Generation on CMOS	
		viii

	6.1.5	A mm-Wave VCO with Switchable Triple-Coupled Transformer	113
Refe	erences	5	114
Арр	endix.		129
А	.1 M	Magnetic Resonance Force Microscopy Imaging (MRFM)	129
А	.2 A	Semi-Quantitative Mapping Between Cooling and Device Scaling	130
А	.3 S	emi-Quantitative Temperature Prediction	139

## List of Tables

Table 2.1 Comparison with the prior art	34
Table 2.2 Performance summary of the fabricated cryogenic VCO at multiple temperatures	38
Table 3.1 Table of Comparison with the State-of-the-Art	62
Table 4.1 Performance Comparison of Millimeter-Wave Signal Sources in CMOS	78
Table 5.1 Performance summary of 60-GHz VCOs in 65-nm CMOS 1	09

# List of Figures

Figure 1.1	mm-wave spectrum	applications	ý
	min wave speed and	"Pp"	

Figure 2.1 $f_T$ versus the collector current for SiGe HBT with L <sub>E</sub> =3 $\mu$ m20
Figure 2.2 $f_{max}$ variations as a function of transistor size
Figure 2.3 The transconductance sensitivity as a function of emitter length versus base-emitter voltage
Figure 2.4 Gummel characteristics for the employed HBTs from BiCMOS9MW and BiCMOS055 processes
Figure 2.5 Change in the trend of Gummel characteristics due to (a) scaling (b) cooling with respect to room-temperature condition
Figure 2.6 The employed setting to achieve maximum PAE
Figure 2.7 Conventional cross-coupled structure
Figure 2.8 Pictorial representation of the base current in time
Figure 2.9 Base-voltage waveform evolution in time
Figure 2.10 The equivalent transistor embedding to achieve maximum PAE in oscillator 29
Figure 2.11 A simulated sample of base-voltage waveform evolution in time
Figure 2.12 The oscillation build-up in time (a) when the transistor's gain is not sufficient (b) when the oscillation swing reach the required $S_{Bf}$
Figure 2.13 The equivalent series resistance at the base to determine the input drive condition31
Figure 2.14 Forced-I <sub>B</sub> mode implementation to avoid NDR
Figure 2.15 The schematic view of the proposed efficient VCO
Figure 2.16 The schematic view of the proposed efficient VCO
Figure 2.17 The measurement setup for the proposed VCO at room temperature
Figure 2.18 fosc vs. control voltage at various temperatures
Figure 2.19 Measured DC-to-RF efficiency at the CPW output vs. V <sub>BE</sub>

Figure 2.20	Measuered phase	noise vs. foffset	for various tempe	eratures
0	1	011000	1	

Figure 3.1 Conceptual representation of the possible $i_d$ waveforms for various bias levels (a) The ideal case (b) When the transistor turns off (c) Closer to the ideal case (d) Resulted waveform beyond the optimum case
Figure 3.2 Simplified model of the mm-wave transistor in the circuit when the transistor is turned off
Figure 3.3 The fundamentla component of $(i_D)$ and DC component of $I_D$ as a function of the gate- bias level $(V_G)$
Figure 3.4 Gate-drain voltage of the single transistor
Figure 3.5 The output current of the transistor that also shows the corresponding cases in Fig. 3.1
Figure 3.6 The output waveforms of the final oscillator including $V_G$ , $V_D$ , and $I_D$
Figure 3.7 Equivalent negative resistance for both of the single-transistor and cross-coupled structure as a function of $C_8$
Figure 3.8 Effect of introduing an additional capacitor at the transistor source on the efficiency and power gain in the open-loop configuration
Figure 3.9 Schematic of the implemented high-efficiency fundamental VCO, the output buffer, and the matching network
Figure 3.10 The measured phase noise of the proposed VCO at 89.75 GHz offset frequency. The simulated PN at the same offset frequency is also denoted by the red dotted line for comparison
Figure 3.11 (a) Measured oscillation frequency $(f_{osc})$ and the calculated efficiency vs. $V_{tune}$ (b) Simulated oscillation frequency and the corresponding calculated efficiency vs. control voltage $(V_{tune})$
Figure 3.12 The meausred efficiency for different levels of supply voltage ( $V_{DD}$ ) with the corresponding frequency of oscillation ( $f_{osc}$ )
Figure 3.13 The die micrograph of the fabricated fundamental 91-GHz VCO

Figure 4.2 PAE versus $V_B$ for several voltage amplitudes on the gate	6
Figure 4.3 $P_{dc}$ versus $V_B$ for several swing levels	7
Figure 4.4 Gate voltage and drain current waveforms for different levels of swing and bias voltag	;e 8
Figure 4.5 Pictorial representation to implement the AFN for an increased swing	0
Figure 4.6 Z <sub>in</sub> vs. frequency	0
Figure 4.7 DC-to-RF efficiency versus frequency for the proposed AFN-based oscillator structur and the conventional cross-coupled oscillator	•е 1
Figure 4.8 The schematic view of the proposed 85.75 GHz VCO	2
Figure 4.9 The measurement setup for the proposed VCO	3
Figure 4.10 The annotated die micrograph of the fabricated fundamental 89.3 GHz VCO7	4
Figure 4.11 The measured oscillation frequency and output-power versus f <sub>tune</sub>	5
Figure 4.12 (a) The measured oscillation frequency and DC-to-RF efficiency vs. $V_{tune}$ (b) Th simulated oscillation frequency and efficiency vs. control voltage	le 5
Figure 4.13 The measured phase noise at 89.3 GHz7	6
Figure 5.1 A high-frequenc circuit model of an integrated inductor	2
Figure 5.2 The high-frequency circuit model of the propposed triple-coupled transformer 8	2
Figure 5.3 A simplfied circuit model of the proposed switachble triple-coupled transformer 8	3
Figure 5.4 Q <sub>eff</sub> and C <sub>2</sub> versus k <sub>23</sub> for given design parameters	7
Figure 5.5 Simplified pi-model of the proposed transformer when (a) $k_{23}=0$ (b) $k_{23}=1$	8
Figure 5.6 Simplfied pi-model of the proposed triple-coupled transformer	8
Figure 5.7 Effective quality factor versus L <sub>2</sub> with a constant L <sub>3</sub> 9	0
Figure 5.8 Effective inductance variations versus $L_2$ with a constant $L_3$	1
Figure 5.9 Circuit implementation of the differential switch	2

Figure 5.10 Effective inductance variations versus the off-state parasitic capacitance of the switch
Figure 5.11 Effective quality and effective inductance variations versus the off-state parasitics capacitance of the switch
Figure 5.12 Effective quality factor and effective inductance variations versus the coupling factor between the loads
Figure 5.13 Quality factor comparison between the proposed model and simualtion of the transformer
Figure 5.14 Flowchart of the summarized design procedure
Figure 5.15 Schematic of the proposed VCO
Figure 5.16 Simulated Q-factor over all switch states
Figure 5.17 Simulated effective inductance over the entire frequency tuning range 100
Figure 5.18 Frequency allocation of the four switch states based on effective quality factor 102
Figure 5.19 Simualtion results of the proposed geometry for different configurations (a) k (b) Q
Figure 5.20 The cross section view of the proposed switchable transformer
Figure 5.21 The top view of the proposed switchable transformer
Figure 5.22 Micrograph of the fabricated mm-wave VCO 107
Figure 5.23 The measured PN at 10-MHz offset over the entire frequency tuning range 107
Figure 5.24 The measured output power over the entire frequency tuning range
Figure 5.25 Measured tuning range versus control voltage 108

Figure A.1 The details and the primary principals of an MRFM imaging experiment	115
Figure A.2 Transit time components versus the vertical profile of a bipolar transistor	117
Figure A.3 Si and SiGe HBT schematic band diagram in the presence of Ge grading	118

Figure A.4 (a) Evolution of the current gain cut-off frequency vs. the inverse of collector current
density in STMicroelectronics SiGe BiCMOS technologies. The plot is reproduced from the data
in [14]. (b) The peak $f_T$ values of the plot in (a) versus the minimum emitter width of the transistors
Figure A.5 Published $f_{max}$ of SiGe HBTs with the corresponding cooled prototypes versus
temperature

## List of Abbreviations

BiCMOS	Bipolar/Complementary Metal-Oxide Semiconductor
CMOS	Complementary Metal-Oxide Semiconductor
CPW	Coplanar Wave Guide
DNP	Dynamic Nuclear Polarization
f <sub>max</sub>	Maximum Oscillation Frequency
FOM	Figure of Merit
FOM <sub>T</sub>	Figure of Merit with Tuning Range
fosc	Oscillation Frequency
$f_T$	Transit Frequency
FTR	Frequency Tuning Range
GCPW	Grounded Coplanar Wave Guide
gm	Transconductance
GSG	Ground Signal Ground
HBT	Heterojunction Bipolar Transistor
HFSS	High-Frequency Structure Simulator
IC	Integrated Circuits
IF	Intermediate Frequency
LNA	Low Noise Amplifier
LO	Local Oscillator
LTE	Low-Temperature Electronics
LTE	Long-Term Evolution
MIM	Metal Insulator Metal
mm-wave	Millimeter-Wave
MRFM	Magnetic Resonance Force Microscopy
MRI	Magnetic Resonance Imaging
NDR	Negative Differential Resistance
PA	Power Amplifier
PAE	Power Added Efficiency
PCB	Printed Circuit Board
PLL	Phase-Locked-Loop
PN	Phase Noise
Q	Quality Factor
RF	Radio Frequency
RFC	RF choke
RX	Receiver
SiGe	Silicon Germanium
THz	Terahertz
TL	Transmission Line
TX	Transmitter
VCO	Voltage Control Oscillator
VGS	Gate-to-Source voltage
Vth	Threshold Voltage

### Acknowledgements

During my graduate studies at UBC, I have had the privilege to interact with and learn from many brilliant and talented individuals who have enriched my PhD experience. I offer my enduring gratitude to the faculty, staff and my fellow students at UBC, who have inspired me to continue my work in this field. I would like to thank my adviser Professor Shahriar Mirabbasi who has always been a constant source of encouragement and support all along.

I would like to express my profound gratitude to Professor Ehsan Afshari, a terrific teacher, friend, adviser, brother and more importantly, an infinite source of never-ending kindness and support. He is just an amazing person and I am very fortunate to know him. I would also like to thank Professor John Madden for taking part in my dissertation exam committee, and his time and feedback.

I owe particular thanks to Dr. Ali Mostajeran who has been a brilliant collaborator and great friend. I enjoyed our discussions on various topics. I interacted with a great group of students at both UBC and University of Michigan including Milad Haghi Kashani, Mohammad Hossein Naghavi, and Sajjad Ohadi. I would like to thank them all. My warmest gratitude goes to Professor John Marohn and Dr. Corinne Isaac who have been brilliant collaborators and great colleagues.

Special thanks are owed to my parents and my brother, who have supported me throughout my years of education, both morally and financially.

Most important of all, I would like to thank my best friend, wife, and my love, Tina, who has been patiently with me in every step of the way in the past five years. I would like to thank her for her endless support and unconditional love.

**Dedicated to My Family** 

## **Chapter 1: Introduction**

What is millimeter-wave technology? 4th generation (4G) long term evolution (LTE) wireless technology currently employs lower frequency bands of the electromagnetic (EM) spectrum, generally below 5 gigahertz (GHz), to deliver data at improved speeds as compared to previous generations of mobile communication protocols. The speeds achieved via 4G LTE have dramatically transformed the way we communicate and manage our daily lives. However, as technology evolves, so do our expectations on how we need to harness and use it. As people and societies become more and more reliant on digital data and technologies, the demands created by increasingly data-intensive applications necessitate to reimagine the fundamentals of how the data flows. To this end, in the recently introduced 5<sup>th</sup> generation (5G) wireless systems, higher frequency bands— such as, the 28 and 39 GHz frequency bands—are being considered. These frequencies are within the millimeter-wave (mm-wave) band. In general, frequencies within 30 to 300 GHz are referred to as mm-wave frequencies. These higher frequency bands offer an increased absolute bandwidth, and thus facilitate carrying larger amounts of data at higher speeds and with a negligible (almost) zero latency, or lag. To intuitively explain, we can think of mm-wave spectrum as a superhighway where fleets of trailers transport gigantic shipment packs of data around in all directions with no restriction on speeds. Higher frequencies create the condition where tens of thousands of gigantic lanes can be put together instead of the individual small lanes. This is an ideal choice for accommodating a massive increase in data demands brought in by data-intensive applications such as mobilefirst users, connected homes, cloud gaming systems, self-driving vehicles, and IoT (Internetof-Things) sensors. Historically, mm-wave bands have been used almost exclusively for

government and non-commercialized applications due to the costly nature of semiconductor technologies such as Gallium Arsenide that could operate at such high frequencies. In the recent years, the maturation of silicon technologies such as CMOS and SiGe BiCMOS has created the possibility for both commercialized and non-commercialized products uniquely suited for mmwave operation. This has been pushing silicon deeper into mm-wave realm. Despite great advances, there are still bottlenecks resulting from the less-explored aspects of the highfrequency circuit design that need to be addressed by the low-cost and versatile CMOS electronics. For instance, in the context of battery-powered wireless/wireline connectivity solutions low power operation is one of the important design objectives and operation at high frequency while keeping the power consumption as low as possible is a nontrivial task.

#### 1.1 mm-Wave Technology Overview

Over the last decade, integrated technology has witnessed a surge in interest toward mm-wave frequencies (30 to 300 GHz), due to the increasing number of promising applications implemented in this frequency band. With an increased available absolute bandwidth, the lower frequency bands into mm-wave spectrum (<110 GHz) are presently being explored for most notably a variety of communication and imaging applications [1]. Imaging and bio/molecular spectroscopy were the first and main applications of mm-wave frequencies. Other examples include, automotive radar at 77-GHz band with a massive commercialized viability [2], passive radar imaging in W-band (75 to 110 GHz) [3], and multi-Gbps (giga-bit-per-second) wireless communication at 60 GHz [4], [5]. The effective short wavelength on silicon in such a band (in the order of millimeter) facilitates on-chip antennas design, thereby enabling the implementation of multi-element transceivers with electronic beamforming (mm-wave phased-array) that overcomes the large inherent path-loss and

fading at mm-wave frequencies [6]–[8]. This can improve range and capacity by providing access to spatial diversity and point-to-point wireless links. The critical requirement to implement such systems, especially for proliferation into mobile devices, is the power efficiency of the mm-wave links. Solutions that consume several watts of power to realize such links cannot be readily integrated into portable devices. In this context, the ability to generate low-power millimeter-wave signals in low-cost silicon technologies, for instance, can enable inexpensive in-home treatment options as a mm-wave medical benefit [9], [10]. Fig. 1.1 illustrates a few examples of mm-wave spectrum opportunities.



Figure 1.1 mm-wave spectrum applications

#### 1.2 mm-Wave Signal Generation

Despite all recent advances to realize highly integrated systems, due to the close proximity of the operating frequency with the transistors'  $f_{max}$  (maximum oscillation frequency), high-power signal generation at mm-wave frequencies is still a major challenge in solid-state electronics. In the context of CMOS, the challenges associated with signal generation is primarily attributed to three factors: (a) Despite the aggressive scaling trend in CMOS technology, the maximum oscillation frequency  $(f_{max})$  of transistors are still below 300 GHz. This number is even lower especially when the device interconnects are included. This imposes a theoretical limitation on the device which no fundamental oscillation nor power amplification is conceivable beyond [11], (b) As the gate oxide thickness shrinks in advanced technology nodes, the breakdown voltage of the device diminishes. This prohibitively reduces the maximum allowable voltage swing of the device, thereby limiting the generated output power, (c) The passive components which are fabricated in a CMOS process are very lossy, particularly with the presence of the lossy silicon substrate. This is originated from the slim nature of the top metal layers and the thick silicon substrate that collectively deteriorate the quality-factor (Q) of such elements. Due to these drawbacks, compound semiconductors (e.g., bulky group III-V solutions) have been traditionally used to demonstrate high-frequency signal generation at mm-wave and terahertz frequencies provided the high transit frequency  $(f_T)$  offered by these devices [12]–[14]. This solution, however, is very costly and inefficient and does not lend itself into integration. In the recent years, SiGe and CMOS transistors have also been employed to demonstrate signal generation using fundamental and push-push oscillator structures [15]–[20]. Even with demonstrating decent performances, the oscillation frequency  $(f_{osc})$  in all of these works remains well below the transistors' maximum oscillation

frequency ( $f_{osc} \sim 0.5 * f_{max}$ ). The question arises is whether these oscillators are capable to exploit the full capacity of the devices in terms of the signal frequency and output power. In [21], the authors address this fundamental question by investigating the effect of oscillator topology and the quality factor of the passive components using the activity condition of transistors [22]. The result is an oscillation frequency very close to the  $f_{max}$  with a decent generated output power. This facilitates on-chip high-power signal generation at frequencies close to or beyond  $f_{max}$  which is not a trivial task to accomplish. The limited  $f_{max}$  of today's silicon CMOS technology at mmwave frequencies, rules out the choice of generating the high-frequency signal independently and then utilizing a power amplifier (PA) to efficiently boost the close-to- $f_{max}$  generated signal's power. Motivated by this finding, we investigate circuit techniques required to realize highfrequency fundamental oscillators.

### 1.3 mm-Wave Signal Generation at Cryogenic Temperatures

Today's need for cryogenic electronics are very diverse and include a set of varied applications such as deep-space probes and planetary mission sensors, cooled IR (Infrared) detectors arrays, ultra-low-noise radio-astronomy receivers, quantum computing, and health imaging systems. There are clearly compelling reasons (e.g., low cost) that push toward utilization of conventional electronics to implement such applications. It is a generally true that a cryogenic environment presents a serious reliability issues to conventional electronics, requiring the designer to apply changes to the circuit topology and/or system architectures. Recently, chip-scale high-frequency oscillators have been developed for magnetic resonance spectroscopy and imaging applications. Many of the associated experiments in such applications either take place at cryogenic temperatures or would benefit from the use of low-noise oscillators operating at very low temperatures [23]-[31]. Several chip-scale radio-frequency (RF) transmitters and receivers suitable for nuclear magnetic resonance (NMR) spectroscopy and imaging measurements have been demonstrated so far [23]-[25]. There are also precedents for exciting and detecting electronspin resonance (ESR) using single-frequency CMOS circuits operating at frequencies ranging from 9 GHz to 146 GHz and at temperatures of 300 K, 77 K, 30 K, 10 K and 4 K [26]-[30]. A variablefrequency chip-scale 14-GHz source has recently been developed for use in electron spin resonance experiments at room temperature [31]. Even with the above developments, there is still a high demand for increasing the sensitivity in the imaging experiments. For instance, a highresolution microwave-enhance magnetic resonance force microscopy (MRFM) imaging experiment calls for the design of higher frequency chip-scale microwave sources that oscillate up to 200 GHz, are tunable, have an output power of 100s of microwatts, and operate down to 4.2 K. Lack of device models that are valid at such low temperatures is one of the major obstacles for designing cryogenic circuits. Transistor models in simulation packages are typically optimized for operations at or near room temperature. These tools rarely include low-temperature models below -50 °C. Thus, they become ineffective for designs at cryogenic temperatures (-150 °C). In [32], test structures are fabricated and measured to develop a low-temperature model for CMOS transistors. The developed custom-model is then incorporated in to the simulation packages to design circuit blocks. This approach, in most cases, may be troublesome, time-consuming, and inefficient. Recently, people have utilized silicon-germanium hetero-junction bipolar transistors (SiGe HBT) as the inherently "environmentally invariant" platform to design cryogenic circuit blocks. In this approach, given the knowledge that HBT functions well across a wide range of temperatures [33], the circuit is designed at room temperature and expected to operate even

superiorly at cryogenic temperatures [34]. This method also entails its own difficulties as the method does not design for the targeted optimized performance at low temperatures. Motivated by the above findings, we investigate procedures for designing circuit blocks for operations at low temperatures that sidestep the need to develop and test low-temperature transistor models and at the same time effectively optimize the circuit performance for low temperatures.

#### 1.4 mm-Wave High-Power Signal Generation

As far as signal generation for mm-wave applications is concerned, a signal source that can generate high radiation power to overcome large propagation loss at this frequency range is a highly desirable building block. Typically imaging, radar, and communication application require about 1 to 10 mW of power (equal to 0 to 10 dBm). As the frequency of operation in such applications normally lies within the proximity of the transistor cut-off frequency  $(f_T)$ , or in most cases beyond this frequency, therefore, no circuit block can be realized to proceed and amplify the output power of the generated signal. This mandates designing a power-oscillator as an alternative approach. To address the issue of power generation close to  $f_{max}$  in CMOS, the device nonlinearity and harmonic generation are the schemes that have been utilized in prior art. The signal sources based on such principle can be partitioned into two categories, first frequency multipliers and second harmonic VCOs. The frequency multipliers normally demonstrate high bandwidth. They, however, introduce a large power conversion loss that requires implementation of power-hungry PAs inside the multiplier chain to boost the power level [35]. This increases the silicon area as well as the DC power consumption. Additionally, the power variation increases across the bandwidth due to inter-stage mismatches. These multipliers also need a large-power and widetuning-range fundamental signal source to be driven, introducing another challenge. In comparison, the second nonlinear circuit category, i.e., the harmonic oscillator, has the advantage of being self-sustainable, compact, and power efficient with no inter-stage mismatch issue. The output power is competitive in such oscillators to that of the frequency doublers, especially with multi-cell power combining solutions. This, however, is also very inefficient in majority of designs. The state-of-the-art value for DC-to-RF efficiency of such sources barely exceeds a few percent [36], meaning that for achieving 1 mW of power almost 100 mW must be wasted as DC power. Therefore, it is desired to adopt novel design approaches to design high-power single block fundamental oscillators and fully exploit the transistors' capabilities while achieving a high value of DC-to-RF efficiency for power generation.

#### 1.5 High-Efficiency Broad-Band mm-Wave Signal Generation

In addition to the signal power level and efficiency, another challenge in CMOS mm-wave sources is the output frequency bandwidth. It is especially important for material identifications using mm-wave spectroscopy. For example, prior research shows that many types of materials exhibit vibrational resonance between 200 GHz to 300 GHz. To obtain such spectrum, a broadband radiation source is required. As another example, many frequency-modulated continuous wave (FMCW) radar application require a wide tuning range VCOs. In such imaging radars the range resolution determines the minimum distance between two objects which can be resolved clearly. In LC-resonator-based voltage-controlled oscillators (VCO), varactors are placed in the resonator in order to tune the oscillating frequency. Although this tuning method works well at radio frequencies, it achieves low tunability at lower mm-wave frequencies due to the limited C<sub>max</sub>/C<sub>min</sub>. At these frequencies the varactor quality factor is low. This lowers the achievable output power and degrades the phase noise performance. Moreover, as the operation frequency increases, the

parasitic capacitances become comparable with the tuning varactors, thus dominating the tank and limiting the tuning capability. These challenges impose an important trade-off in the design of high frequency oscillators. Therefore, CMOS oscillators above 100 GHz with high output power should not use varactors and as a result their frequency cannot be tuned or, otherwise, tunable oscillators at these frequencies must yield very low output powers due to the use of tuning varactors. Consequently, power generation along with frequency tuning above 100 GHz is normally realized by employing frequency multipliers. Frequency multiplication per se requires a high-power external source which is not desirable in a fully integrated mm-wave source, in addition to the problems mentioned earlier. To address this challenge, prior works at lower frequencies (< 60 GHz) propose tuning methods for oscillation frequency without using varactors. The magneticallytuned and the transconductance-tuned VCOs are two examples where instead of capacitive tuning, the oscillation frequency is changed by manipulating the inductive loads. In such approach, the inductor of the tank is effectively tuned through using different circuit techniques. These techniques still require additional active devices inside the oscillator tank in many cases that can undermine the power efficiency or the phase noise performance. A recent work employs an interpolative-phase tuning technique in an LC ring oscillator at the mm-wave frequency range to enhance the tuning range [37]. All these techniques have been used to generate output powers well below the cut-off frequency of the transistors. We will demonstrate design techniques so that the trade-off between frequency tuning and power efficiency in conventional fundamental VCOs is largely resolved.

#### 1.6 Outline

The objectives, contributions, and organization of the thesis are as follows:

- Circuit design techniques are employed to realize signal generation across a wide range of temperatures, with a focus on cryogenic temperatures in Chapter 2. This paves the way for implementation of many cryogenic applications.
- 2. Design techniques are introduced to aim for developing efficient high-power oscillators at operating frequencies close to maximum oscillation frequency ( $f_{max}$ ) in Chapter 3. As the principle objective, we investigate the fundamental behavior of power amplifiers' operation to arrive at efficient high-frequency signal generation.
- A large-signal understanding of transistor operations is developed in Chapter 4 to realize monolithic oscillators with more than 10% DC-to-RF efficiency operating beyond 100 GHz with wide tuning range.
- 4. In Chapter 5, utilizing a proposed inductive tuning method, we present a low phase noise wide tuning range 60-GHz VCO. The proposed techniques can be used in many highfrequency communication and imaging applications.

We conclude the thesis in Chapter 6.

## Chapter 2: A SiGe HBT mm-Wave Oscillator for MRFM Imaging

#### 2.1 Introduction

Recent emerging applications of cryogenic electronics have drawn considerable attention toward this important growing market. Applications ranging from high-sensitivity cooled sensors and detectors, satellite communication systems, medical electronics (e.g., CT scanners), and very low noise receivers for astronomy demand a comprehensive study to explore cryogenic electronics [38]. Another area of interest is magnetic resonance imaging (MRI) where the three-dimensional structure of biomolecular complexes (e.g., proteins) can be studied at the nanometer scales. By mechanically detecting magnetic resonance, magnetic resonance force microscopy (MRFM) has demonstrated the ability to perform atomic scale MRI under cryogenic condition [39]. This is realized by imaging electron spin radicals affixed to a single copy of a molecule of interest [40] or by imaging nuclear spin resonance [39]. The current sensitivities in a proton imaging experiment can be even further increased by introducing the dynamic nuclear polarization (DNP) effect [41]. DNP enhances the proton polarization by letting spin polarization transfer from highly polarized electron spins to the lesser polarized nuclear spins. An electron spin polarization of near unity is achievable at temperatures near or below 4 kelvin and in static magnetic fields  $(B_0)$  above 6 T [42]. However, the electron spin resonance frequency required to saturate electron spins under such conditions exceeds 150 GHz. Record nuclear spin polarization enhancements (exceeding 100×) have been achieved at static fields greater than 7 T, where the electron-spin resonance frequency exceeds 200 GHz [43], [44]. While commercial sources operating at these high frequencies are available, the technical challenges associated with delivering 200 GHz microwaves to a vibration-sensitive, cryogenic MRFM experiment are still enormous [45]. The fundamental details of an MRFM imaging experiments has been outlined in the Appendix section (A.1). In this context, several groups have demonstrated chip-scale radio-frequency CMOS transmitters and receivers suitable for nuclear magnetic resonance spectroscopy [45] and/or imaging measurements [46]- [48]. Fixed-frequency CMOS oscillators have been used to excite and detect electron-spin resonance at frequencies ranging from 9 to 146 GHz and at temperatures of 300 K [49], 77 K [50], 30 K, 10 K [51] and 4 K [52]. An on-chip 14-GHz VCO has recently been developed for use in electron spin resonance (ESR) at room temperature [53]. However, none of the circuits and systems described in these works are developed as a custom-designed chip specifically optimized for operation at low temperatures. This work proposes a procedure to design a VCO optimized for operation at cryogenic temperatures. The proposed method sidesteps the need for fabrication and measurement of the test structures required to develop low-temperature transistor models. The required specifications for a single-molecule DNP-enabled MRFM experiment entails the design of a VCO oscillating at up to tens of GHz and generating 100s of microwatt of power at 4.2 K.

#### 2.2 Low-Temperature Electronics

#### 2.2.1 Magnetic Resonance and Sample Considerations

In all MRFM experiments carried out to date, microwaves are generated at room temperature and guided to the vacuum/cryogenic sample space using waveguides or coaxial cables. Using a *chip-scale* source to generate the microwaves in close proximity to the sample would simplify the MRFM experiment considerably and facilitate the delivery of the high-frequency irradiation needed to achieve both single-nitroxide-spin detection and high-efficiency DNP. In this section, we analyze the frequency, power, and efficiency required for the source to be used in an MRFM experiment. The <sup>1</sup>H- and electron-spin signal in force-gradient detected magnetic resonance experiments like those performed in [41] is proportional to the spin polarization. At achievable magnetic fields and temperatures, proton spin polarization is  $p_p \ll 1$  at thermal equilibrium. To increase the polarization of a sample's protons *via* the DNP effect, a paramagnetic dopant is added to the sample, microwaves are applied to saturate the electron-spin transitions in the dopant, and magnetization flows from electron spins to nuclear spins. The proton polarization achievable in a DNP experiment is, at its upper limit, equal to the electron-spin polarization. Thus, we want the electron-spin polarization to be as close to unity as possible, ideally in the range from 0.1 to 1. The polarization of the sample's electron spins ( $p_e$ ) is,

$$p_{\rm e} = \tanh\left(\frac{\hbar \,\gamma_{\rm e} B_0}{2k_{\rm B}T}\right) \tag{2.1}$$

with  $\hbar$  Planck's constant divided by  $2\pi$ ,  $\gamma_e = 1.76 \times 10^{11} \text{s}^{-1} \text{T}^{-1}$  the gyromagnetic ratio of the electron,  $B_0$  the magnetic field,  $k_B$  Boltzmann's constant, and T temperature. The resonance (Larmor) frequency of electron-spin transitions is given by,

$$\omega_0 = \gamma_e B_0 \tag{2.2}$$

Therefore, the associated resonance field for  $f_{osc} = 40$  GHz is  $B_0 = 2 \pi f/\gamma_e = 1.4$  T. Thus, at T = 4.2 K and  $B_0 = 1.4$  T,  $p_e$  becomes ~0.3, within our target polarization range.

The irradiation intensity required to observe signal depends on the magnetic resonance technique used to manipulate the electron spin magnetization and on the efficiency of the coil or waveguide used to generate the resonant oscillating magnetic field. The irradiation intensity is specified in terms of the magnitude  $B_1$  of the irradiation's transverse oscillating magnetic field. To achieve saturation requires a  $B_1$  exceeding,

$$B_{\rm sat} = \frac{1}{\gamma_{\rm e} \sqrt{\tau_1 \tau_2}} \tag{2.3}$$

with  $\tau_1$  the spin-lattice relaxation time and  $\tau_2$  the spin dephasing time. For a nitroxide-radical dopant like the one used in the DNP experiments in [41],  $\tau_1 = 1.3$  ms and  $\tau_2 = 450$  ns at T = 4.2 K and  $B_0 = 0.6$  T; the computed saturation field is  $B_{sat} = 2 \mu$ T. To flip an electron spin using a short pulse of irradiation requires a  $B_1$  exceeding,

$$B_{\text{pulse}} = \frac{1}{\gamma_{\text{e}} T_2} \tag{2.4}$$

which gives  $B_{pulse} = 80 \ \mu\text{T}$  for the representative nitroxide radical and waveguide discussed. In the experiments in [41], a coplanar waveguide (CPW) was used to generate the transverse oscillating magnetic field. We carried out SONNET simulations to obtain a lower-bound estimate of the microwave power (*P*) required to achieve a given  $B_1$  using such a waveguide. Simulating the performance of a 50- $\Omega$  CPW on silicon whose centerline was 5  $\mu$ m wide and 300  $\mu$ m long with an input power of  $P = 300 \ \mu$ W gives a calculated  $B_1 = 54 \ \mu$ T. The calculated coil constant was thus  $c = 3.1 \ \mu\text{T}/\sqrt{\mu\text{W}}$ . Given  $P_{\text{sat/pulse}} = (B_{\text{sat or pulse}}/c)^2$ , the input power required for saturation is  $P_{\text{sat}} = 0.4 \ \mu$ W. The corresponding power for a pulsed magnetic resonance experiment is  $P_{\text{pulse}} = 670 \ \mu$ W. The saturation condition depends on the electron-spin relaxation times, which are not a strong function of the applied magnetic field for the nitroxide radical used in [41]. Thus, the estimated  $P_{\text{sat}}$  is approximately the waveguide input power required for any nitroxide DNP experiment at 4.2 K, carried out in the 20 to 250 GHz frequency range.

With the chip-scale microwave source operating in close proximity to the sample, heat generated by the source becomes a concern. Any power lost in the conversion between the dc

input and the ac output of the chip shows up as heat. Heat transmitted to the sample will lower the spin polarization, based on (2.1), and heat transmitted to the cantilever will increase its thermomechanical motion; both of these changes decrease the MRFM experiment's signal-to-noise ratio. Pulsed magnetic resonance experiments are the most demanding. Given that the CPW is interposed between the IC and the sample and taking into account the losses associated with wirebonds from the IC to CPW, with a DC-to-RF efficiency exceeding 40% at cryogenic temperatures, an output power of  $P_{pulse} = 400 \,\mu\text{W}$  requires around 50 mW of dc input power and generates 40 mW of heat. Even hundreds of milli-watts of heating is still well within the cooling budget of a cryogenic MRFM experiment operating at T = 10 K.

Using a chip-scale microwave source in a magnetic resonance experiment raises the concern of 'bleed-through' while the oscillator is in the off state. In the electron-spin-detection MRFM experiments of [41], bursts of saturating microwaves were applied periodically; the microwave duty cycle – the ratio of the microwave-on to microwave-off times – was kept low to minimize heating. With a typical pulsed-source on/off power ratio of 40, the low-level irradiation present during the "off" period could cause a deleterious saturation of the sample's electron spins. This unwanted saturation can be eliminated by shifting the microwave source's off-resonance operating frequency during the "off" period. For a 40 GHz oscillator, with a tuning range of 2 GHz, this change in irradiation frequency is equivalent to a field sweep of  $\Delta B_0 = 71$  mT. This  $\Delta B_0$  is two to three orders of magnitude larger than the saturation linewidth and therefore more than sufficient to eliminate the possibility of off-state saturation.

Sweeping the magnetic field rapidly and adiabatically through resonance is a particularly versatile and power-efficient way to invert spin magnetization. The expected switching time of the oscillator presented here is on the order of ~nanoseconds. The large and rapid frequency sweep

created by an on-chip oscillator facilitates using adiabatic rapid passages to manipulate electronspin magnetization. To maintain magnetization during the sweep, a  $B_1$  large enough to meet the adiabatic condition,

$$B_1^2 \gg \frac{1}{2 \pi \gamma_{\rm e}} \frac{\Delta B_0}{\Delta t} \tag{2.5}$$

is required. In (2.5),  $\Delta t$  is the field-switching (or frequency-switching) time. To maintain magnetization during a sweep of width  $\Delta B_0 = 71$  mT and duration  $\Delta t = 1 \,\mu s$  requires  $B_1 \gg$ 630  $\mu$ T. This  $B_1$  is comparable to the pulsed-field value. One additional concern is that phase noise could lead to a decay of spin-locked magnetization during the passage through resonance. The decay rate on resonance is estimated by,

$$k \sim (\gamma_{\rm e} B_1)^2 P_{\delta \phi}(\gamma_{\rm e} B_1) \tag{2.6}$$

where  $P_{\delta\phi}(\gamma_e B_1)$  is the oscillator phase noise, evaluated at the frequency offset of  $\omega = \gamma_e B_1$ .

#### 2.2.2 Cryogenic Device Physics Considerations

Electronics that are intended to be used for extreme environments (e.g., very low/high temperatures) must be "hardened" against the detrimental effects thereof. Hardening is accomplished through changing the underlying fabrication flow and/or incorporating circuit techniques to mitigate the unwanted phenomena. SiGe HBT represents a platform that offers a great potential for cryogenic environment electronics [33]. Unlike a silicon bipolar-junction transistor where cooling leads to an exponential drop in the current gain ( $\beta$ ) and a concomitant increase in the base resistance due to carrier freeze-out, SiGe HBT offers a solution that naturally benefits from cooling [33]. Cooling boosts the speed of SiGe HBTs by magnifying the favorable

effects of bandgap engineering. Cooling also reduces the transistor-limiting parasitics such as base resistance and junction capacitances, while greatly improving transit and charging times. SiGe HBT reliability is typically improved at low temperatures. This is in contrast to conventional CMOS that, for instance, suffers from enhanced hot carrier damage at cryogenic temperatures. From a circuit-level perspective, cooling improves almost all of the transistor attributes including current gain, transconductance  $(g_m)$ , transit frequency  $(f_T)$ , maximum oscillation frequency  $(f_{max})$ , and broadband noise. However, unusual phenomena that can be observed in SiGe HBTs at very low temperatures, including higher leakage, barrier effect, negative-differential-resistance (NDR) effect, and collector-current kink in the Gummel characteristics [54] must be considered. These effects should remain second-order if the circuit is designed properly. Since cooling reinforces the desired effects of bandgap engineering, cryogenic operation is typically utilized as a "scaling lever" to gain better insight into the scaling limits in SiGe HBT technology. In this context, cooling is demonstrated to be qualitatively similar to combined vertical and lateral scaling at room temperature with respect to dc, ac, and noise-performance metrics of SiGe HBTs [55]. That is, cooling impacts the terminal characteristics of transistors in the same manner as scaling. In other words, the transistor is virtually scaled by cooling. Unlike scaling, cooling can improve the transistor metrics without affecting the breakdown voltages, i.e., the reduction in  $\tau_{BE}$  (baseemitter transit time) is achieved without aggressive collector scaling.

Transistor models in simulation packages are typically optimized for operations at or near room temperature. These tools rarely include low-temperature models below -50 °C. Thus, they become ineffective for designs at cryogenic temperatures (< -150 °C). In [56], test structures are fabricated and measured to develop a low-temperature model for CMOS transistors. The developed custom-model is then incorporated in to the simulation packages to design circuit

blocks. This approach, in most cases, may be troublesome, time-consuming, and inefficient. Here, we propose using the relationship between cooling and transistor scaling to intelligently design circuits without the need for low-temperature models. In short, SiGe HBTs can be thought of as the enabling technology for cryogenic electronics that can be utilized at room temperature to yield better performance results at low temperatures [33]. For a detailed discussion on the lowtemperature electronics and the equivalency between scaling and cooling, theoretical derivations and reported experimental results have been compared extensively in the Appendix section (A.2). SiGe technology is typically practiced as a BiCMOS platform (SiGe technology exists in a BiCMOS implementation as an adder process module to the core CMOS platform), thus, care must be exercised with respect to the behavior of CMOS transistors within the BiCMOS platform at cryogenic temperatures. The threshold voltage in CMOS transistors rises with cooling, diminishing the achievable current drive and  $g_m$  and degrading the linearity. CMOS transistors also suffer from an enhanced hot-carrier damage at reduced temperatures [33]. Nevertheless, CMOS is well-behaved at low temperatures in other respects [33]. Examples of SiGe HBT circuits operating cryogenically can be found in [57]–[58].

#### 2.2.3 Circuit Design Given the "Hypothetical Ongoing Scaling"

Due to the favorable impacts of cooling, in almost all works where cryogenic circuit blocks are implemented in a SiGe platform, the circuits are designed at room temperature with results expected to improve at low temperatures. In this context, the circuit output is not specifically optimized for a target temperature. In [34], for instance, a low-noise amplifier (LNA) designed at room temperature to achieve a low-power performance exhibits a low-noise performance at cryogenic temperatures. This finding shows the importance of having low-temperature models, if
aiming to optimize a circuit parameter for cryogenic operation. For the purposes of circuit design, the effect of reducing the temperature on a transistor's metrics can be quantified given the relation between cooling and scaling. To do so, we begin by investigating whether scaling from one technology node to the next impacts transistors to the same extent as reducing the temperature from 300 to 4 K.

A literature review suggests that cooling more strongly impacts the transistors than scaling [60]. A well-balanced increase of about 60% for both  $f_T$  and  $f_{max}$  at 48 K is reported in [60]. The increase in  $f_T$  is almost uniform for transistors with different vertical profiles and/or lateral layout while devices with thin base layers and high collector doping level exhibit a moderate increase in  $f_{max}$ . To improve  $f_{max}$ , in addition to  $L_E$ , the emitter width ( $W_E$ ) of the transistor must also be scaled to reduce collector-base capacitance and base resistance as per (lateral scaling is impactful on  $f_{max}$  and vertical scaling on  $f_T$ )

$$f_{max} = \sqrt{\frac{f_T}{8\pi R_B C_{BC}}} = \frac{1}{4\pi \sqrt{\tau_{EC} \tau_{BC}^*}}$$
(2.7)

where  $R_B$  is the base resistance,  $C_{BC}$  is the base-collector capacitance,  $\tau_{EC}$  is the emitter-tocollector delay, and  $\tau_{BC}^* = R_B C_{BC}$ . The  $f_T$  and  $f_{max}$  parameters are improved by ~40% in [61] due to technology scaling from 0.13-µm BiCMOS9MW to 55-nm BiCMOS055. In the BiCMOS055 process, high-speed SiGe HBTs exhibit a superior performance as a result of vertical and lateral scaling and an improved architecture.

In each technology, the peak  $f_T$  and  $f_{max}$  are achieved if the transistors are biased with optimum collector current density or have the proper corresponding size (In the SiGe HBT case, this size corresponds to the emitter length (L<sub>E</sub>) of the transistors) [55], [61].

19

Fig. 2.1 presents the same trend as that of seen in the prior works for the current gain cut-off frequency versus the collector current for the available processes. In this figure, the transit frequency ( $f_T$ ) augmentation is achieved with a lower collector current for the scaled SiGe HBTs. In addition to  $f_T$ , to engineer the transistor to yield a high level of performance at high frequencies, the optimum collector current for the peak  $f_{max}$  or U (unilateral power gain or Mason invariant) is required [63].



Figure 2.1  $f_T$  versus the collector current for a SiGe HBT with L<sub>E</sub> = 3  $\mu$ m



Figure 2.2 f<sub>max</sub> variations as a function of transistor size.

Fig. 2.2 plots the  $f_{max}$  as a function of the size of SiGe heterojunction bipolar transistors. As indicated, arbitrarily small emitter lengths are not capable to achieve the peak  $f_{max}$ . This sets a minimum on the emitter length for a SiGe HBT to fully benefit from scaling and consequently cooling. To elaborate on this; the effect that cooling has on a SiGe HBT performance level is equivalent to that of scaling if that specific same level of performance can be achieved due to scaling, given a certain size of a transistor. For instance, with a too small-size transistor, e.g.,  $L_E < 3 \mu m$ , the "peak"  $f_T$  and  $f_{max}$  values are not resulted whatsoever in a process. This implies that cooling and scaling are no longer similar in, at least, some performance respects in this region.

In Fig. 2.2, the shaded area ( $L_E > 3 \mu m$ ) marks the region where cooling and scaling can be potentially similar/equivalent. This splits the design space into two sub-regions in terms of the transistor performance; the equivalency region (shaded area on the plot) and the semi-equivalency area. By way of explanation, the scaling-cooling equivalency has been observed for the cases where we can fully gain from scaling benefits. In this context, more or less, cooling makes the transistor undergo a continuous "virtual" scaling while in operation toward lower temperatures. The superior high-speed performance in BiCMOS055 process is resulted by vertical and lateral scaling together with an improved transistor architecture. The devices provided by both processes enjoy from a double-polysilicon fully self-aligned (FSA) structure using selective epitaxial growth (SEG) of the SiGe:C base and a selective implanted collector (SIC) to achieve a balanced set of  $f_T$ and  $f_{max}$ , as mentioned earlier [55].

Figure 2.3 illustrates the ac transconductance versus the emitter-base voltage for several SiGe HBTs with various emitter lengths ( $L_E$ ). The transconductance curves exhibit smoother peaks and lower sensitivity for small emitter lengths. That is, the transconductance level is less prone to changes with variations in  $V_{BE}$  which is a helpful fact for a cryogenic design where  $V_{BE}$  needs to

be adjusted due to cooling. This significantly assists with having a more predictable loop gain for oscillation build-up at cryogenic temperatures.

According to this plot, for instance,  $V_{BE}$  must be reduced by almost a volt to accommodate the same level of peak  $g_m$  while the device is being cooled for a large  $L_E$ . This is opposite to the fact that  $V_{BE}$  needs to increase at low temperatures to supply the same I<sub>C</sub>. Thus, the smallest transistor size which happens to be in the equivalency region is highly desirable. This SiGe HBT is expected to yield a lower sensitivity with different  $V_{BE}$  levels.



Figure 2.3 The transconductance sensitivity as a function of emitter length versus base-emitter voltage.

In the simulations, the transistor layouts are changed from a CBE to a CBEBC structure to reduce base and collector resistances for a more balanced  $f_T$  and  $f_{max}$  numbers [55]. This would make the results of the simulations to be increasingly similar to that of resulted by the cooling effect where an increased, balanced values of  $f_T$  and  $f_{max}$  are resulted at reduced temperatures.

Essentially, higher VBE is needed at cryogenic temperatures to bias the transistor having the same level of collector current density at room temperatures [55]. Figure 2.4 displays the forward

Gummel characteristics for the employed SiGe HBT for the available technology nodes, 0.13-µm BiCMOS9MW to 55-nm BiCMOS055 from STMicroelectronics. As observed, the drive-current capability of SiGe HBTs are enhanced as the transistor scales. In other words, scaled HBTs supply an increased collector current with a constant value of VBE which is in agreement with the fact that scaled devices function with increased operation current densities [64]. This, however, is in sharp contrast with the behavior seen in Gummel characteristics in cooled SiGe HBTs [55].



Figure 2.4 Gummel characteristics for the employed HBTs from BiCMOS9MW and BiCMOS055 processes. Figure 2.5 conceptually compares the resulting Gummel characteristic plots due to scaling and cooling. In this figure, the solid line represents the transistor under normal conditions. As observed, if a SiGe HBT is cooled, there is a necessity to increase  $V_{BE}$  to supply the same level of collector current density as in the room temperature. As a result of cooling, the J<sub>C</sub> plot shifts to the right and this is translated to a higher  $V_{BE}$  for a fixed J<sub>C</sub>.

On the contrary, when a SiGe HBT device is scaled, its current-generating capability improves, and the device can afford a higher  $J_C$  with the same  $V_{BE}$ . This recommends designing the transistors

with a low level of  $V_{BE}$ . This leaves the transistor with an adequate  $V_{BC}$  range even in case that the base voltage is required to increase to secure the corresponding fixed current density level at room temperature. This has to be accounted for with respect to the biasing circuitry at circuit design level. Nonetheless, even with the opposite current trend in biasing, the current gain ( $\beta$ ) still increases with scaling similar to that of a cooled operation [55].



Figure 2.5 Change in the trend of Gummel characteristics due to (a) scaling (b) cooling with respect to room-temperature condition.

In this approach, scaling mimics the effect of cooling down to ~77 K while we aim for an operation near ~45 K (which is the operating temperature close to the IC). In this respect, we investigate whether a  $\Delta T$  reduction in temperature below 77 K would affect the transistor metrics to the same extent as a similar temperature change would do at higher temperatures (i.e., going from 0 to -50 °C). The details has been provided in the Appendix section (A.3). In short, setting the simulation temperature to be at -45 °C in simulations, we can confidently predict and simulate the performance parameters, i.e., the circuit block operation, up down to 55 K. In these plots, the transistor layout has been changed from a CBE to a CBEBC structure to reduce base and collector resistance for a balanced  $f_T$  and  $f_{max}$ .

### 2.2.4 An Efficient High-Power Fundamental Oscillator

Efficiency is the metric that measures how much input power in a power amplifier, including the input ac ( $P_{in}$ ) and power supply dc ( $P_{dc}$ ), converts into desired output power. In drain efficiency, the definition drops the input power needed to drive the amplifier and is the ratio between the output power and dc input power. An amplifier can have a high drain efficiency while its genuine "efficiency" is low due to a relatively low power gain ( $P_G$ ). In those cases, drain efficiency becomes a misleading indicator and power-added efficiency is rather considered to present a device's potentiality for efficient power conversion and delivery. In this context, PAE is a measure that shows a transistor talent for efficient operation when it is stable as, at very high frequencies, a transistor's power gain significantly falls with frequency. As such, even if the drain efficiency is high the amount of that is needed to derive the input becomes considerable. This is more crucial for bipolar transistors where the input power can be significantly higher compared to CMOS transistors. This all suggests PAE as the most relevant metric if one is to find the efficiency limits of the transistors.

To design a high-efficiency power amplifier, therefore, the first step is to find the maximum achievable PAE (PAE<sub>max</sub>) for transistors given an available process at hand. For a single-transistor stage, the maximum achievable PAE is resulted if high-order passive networks consisting of a large number of ideal energy-storing elements and bias nodes are utilized, similar to the concepts in class-F power amplifiers. In this setting, a large number of passives facilitate harmonics utilization, i.e., harmonic wave-shaping, to enhance efficiency. The next phase is to replicate this ideal condition for transistors in power amplifier structures using the actual lossy components, approaching the actual efficiency limits of the device. The same idea can be effectively employed

to design high-efficiency oscillators with some additional considerations. First, the loop gain for start-up condition must also provide by the design and secondly, one should ensure that  $P_{in}$  is greater than zero. The second condition implies that the transistor is internally stable in the proposed oscillator topology. In [63], for instance, the authors propose a method to internally destabilize the transistor by putting a capacitor at the source and increase the unilateral power gain. Obtaining the maximum achievable PAE for the available 0.13-µm BiCMOS process, 80%, we notice a significant difference between the state-of-the-art efficiency numbers reported so far on oscillators (16% DC-to-RF efficiency at 40 GHz) with the achieved number. This is despite that there are decent works on power amplifiers with PAE = 45% at 30 GHz on CMOS. This well exhibits the gap in oscillators and implies that the full capacity of the transistors in prior designs are not effectively exploited overall. With the setting shown in Fig. 2.6, we find the maximum achievable PAE for a transistor size of  $L_E = 6 \mu m$ .



Figure 2.6 The employed setting to achieve maximum PAE

For the employed SiGe HBT, the base bias voltage ( $V_B$ ) and the voltage swing on the base ( $S_B$ ) primarily controls  $PAE_{max}$  as they have key roles in power maximization and power consumption. Based on the simulations, the maximum PAE of 70% is obtained for the employed transistor with 700 mV voltage swing ( $S_B$ ) on the base and  $V_B = 0.3 V$ . With a typical conventional cross-coupled structure as shown in Fig. 2.7, these conditions, however, cannot lead to oscillation and based on the calculations a  $V_B = 0.75 V$  (independent of the voltage amplitude) is necessary to provide the required g<sub>m</sub>.



Figure 2.7 Conventional cross-coupled structure

This bias voltage also gives rise to a voltage swing which is not close to the ideal swing (i.e., 700 mV). In bipolar transistors, if  $P_{in} > 0$ , the base draws current when the transistor is on. This gives a periodic waveform for the base current. The pictorial representation of the resulted base current i<sub>B</sub> is presented in Fig. 2.8. In this waveform, the dc value, I<sub>ave</sub>, is a function of both  $V_B$  and  $S_B$ . The figure shows that in a bipolar-based oscillator topology, I<sub>ave</sub> can increase as the oscillation builds up. As a result, the voltage bias on the base is lowered in the steady state as shown in Fig. 2.9.



Figure 2.8 Pictorial representation of the base current in time.



Figure 2.9 Base-voltage waveform evolution in time

In this figure,  $V_{Bi}$  represents the initial state where the oscillation waveform forms its initial cycles and  $V_{Bf}$  represents the base voltage in the steady-state. Von represents the on-voltage of the transistors. With larger voltage swing on the base, the transistor draws more current and, as a result, the base voltage average decreases. Such an operation type can inherently bring the transistor closer to the ideal condition. In this situation,  $I_{ave}$  can be provided through biasing the base using a resistor,  $R_B$ , as shown in Fig. 2.10. Knowing the ideal conditions for the start-up and the maximum PAE, the initial base bias voltage ( $V_{Bi}$ ) must be 0.75 V and the final bias voltage ( $V_{Bf}$ ) and  $S_{Bf}$  in the steady-state mode must be equal to 0.3 V and 700 mV, respectively. As such, we modify the transistor embeddings and manipulate the decoupling capacitor ( $C_d$ ),  $R_B$ ,  $V_B$ , and the transistor size to set the above values for when the signals are small and when they are large. In this circuit, number of degrees of freedom are greater than the required unknowns ( $V_{Bf}$  and  $V_{Bi}$ ). As at PAE<sub>max</sub>, the swing on the drain is maximum,  $V_B$ ,  $C_d$  and  $C_{\pi}$  set the desired  $S_B$ . Given that,

$$V_{Bf} = V_{Bi} - R_B \times i_B|_{dc} = V_{Bi} - R_B \times I_{ave}$$

$$\tag{2.8}$$

The biasing circuitry must employ a resistive load to sustain the dc drop across. In this context, an RF choke (RFC), cannot be employed to yields the maximum efficient condition. If the ideal  $V_B$  and  $S_B$  happen to be close to the initial swings and bias voltage, then RFC can be a possible solution

for biasing. Figure 2.11 displays the simulated base-voltage waveform as it evolves in time which confirms the above conclusion.



Figure 2.10 The equivalent transistor embedding to achieve maximum PAE in an oscillator.



Figure 2.11 A simulated sample of base-voltage waveform evolution in time

If the transistor gain is not large enough to generate the required swing levels at the steady state, the optimum  $S_B$  for maximum PAE cannot be achieved. Figure 2.12 compares two cases where in (a) the transistor gain is sufficiently high that it gives swing levels close to the ideal condition with that in (b) where the oscillation level cannot reach that of the optimum condition. To supply the required gain, a tail current source is designed to yield the needed  $g_m$ . Transistor size can also vary while it is preferred to shrink it for maximum frequency tuning. In this structure, the tail current source consumes no more than 0.3 mW.



Figure 2.12 The oscillation build-up in time (a) when the transistor's gain is not sufficient (b) when the oscillation swings reach the required  $S_{Bf}$ .

The input-drive condition of a transistor governs the intensity of the NDR-induced instabilities at cryogenic temperatures [54]. A transistor can be driven by an ideal current source (forced- $I_B$ ) or an ideal voltage source (forced- $V_{BE}$ ) at the two extreme input-drive cases. Equivalently, assuming a series resistance ( $R_{B,series}$ ) between the transistor base and the voltage source (Fig. 2.13),  $R_{B,series} = 0$  represents the constant-voltage mode and  $R_{B,series} = \infty$  typifies the constantcurrent condition. In practical cases, circuits operate in an intermediate state. The operative NDR in a forced- $I_B$  derive is reduced as each generation of a technology scales the device. In contrast, the constant-voltage input NDR is increasingly worsened with scaling [54]. Thus, to avoid the NDR effect, the transistor must be driven to operate in a forced- $I_B$  mode. From a circuit-design perspective, the output impedance of the preceding stage or the biasing circuitry should be incorporated in the design to assess the NDR intensity. In Fig. 2.14, with a high  $V_{CE}$ , a voltage divider is designed for biasing the base that gives  $R_{B,series} = R_{eq} = R_1 \parallel R_2$ . Choosing large values of  $R_1$  and  $R_2$  (on the order of K $\Omega$ ) increases  $R_{B,series}$ . Biasing the transistors with a small  $V_{BE}$  can alleviate the concerns of NDR entirely [54].



Figure 2.13 The equivalent series resistance at the base to determine the input derive conditions.



### Figure 2.14 Forced- $I_B$ mode implementation to avoid NDR.

Passive circuit components require especial attention at low temperatures. In particular, capacitive changes are dominant at reduced temperatures while changes in dielectric constant are not remarkable. Cooling positively impacts the conductivity of metallic interconnects by lowering the inherent resistivity. The behavior of semiconductor-based resistors deviates from that at 300 K, depending on the type. Polysilicon-based resistors show no significant change due to cooling [56], unlike N-well resistors where changes are large.

### 2.3 Implementation and Circuit Design

The proposed 40-GHz fundamental VCO is implemented in a 0.13-µm BiCMOS process while being simulated in a 55-nm BiCMOS process. Figure 2.15 displays the schematic of the proposed VCO. The core oscillatory structure is followed by two amplifiers at the outputs and utilizes a tail current source to generate the required swing level and improve the phase noise of the circuit. The tail current source (I<sub>T</sub>) is implemented using CMOS transistors where they operate at the edge of saturation over the entire oscillation period, consuming the minimal power ( $\leq 0.3$  mW). The core transistors will have L<sub>E</sub> = 6 µm in Nx-(CBEB)-C configuration to increase  $f_{max}$ . In Fig. 2.10,  $C_d$ ,  $C_a$ ,  $C_{var}$ , and size of the transistor are co-designed to maximize the tunability for the oscillator while not degrading the efficiency significantly. This imposes an extra constraint for efficiency enhancement such that  $C_d$  and  $C_a$  are shifted form the theoretical optimum values. With  $C_{max}/C_{min} = 2.8$ , the quality factors of the varactors do not exceed 6, so  $C_{var}$  is set to be 120 fF.

At the amplifier stage, the transistors are sized to take on a larger value,  $L_E = 8 \mu m$ , and the power amplifier is designed to yield the maximum PAE. Simulations shows a PAE of more than 40% for the output stage. The inductive loads required for the core tank and the outputs are implemented as CPW transmission-lines (TLs) with ground shielding. The lengths and dimensions of the TLs are simulated using an electromagnetic (EM) solver to maximize Q. NMOS current mirrors are used to bias the tail current source. To account for the parasitics contributed by the RF pad, this component is also simulated in HFSS to obtain an equivalent capacitance to the substrate (in this case, 30 fF at 40 GHz). The inductance of traces and other interconnects are also carefully modeled and included in the simulations. Remarks: most of the design variables have been optimized for the low-temperature performance. However, the large number of bias points provided by the base voltage and the current source (I<sub>T</sub>) permits to tune the performance such that it is boosted at room temperature as well.



Figure 2.15 The Schematic view of the proposed cryogenic efficient VCO.

#### 2.4 **Prototype Measurement and Results**

The implemented fundamental VCO is tested in chip-on-board configuration through on-chip probing at room temperature. The frequency and power measurement setup is shown in Fig. 2.16. For the measurement of the oscillation frequency ( $f_{osc}$ ) and the phase noise (PN), we have used 8565E Agilent 50-GHz spectrum analyzer (SA). The output of the chip is connected to the SA using low-loss microwave cables through a GSGSG 60-GHz Cascade Infinity probe with a power loss of ~1 dBm. The same setting is utilized to measure the output power. The power levels in this plot are those after de-embedding the loss from the cables and the probe used in the measurement. Table 2.1 summarizes the measured performance results at room temperature.

The low-temperature measurements are done using a vacuum electrical characterization probe station. Figure 2.17 presents the measurement setup along with the annotated chip micrograph of the cryogenic VCO.



Fig. 2.16 The measurement setup for the proposed VCO at room temperature

reference	process	Freq. (GHz)	TR (%)	Eff (%)	Pout (mW)	P <sub>DC</sub> (mW)
[65]	45-nm CMOS SOI	45.7	4.3	15.7	6.62	42.2
[66]	90-nm CMOS	46.2	10.4	13.6	11.5	84.4
[67]	90-nm CMOS	30	20	4.6	1	21.7
[68]	0.12-µm BiCMOS	22.4	5.5	1	0.28	27
This work	0.13-μm BiCMOS	40.4	11.2	25.5	14	55

### Table 2.1 Comparison with the prior art

The PCB-mounted chip is heat sunk to the probe-station's copper cold finger in vacuum and the VCO is coupled through aluminum wire bonds to a lithographically defined coplanar waveguide first to verify if the fabricated VCO features the desired expected performance for the experiment. In this setting, the CPW is tapered to a micro wire in the center [41]. The narrow micro wire concentrates the microwaves delivered in an MRFM experiment to generate a micro teslastrength  $B_1$  field. The VCO's inverting output is left unconnected. The output of the CPW is delivered to a 50-GHz spectrum analyzer (Agilent 8565E) using a 60-GHz Cascade GSG Infinity Probe and low-loss coaxial cables. The chip's temperature is measured using a sensor located on

the underside of the probe station's cold finger and is regulated using a feedback controller. Each temperature set point is reached by adjusting a heater and controlling the flow of liquid helium beneath the probe's cold finger. At each temperature setup point, the system is allowed to settle for >15 minutes, permitting the copper cold plate and the PCB/VCO/CPW setup to equilibrate.



Figure 2.17 The cryogenic measurement setup and the chip micrograph of the VCO.

Fig. 2.18 shows the measured VCO's oscillation frequency versus the control voltage ( $V_{tune}$ ) at several temperatures **at the output of the CPW**. The VCO consumes almost 55.8 mW of dc power and has a tuning range of more than 5.4% at low temperatures. The measured data shows a slight shift in the frequency of operation moving from 300 K to 12 K. The reduction in the capacitance of MIM capacitors is mainly responsible for this frequency shift (the inductance's temperature coefficient is expected to be on the order of ppm/°C [69]). The electron spin resonance field corresponding to the VCO's center frequency at room temperature is 1.29 T; the associated proton nuclear magnetic resonance frequency at this field is 54.7 MHz. The relative 5.4% tuning range of the proposed VCO at all temperatures corresponds to a field-sweep range of 70 mT, more than sufficient for use in an electron spin resonance-MRFM experiment.



Figure 2.18  $f_{osc}$  vs. control voltage at various temperatures.

Fig. 2.19 plots the measured DC-to-RF efficiency versus  $V_{tune}$ . This data was collected directly from the output of the CPW; no calibration techniques are utilized to de-embed the effect of loss. From a separate measurement of the transmission losses through the CPW and SMA cables used in the measurements, the setup is conservatively estimated to have an attenuation of ~14 dB (~25×) compared to a direct measurement of the VCO, without the losses from the wire-bonds.



Figure 2.19 Measured DC-to-RF efficiency at the CPW output vs.  $V_{BE}$ .

The measurements also show that output power exceeds 300  $\mu$ W at the output of the CPW. Based on SONNET simulations this power should produce  $B_1 \sim 54 \mu$ T of transverse oscillating magnetic field. This  $B_1$  is large enough for pulsed experiments and more than sufficient for electron spin resonance and DNP experiments. Replacing the 1 mm-long wire bonds with a low-temperature flip-chip connection could improve the  $B_1$  by an order of magnitude.

The commonly used X5R and X7R on-board capacitors and potentiometer used on the board are known to exhibit large variations at extremely low temperatures (up to a 90% change from the nominal value [56]), leading to significant changes in the bias currents. In this context, the bias points should be manually adjusted to compensate for such changes. Low-temperature capacitors such as NP0/COG also show large variations at cryogenic temperatures. These capacitors are unable to filter the supply noise near 4 K [56]. Figure 2.20 shows the measured phase noise (PN) near 36 GHz.



Figure 2.20 Measured phase noise vs.  $f_{offset}$  for various temperatures.

The higher PN at low offsets, down to mid-range cryogenic temperatures (i.e., 77 K), confirms that the on-board capacitors cannot filter the supply noise. In contrast, PN at lower temperatures 37

behaves as expected, resulting from an additional increase in the bias current. For a field strength of  $B_1 = 1 \text{ mT}$ , the relevant offset frequency is  $\gamma_e B_1 = 28 \text{ MHz}$ . For a phase noise at this offset of -100 dBc/Hz,  $P_{\delta\phi} = 1.0 \times 10^{-10} \text{ Hz}^{-1}$  and the electron-magnetization decay time is  $\tau = 1/k = 13 \ \mu s$ . At high offset and low temperature, the phase noise approaches -120 dBc/Hz and  $\tau$  becomes 1.3 ms which avoids magnetization loss from phase noise during an adiabatic sweep.

To verify the accuracy of the proposed cooling-scaling-mapping based approach, we also measured the performance of the chip without the presence of the CPW at low temperatures utilizing the exact same measurement setting. The cryogenic measurement results of the fabricated VCO is presented in table 2.2.

Temp. (K)	Freq. (GHz)	TR (%)	Eff (%)	P <sub>out</sub> (mW)	$P_{DC}(mW)$
77	37.7	6.2	40	22.1	55.2
50	37.8	6.1	44.9	24.9	55.5
20	37.6	6.1	51	28.4	55.8
12	37.4	6.15	50.2	28	55.8

Table 2.2 the performance summary of the fabricated cryogenic VCO at multiple temperatures

Simulations show a peak DC-to-RF efficiency of 34% at 77 K and 36% at 55 K which is slightly smaller than the measured values. However, the accuracy of the proposed design methodology is good enough for the purpose of circuit design, given no low-temperature transistor model is available. The larger-than-expected estimated efficiency numbers can be attributed to the enhancement in the tank's Q. At low temperatures, magnetically induced eddy currents from the

tank barely flow in the high-resistive substrate because of carrier freeze-out. We postulate that the freeze-out reduces the substrate loss and prevents the degradation of the Q.

### 2.5 Conclusions

In this work, we develop a semi-quantitative approach for low-temperature circuit design. The proposed approach is capable to acceptably predict the circuit performance at low temperatures without resorting to low-temperature transistor models. Once the low-temperature design procedure is established, we utilize circuit techniques to boost the efficiency of the oscillator. The proposed VCO employs SiGe HBTs to operate cryogenically at frequencies up to tens of gigahertz. The result is a 40-GHz efficient cryogenic VCO, implemented in a 0.13- $\mu$ m BiCMOS process. With 300  $\mu$ W output power, 5.4% tuning range, and 25% DC-to-RF efficiency at the output of the CPW in the experiment setting, the VCO satisfies the requirements for exciting electron spin resonance and initiating dynamic nuclear polarization in an MRFM experiment.

# Chapter 3: Efficient Signal Generation Close-to-fmax

### 3.1 Introduction

The demand for signal generation at mm-wave frequencies has demonstrated an upward trend in the recent years. This increasing demand can be attributed to the numerous applications that are being implemented in the mm-wave frequency band including high data-rate wireless communication in 60-GHz band, automotive radar at 77-GHz band, and passive radar imaging in W-band (75-110 GHz) [70]. Millimeter-wave sources, especially those that can operate at the frequencies in W-band and above are needed for advanced imaging, remote sensing and radar applications. W-band radiation, in particular, offers an interesting balance between spatial resolution, material penetration, and atmospheric absorption for these applications. Depending on the application and the required resolutions the bandwidth can be different. However, output power and efficiency are the critical parameters at such high frequencies.

On-chip power generation at frequencies close to or beyond  $f_{\text{max}}$  is not a trivial task to accomplish. The limited  $f_{\text{max}}$  of today's silicon CMOS technology at mm-wave frequencies, rules out the choice of generating the high-frequency signal independently and then utilizing a power amplifier (PA) to efficiently boost the close-to- $f_{\text{max}}$  generated signal's power. This calls for alternative power-oscillator design approaches that are capable of enhancing the output power of generated signals. In such schemes, the oscillator is typically the "power-hungry" circuit block of the system, requiring its efficiency to be improved. As an alternative, the signal can be generated at 45 GHz, then use a 45 GHz PAs followed by bias-optimized frequency doublers to achieve the target operating frequency. This is practically challenging. The frequency multipliers normally have both high output power and bandwidth. However, these multipliers suffer from large power

conversion loss. Hence, they need power hungry power amplifiers to boost the power level. This increases the power consumption and silicon area and largely deteriorates the efficiency that is the main concern of the proposed approach. In terms of phase noise, there are two sources of noises, the noise from the passives and the noise from the active devices. It is practical to achieve a better quality-factor at 45 GHz, but it is not significantly larger, based on the simulations it differs by a factor of about 12%. In terms of noise performance of the device itself, the noise performance is better at 45 GHz in general. However, we need to consider that the noise of the doubler needs to be considered for the overall phase noise performance. Specially since the doubler is a non-linear component which folds the noise around the harmonics back to the fundamental and can increase the noise level and hence the overall PN. This is especially important for phase noise at higher offset frequencies which might limit the phase noise to the noise level of the doubler. We should also note that the doubler is usually optimized for the lowest conversion loss, hence it is not optimized for noise performance and can have a high noise level. For a VCO at 90 GHz, we do not need to deal with the noise of the frequency doubler which makes it easier to manage the noise contribution of the devices to the phase noise.

Several studies have been carried out to address the design of high-efficiency oscillators at mmwave frequencies [71]–[75]. The voltage-controlled oscillator (VCO) in [71] utilizes a band switching technique and controls the effective gate inductance to increase the tuning range. The resulted VCO demonstrates a peak DC-to-RF efficiency of 3.7% as a by-product. In this design, the output power drops as the bias changes to tune the frequency. In [72], a varactor-less technique is employed that tunes the circuit with minimal degradation of the generated output power. In [73], a more inclusive approach is adopted to systematically maximize the available power of a single transistor regardless of the circuit topology and in [74] the method is improved by standardizing the methodology of arriving at the passive network for a multistage ring oscillator. Despite being the common approach in most of the recent works, this method does not maximize the efficiency as it discards the DC power consumption and only focuses on maximizing the generated small-signal power. In [75] a modified differential Colpitts oscillator employs larger transistors to enhance linearity and then exploits the improved linearity to achieve a DC-to-RF efficiency of 4.9%. Yet, the differential output of the VCO in this design needs a supplementary lossy network for power combining.

In this work, we will primarily focus on maximizing the efficiency of oscillators. Inspired by switching-type PAs, we propose a high-efficiency oscillator at mm-wave frequencies. The result is a fundamental oscillator operating at a center frequency of 91 GHz and achieving a state-of-the-art DC-to-RF efficiency of 6.1% while generating a maximum measured output power of 4.5 dBm. The proposed oscillator attains the phase noise figure of merit (FoM) of -169.6 dBc/Hz at 1-MHz offset frequency.

### **3.2** A High-Efficiency Class-E-Like Oscillator Topology

### 3.2.1 Drain Efficiency in A Single-Transistor Power Amplifier

The drain efficiency of a single-transistor power amplifier ( $\eta$ ) is defined as:

$$\eta = \frac{P_{L-f0}}{P_{DC}} = \frac{P_{L-f0}}{P_{dissipated-transistor} + P_{disspated-passives} + P_L}$$

$$\approx \frac{P_{L-f0}}{P_{dissipated-transistor} + P_{disspated-passives} + P_{L-f0}}$$
(3.1)

where  $P_{L-f0}$  denotes the fundamental output power,  $P_L$  is the overall power delivered to the load, and  $P_{DC}$  is the average DC power from the supply voltage. Assuming that the output power of harmonics is much lower than the fundamental power, the latter approximation holds. If  $P_{dissipated-passives}$  is negligible, and  $P_{dissipated-transistor}$  substituted by the corresponding instantaneous currents and voltages,

$$\eta \approx \frac{P_{L-f0}}{P_{dissipated-transistor} + P_{L-f0}} = \frac{P_{L-f0}}{\frac{1}{T} \int_{0}^{T} I_{D}(t) V_{D}(t) dt + P_{L-f0}}$$
(3.2)

where  $I_D(t)$  and  $V_D(t)$  are the transistor drain current and voltage waveforms respectively. To reduce power consumption, ideally, it is desired to shape  $I_D$  and  $V_D$  waveforms to minimize the time during which the transistor simultaneously carries a large current and sustains a large voltage. This basic approach has been extensively investigated in the context of switching amplifiers. Class-E stages are nonlinear switching PAs that achieve high efficiency while delivering full power [76]. Nonetheless, switching PAs in CMOS still need to be explored at mm-wave frequencies. The lack of capability to shape a sharp pulse-like drain current at mm-wave frequencies and beyond limits utilization of the switching concepts to enhance efficiency. This is essentially due to the presence of parasitics which prohibit hard switching. Besides, the tightly coupled nature of input and output makes harmonic shaping of voltages and currents challenging. Thus, at mm-wave frequencies, one can merely borrow the class-E principals to practically implement a "switch-like" PA [77]. To maximize the ratio in (3.2), one technique is to lower the DC component of  $P_{dissipated-transistor}$  while keeping  $P_{L-f 0}$  constant. This can be achieved, if the transistor is turned off for a certain period of time while its drain current waveform shape and amplitude is preserved. The low-frequency oscillator works that employ similar waveform shaping approaches for improving aspects of oscillator performance have been presented in [78]–[80].

To explain the idea, an idealized representation of the transistor output current is shown in Fig. 3.1. When the transistor is on for nearly the entire period,  $I_D$  waveform ideally resembles a sine-shape wave, Fig. 3.1(a). If the gate-bias voltage ( $V_G$ ) is lowered below  $V_{th}$ , the transistor is turned off for a portion of the time and the corresponding  $I_D$  waveform becomes as shown in Fig. 3.1(b). In this figure, P marks the point where the transistor is turned off or equivalently the position that  $I_D$  waveform departs its ideal sine shape. Location of P is a function of  $V_G$  and specifies the initial charges on the parasitics once the transistor is powered off.



Figure 3.1 Conceptual representation of the possible I<sub>D</sub> waveforms for various bias levels. (a) The ideal case; (b) When the transistor turns off; (c) Closer to ideal case (d) Resulted waveform beyond the optimum case.

While being off, the transistor presents a high-impedance path to the rest of the circuit and as a result, an increased portion of  $I_D$  must be discharged through the path consisting of  $C_{GS}$  and  $C_{GD}$ , shown by Path-I in Fig. 3.2, the current would discharge to ground with a slope (S) proportional to the time constant of the parasitic network ( $\tau$ ). If  $V_G$  is reduced while the drain voltage remains

almost unchanged (and consequently  $C_{DB}$  stays the same), the location of P is changed. If  $\tau$  is also modified, current can be quickly discharged as if the  $I_D$  trajectory resumes its original sine-shape waveform from the point (P) that it had initially stopped when the transistor was turned off. Subsequently, the waveform in Fig. 3.1(c) is resulted where it is notably similar to its ideal counterpart in Fig. 3.1(a), but with a lower DC value ( $V_{G3} < V_{G1}$ ).



Figure 3.2 Simplified model of the mm-wave transistor in the circuit when the transistor is turned off.

This ensures the same level of generated fundamental component of the drain current  $(i_D)$  with a lower DC value of  $I_D$ . Hereafter, if  $V_G$  is reduced any further, the transistor is kept off for a longer time interval and  $i_D$  is significantly attenuated, giving rise to the waveform in Fig. 3.1(d). Likewise, any further increase beyond the optimum value of  $V_G$  does not enhance the overall efficiency as it mostly contributes to the growth of DC component, similar to the case in Fig. 3.1(a). The simulated  $i_D$  (fundamental component) and DC component of  $I_D$  for the NMOS transistor (W/L =  $15\mu$ m/0.13 $\mu$ m) used in the oscillator are plotted in Fig. 3.3 for various levels of  $V_G$ . As seen, the DC component of the current consumption is reduced by approximately a factor of 1.7 for  $V_G = 0.8$  V compared to the case where the transistor remains on for nearly the entire period ( $V_G$ = 1.2 V). At the same time, the generated fundamental component reaches its peak at  $V_G = 0.8$  V.



**Figure 3.3 The fundamental component (i**<sub>D</sub>) and DC component of I<sub>D</sub> as a function of the gate-bias level (V<sub>G</sub>). The waveforms in Fig. 3.1 may seem not similar to the class-E like waveforms of reference [76] where a concerted effort has been made to ensure that the switch current and voltage waveforms have zero overlap (The residual overlap is due to finite switch loss at mm-wave). As such the relationship of the Fig. 3.1 waveforms, and the final VCO topology might seem to be unclear as to the class-E or class-E like PAs. However, In [77], the fundamental limits on achievable performance in stacked CMOS Class-E-like PAs are analyzed and presented in Fig. 2 of the paper. Quoted from this paper, "In practical implementations, the main non-ideality that causes deviation from these limits is soft switching of the stacked devices due to the lack of square wave drives".

Therefore, soft switching debilitates the transistor to generate pulse–shape output currents required to realize an improved efficiency. In contrast, hard switching would make the generation of the pulse-like currents possible. In the absence of square wave drives, we can modify the output current waveforms to resemble that of an ideal sharp pulse wave achieved by hard switching. In real cases, the output current of the transistor at mm-wave frequencies does not look like what presented in Fig. 3.1(a) which is an ideal sine-shape wave. In contrast, it is conceptually very

similar to the case presented by Fig. 3.1(d) mainly due to the parasitics. This exaggerated representation of the output current significantly departs from the sharp pulse-like ideal current waveform. Figure 3.1 pictorially represents an example where a sine-shape output is considered as the ideal output, outlining the overall concept and suggesting how a deteriorated output waveform can be modified back to approach the ideal sine-shape case in the presence of extreme non-idealities. Once this is established, we can further manipulate other effective parameters to go toward creating an ideal pulse-like waveform. This is doable by placing a capacitor at the source of the transistor, as discussed subsequently. The first step to ensure ideal zero overlap condition is to favourably shape the output current waveform and next form the output voltages. The first part that is the idealization of current signals takes place in this work.

In addition, it appears that the voltage and current sinusoids in Fig. 3.1 still overlap. By changing the gate bias, the duty-cycle is altered so that the transistor is off for a finite time. This may seem closer to a duty-cycle reduction in linear PAs like class AB, B and C. This, however, is inevitable as the oscillation frequency is very high. This mainly stems from the tightly coupled nature of the input-output and the non-idealities outlined earlier. However, the overlap is reduced by shaping the output current waveforms. Having a more pulse-like sinusoid waveform, the current reduces its overlap with the voltage. In PAs that are more linear (such as class-AB, B, and C) the transistor is to remain a current source. Then the minimum required drain voltage and the maximum value of the gate voltage must be carefully designed such that the transistor does not enter triode region. This translates to a lower efficiency in theory due to the minimum required drain-source voltage. In contrast, the gate-drain voltage of the transistor can take on values as close as to  $0.2 (V_{th})$  or even larger in a switching amplifier to allow the transistor to enter the triode region.

As shown in Fig. 3.4, the transistor operates in the triode region for almost more than 35% of the period. Of course, these concepts poorly retain their original meaning at mm-wave frequencies in the neighborhood of 90 GHz. Figure 3.5 compares the output current of the transistor with the corresponding conceptual waveforms. As observed, at optimum value of  $V_{G-bias}$  ( $V_G=0.8$  V), the current waveform looks more like a pulse (in this figure,  $V_G=1.2$  V corresponds to the Case-1, and  $V_G=0.4$  V to the Case-4). Figure 3.6 represents the gate and drain voltages and the drain current waveform in the final topology.



Figure 3.4 Gate-drain voltage of the single transistor.

For best efficiency, it might appear that the gate bias should not be chosen just on the basis of the lowest I<sub>fundamental</sub>/I<sub>DC</sub> in Fig. 3.3. It should be based on the bias that maximizes  $P_{fundamental}/P_{DC}$ . These are the same in the presence of passive loss which can be the significant determiner of  $P_{fundamental}$  ( $I_{fund}^2.R_{passive}$ ) at these frequencies. In other words, it is not sufficient to make the current look sinusoidal. The loss in the passives that assist in shaping the current (in this case making it more sinusoidal in Fig. 3.1) at these frequencies can play a critical role in what power is actually delivered at the fundamental. It may even be possible that the loss in the passives that makes the

current sinusoidal may be large enough, that a design point with less fundamental component but lower loss may result in better efficiency.



Figure 3.5 The output current of the transistor that also shows the corresponding cases in Fig. 3.1.



Figure 3.6 The output waveforms of the final oscillator, including VG, VD, and ID.

To provide more details on locating the point of maximum efficiency where  $\eta = P_{fundametal}/P_{DC}$  is maximized especially after including passives that provide the waveform shaping for maximum  $\eta$ , we add the following.

At the target frequency, on-chip inductors normally exhibit lower quality factor as compared to capacitors. This is specially the case when capacitors with small values are used in the design. In regard to the value of the inductive load at the output load, a large value is desired as explained later. Once, the maximum allowable, proper inductor size is chosen based on the maximum achievable quality factor at the frequency of interest (in this case 91 GHz), the rest of the wave shaping is carried out through putting only an additional capacitor at the source of the transistor. The design parameters are selected through finding the optimum values by simulation. As a result, the small number of passives utilized in the design do not introduce much loss. Besides, a large inductive load directs the current toward the load at the output. At mm-wave t-lines are a better choice than inductors. T-lines are implemented by coplanar transmission lines that have a welldefined return path, high isolation, reduced coupling to substrate, and accordingly a lower loss. Improving  $I_{\text{fundamental}}/I_{\text{DC}}$  by primarily reducing  $I_{\text{DC}}$ , can have its own share in efficiency enhancement. Of course, there are other factors that can boost the efficiency that is not of main concern in this design. Hence for simplicity we have neglected the power dissipation in the matching network which we believe can be tolerated as the Q of the T-line is typically large enough. To reiterate, in connection with our design, the small number of passives employed to perform wave shaping do not contribute much loss. This is one of the factors which results in efficiency improvement in this work.

As pointed out, the general shape of  $I_D$  is also correlated to the time-constant of the RC parasitic path when the transistor is off. Hence, the efficiency can still be boosted if the time-constant is further modified. This is accomplished by placing a capacitor ( $C_S$ ) at the source of the transistor. As such, the additional capacitor would be effectively in series with  $C_{GS}$ , further reducing the timeconstant of the parasitic network. This allows  $C_S$  to shape the  $I_D$  waveform, enabling the designer to produce a closer-to-ideal sine-shape signals.  $C_S$  can also increase the delivered power without affecting the DC power consumption. In general, power gain ( $P_G$ ) is a measure expressing how efficiently a transistor operates. A high  $P_G$  implies a higher conversion between the DC power and the generated output power. Therefore, to boost  $P_G$ , one can possibly reduce the input power while keeping  $P_{L-f0}$  constant. If the transistor is pushed to the verge of instability, its input resistance ( $R_{in}$ ) becomes very small. Any further push beyond this point turns  $R_{in}$  negative, delivering power rather absorbing any thereafter. The presence of  $C_S$  at the source affects the input resistance favorably and enhances  $P_G$ .

Fig. 3.7 illustrates the value of the input resistance seen at the transistor gate versus  $C_S$ . As discussed,  $C_S$  reduces the input resistance. Also, Fig. 3.8 shows how the proper choice of  $C_S$  can improve both efficiency and  $P_G$ . Suggested by the figure, choosing a small value  $C_S$  can improve the transistor  $P_G$  and efficiency while it also helps form a pulse-like signal by improving  $\tau$ .



Figure 3.7 Equivalent negative resistance for both of the single-transistor and cross-coupled structure as a function of  $C_s$ .



Figure 3.8 Effect of introducing an additional capacitor at the transistor source on the efficiency and power gain in the open-loop configuration.

## 3.2.2 The Start-up/Loop Gain Condition for Oscillation Build-up

A major difference between the design of an amplifier and an oscillator is the start-up condition. Single-transistor implementation of the proposed oscillator is not feasible at the target center frequency (91 GHz) as it is too close to the  $f_{max}$  (the post-layout simulated maximum oscillation frequency of the employed process is under 116 GHz). Even with the boosting effect introduced by using  $C_s$ , still the transistor cannot sustain oscillation due to the poor Q of the passives. Therefore, having an additional active network around the transistor is critical to "start-up" the oscillation. In this setting, the transistors are configured in a cross-coupled structure. The smallsignal analysis of the equivalent resistance seen from the drains of the transistors yields to,

$$Re\{Z_{in}\} = -\frac{\omega^2 g_m C_{GS} + \omega^2 g_m C_S (C_{GS} + C_S)}{(\omega^2 g_m C_{GS})^2 + (\omega g_m C_S)^2} \approx -\frac{2C_{GS}}{g_m C_S} - \frac{1}{g_m}$$
(3.3)

where  $\omega$  is angular frequency,  $g_m$  is the transistor transconductance,  $C_{GS}$  is the gate-source parasitic capacitance, and  $C_S$  is the proposed capacitor at the source. The latter approximation holds at the target operation frequency. As seen in Equ. (3.3), a small capacitor ( $C_S$ ) at the source can potentially further relax the start-up condition, if  $C_{GS} > C_S$ . The trade-off is that the transistor transconductance can be degraded. Figure 3.8 also shows the effect of such capacitor on the equivalent negative resistance for the cross-coupled structure.

As confirmed by the plot, a smaller  $C_S$  results in a larger negative R. In addition, a negative input resistance for each single transistor in the loop independently results in a higher voltage swing, thereby enhancing the power generation. Also, the presence of the second transistor allows for using smaller size transistors, offering a smaller subsequent  $C_{GS}$  that have the dominant effect on the RC-network time constant.

It seems there are three criteria to select  $C_S$  for the best performance; maximize  $P_G$ , guarantee start up, and adjust time constant of the circuit when the switch is OFF to maintain  $P_L$ . To describe how  $C_S$  is chosen to address these three criteria successfully, the trade-offs are as follow. In general, if a capacitor is placed at the source of an NMOS transistor, the resulted network pushes the transistor toward instability. As observed in Fig. 3.7, reducing  $C_S$  improves the negative resistance seen looking into the drain of the cross-coupled structure. This is also confirmed by Eq. (3.3). At the same time, this capacitor also independently pushes each individual transistor in the cross-coupled configuration to the verge of instability. The blue line represents the equivalent resistance looking into the gate for each of the single transistors in the cross-coupled circuit in Fig. 3.7. A small input resistance translates to smaller input power while the generated output power remains almost intact.  $P_G$ -wise, one can surmise that the power gain is boosted. Thus, a small value of  $C_S$  enhances both the  $P_G$  and the start-up condition. On the other hand, if  $C_{GD}$ ,  $C_{GS}$ , and  $C_S$ considered to primarily form the input capacitive network, a small  $C_S$  can improve the time constant when the transistor is turned off. As such,  $C_S$  would be in series with the other two capacitors, lowering the total equivalent capacitance and reducing the consumed power (I<sub>D</sub>) by the transistor. This effect can maximally enhance the efficiency up to 1% in the final oscillator structure. On the downside, the transistor transconductance reduces which can turn out to be prohibitively low at smaller gate bias levels. This proves to be the trade-off between the V<sub>G,bias</sub> and the value of  $C_S$ . Also,  $C_S$  and the associated feed-line required to bias the source in the proposed configuration dissipate power and thus introduce power loss.

### 3.3 Circuit Implementation Details

Figure 3.9 depicts the schematic view of the proposed VCO architecture. Compared to the conventional cross-coupled structure, the proposed topology benefits from an additional set of capacitors at the source to improve the efficiency. Also, the capacitive connections ( $C_C$ ) between the transistors remove the bias dependency between the gate and drain nodes. A buffer is included at the output to absorb the loading effect. The buffer receives the generated power from the oscillator and transfers the energy to the  $50\Omega$  load. In the buffer, the transistor is no longer expected to yield a large power gain and thus consumes the minimal DC power. A matching network including the feedline  $TL_3$  and the RF pad parasitic capacitance (around 30 fF) is designed at the buffer output. While the transistor being off, the parasitics, output resistive load, and inductive load in the tank form a high-order filter. With no matching at the output, if the whole parasitic network presents an equivalent capacitance of  $C_T$  to the rest of the network, the result can be approximated by a second-order parallel RLC circuit. If L increases and  $C_T$  is reduced, the circuit is pushed to be operating in an under-damped mode (damping factor < 1) and the initial charges can be discharged more quickly. Therefore, the inductive load is designed to take on the largest possible value, given the achievable quality factors at the target frequency and  $C_T$  is further
reduced. A large inductive load can also more effectively direct the current toward the output. Given a Q of 30 achievable at 91 GHz, L is set to be around 80 pH initially. Next, assuming that L is large enough to supply a nearly constant current for design purposes, transistors widths are selected based on the analysis. Gate bias voltages ( $V_G$ ) and the nominal value of the source capacitors ( $C_S$ ) are chosen to be 0.8 V and 60 *fF* based on the considerations in Section 3.2.

In this design, turning off the transistor clips the drain current. Linear PAs typically recover this by using filtering. In this work, the device capacitances are exploited to effectively perform this waveform shaping. Clearly any finite PA choke (tank inductance or t-line when in a VCO configuration) will also play a critical role in shaping this waveform, (not just the RC time constant).



Figure 3.9 Schematic of the implemented high-efficiency fundamental VCO, the output buffer, and the matching network.

How are the values of passives chosen to ensure a sinusoidal waveform for the current? Once the transistor is powered off, the transistor's parasitic capacitance together with the output load, and the inductive load form an RLC circuit. Assuming that the whole parasitic network presents a  $C_T$  to the rest of the circuit and having no matching between the transistor and the load, the RLC circuit can be approximated as a second-order parallel RLC. If the circuit is pushed to be operating in an under-damped mode (damping factor ( $\xi$ ) < 1) as mentioned earlier, the initial charges can be discharged more quickly. As a result, to increase Neper frequency or attenuation ( $\alpha$ ), we design the inductive load to be as large as possible (given the achievable quality factors at the target frequency) and modify the  $C_T$  network subsequently. As a result, given the L set first, the device capacitance and bias voltage is exploited to effectively perform wave shaping. In such a scheme, we can assume that L is large enough to supply a nearly constant current (for design purposes). Therefore,  $C_T$  plays the critical role hereafter.

In Fig. 3.1, it may seem that is not clear why the current reduces linearly when the device is off (which would happen if the inductive load was shorted directly to ground. To comment on this, it is repeated that the decay would be in the form of a natural response of an approximately a second-order parallel RLC circuit and, in case of a large L, an RC-discharging circuit as in Fig. 3.1(c). However, in this figure, we conceptually draw the line to further highlight the effect of time-constant and illustrate how non-idealities can impact the waveform in extreme cases, shown in Fig. 3.1. To bias the transistor source a  $\lambda/4$  transmission line is implemented as an RF choke (RFC). The designed feedline presents an open-circuit condition at the frequency of operation. The transmission lines are implemented as coplanar waveguide with ground shielding. The efficiency boosting source capacitors are implemented using a pair of accumulation-mode varactors ( $C_{var}$ ) to tune the frequency as well. To account for the effect of interconnects, the entire structure is

simulated using HFSS. However,  $C_{var}(C_S)$  in Fig. 3.9 affect the phase noise and negative resistance since  $C_{var}(C_S)$  has a loss. In other words,  $C_{var}$  at 91 GHz would have very low Q. So, why it is a good idea to use a varactor, and not a capacitor across the source terminals? How would the VCO output swing vary across the tuning range?

The main reason for utilizing  $C_{var}$  and not a capacitor at the source is to provide a tuning range. Putting  $C_{var}$  in parallel with the tank can potentially kill the oscillation at such a high frequency. Even if the oscillation sustains, the output power would be small, the efficiency would be poor, and the noise performance would be bad. Of course, employing  $C_{var}$  instead of  $C_s$  degrades the negative resistance and the phase noise performance as the quality factor is very low. However, if the capacitive range  $(C_{max}/C_{min})$  is selected to be small with small  $C_{var}$ , the varactor yields an acceptable quality factor. As such, we can accommodate a tuning range while not diminishing the quality factor directly. In this design, however, due to a high oscillation frequency  $(0.8 \times f_{max})$ , the transistor models become inaccurate and the idea of tuning has not been well worked (inaccurate models have resulted in discrepancies between simulation and measurement). Thus, the measured tuning range is limited. Nevertheless, if the VCO output is considered for the whole frequency range, the output swing does not vary much as the output power maintains almost the same value (to be presented in the next section).

#### **3.4 Measurement Setup and Experimental Results**

The implemented fundamental VCO is tested in chip-on-board configuration through on-chip probing. A WR-08 GSG Picoprobe is used to probe the output of the VCO. The circuit is connected to a Rhode & Schwarz FSW 67-GHz spectrum analyzer (SA) with a WR-8 frequency extender for frequency and phase noise measurements. Figure 3.10 presents the measured PN of the VCO at 1,

3 and 10 MHz offsets from the carrier which features a phase noise of better than -86.2 dBc/Hz at 1–MHz offset. The output power is measured using an Erikson PM4 power meter. The measured oscillation frequency and the calibrated efficiency versus control voltage (V<sub>tune</sub>) are plotted in Fig. 3.11(a). The proposed oscillator consumes 46 mW of DC power and achieves a peak output power of 4.5 dBm with the record efficiency of 6.1%. Figure 3.11(b) illustrates the simulated efficiency and output oscillation frequency ( $f_{osc}$ ) plots versus V<sub>tune</sub>.

The difference between the simulation and measurement can be attributed to inaccuracy of device models at frequencies very close to  $f_{\text{max}}$ . As observed, the tuning range of the VCO is limited. The principal objectives of this work have been improving efficiency and then output power.



Figure 3.10 The measured phase noise of the proposed VCO at 89.75 GHz offset frequency. The simulated PN at the same offset frequency is also denoted by the red dotted line for comparison.

The tuning range has also been considered for the oscillator to meet the system requirements and account for uncertainties introduced as a result of variations in temperature and process. Nonetheless, no specific circuit technique is employed to improve the tuning range in this design.

The varactors are placed at the source of the cross-coupled oscillator to restrict their negative impact on the oscillation, attributed to their limited quality factor.



Figure 3.11 (a) Measured oscillation frequency ( $f_{osc}$ ) and calculated efficiency vs.  $V_{tune}$ ; (b) Simulated oscillation frequency and the corresponding calculated efficiency vs. control voltage ( $V_{tune}$ ).

This, however, cannot handle the negative impact of the varactors on the oscillation entirely. The designed VCO is operating at  $0.8 \times f_{max}$  that is too close to the transistors'  $f_{max}$ . This implies that the transistor models are rather inaccurate. The difference between the measured and the simulated tuning range points to the same fact in this regard. Simulations exhibit a tuning range of more than 2.2%. The inaccuracy in device models is mainly attributed to the significant parasitics contributed by the transistor at a close-to- $f_{max}$  frequency. The tuning range can be improved by selecting smaller transistors in general (that is almost impossible in this design due to proximity to  $f_{max}$ ; we need "start-up") or scarifying a little bit of output power to accommodate some frequency variations. The oscillation frequency also can be lowered a bit to have a more accurate transistor model.

It is noteworthy that the negative effect of varactor's low-quality factor has already been compensated for in the proposed VCO. Furthermore, Fig. 3.12 shows the calculated peak efficiency as a function of the supply voltage level ( $V_{DD}$ ) along with the corresponding oscillation frequencies. The figure suggests that the efficiency is boosted if  $V_{DD}$  is increased.



Figure 3.12 The measured efficiency vs. supply voltage.

Figure 3.13 depicts the chip micrograph of the fabricated VCO. Table 3.1 compares this work with the prior art for two different supply voltage values. The 6.1% power efficiency is reported under 1.8-V V<sub>DD</sub> in 0.13- $\mu$ m CMOS (nominal V<sub>DD</sub> of 1.2 V). We can comment on the voltage swing in the oscillator and the long-term reliability considerations of using such a high-power supply in a power oscillator? For example, if we assume 0.8 V swing (= V<sub>gate,bias</sub>) with anti-phase swing in a cross-couple oscillator this corresponds to a V<sub>gd</sub> voltage of 2.6 V which is more than the typical 2×V<sub>DD</sub> limit on the voltage between any two device node for long term reliability. Although the nominal V<sub>DD</sub> for the 0.13- $\mu$ m CMOS process employed in this work is 1.2 V, based on the design kit the devices can almost reliably operate with the supply voltage of 1.5 V and even 1.6 V.



Figure 3.13 The die micrograph of the 91 GHz VCO.

We have reported the efficiency values for  $V_{DD}$  greater than 1.6 V to have a complete set of measurement results. In regard to "ac-reliability", the gate swing in this design barely reaches values as high as 1.6 peak-to-peak. As such, even assuming an entirely anti-phase swing between the gate and drain voltages (which rarely happens at such mm-wave frequencies), the peak-to-peak  $V_{gd}$  does not exceed the 2×V<sub>DD</sub> limit. The drain voltage swing is around 400 mV in this design. The designed VCO shows the peak-to-peak of ~2.4 V for  $V_{gd}$ .

The standard oscillator figure of merit does not reward high efficiency (DC-ot-RF efficiency), which is the focus of this work. Possibly that is why many authors generally do not report FoM in the comparison table when the focus is the efficiency. The generated output power of this work is significantly higher than the other reported works. That is the case for efficiency too. The only comparable work in terms of power is the VCO in [81] with a lower efficiency, implemented in a CMOS process with a higher  $f_{max}$ . As can be observed, the phase noise performance of the VCO in [81] is very poor. Thus, overall, given the output power and efficiency to be the target

specifications, the presented VCO offers the highest efficiency while having an acceptable phase noise performance. Implementing the VCO with a CMOS process suffering from having transistors with lower value of  $f_{max}$  exhibits the effectiveness of the design method. we have removed the FoM data from the table of comparison to avoid any possible confusion.

Ref.	[71]	[72]	[81]	[82]	This Work	
CMOS Tech	65	65	90	65	0.13	0.13
	nm	nm	nm	nm	μm	μm
Freq. [GHz]	98	105	97.7	94.6	89.8	91
FTR [%]	8.7	9.5	7.2	5.6	0.6	0.5
P <sub>DC</sub> [mW]	21.6	54*	63	14	22	46
Pout [dBm]	-1	4.5‡	4.7	-4	0	4.5
DC-to-RF [%]	3.7	5.2*	4.7	2.9†	4.7	6.1
PN [dBc/Hz]	-90	-92.8	-80	-87	-86.2	-87
@ 1-MHz						
V <sub>DD</sub>	1.8	1.2	1.8	0.8	1.2	1.8

Table 3.1 Table of comparison with the-state-of-the-art

Power consumption of the buffer not included.

<sup>†</sup>Differential output power.

<sup>‡</sup>Use of an off-chip external bias-T for the RF output.

#### 3.5 Conclusion

In this work, we demonstrate the merits of employing a fundamental approach to primarily design a high-efficiency oscillator. Adopting the proposed method, we introduce techniques to successfully design and implement a VCO in a 0.13-µm CMOS process with record DC-to-RF efficiency numbers. The proposed VCO incorporates a capacitor at the source and manipulates the gate voltage of the transistors to improve the efficiency and relax the start-up condition. The

presented VCO achieves a DC-to-RF efficiency of 6.1% which to the best of the authors' knowledge is the highest among the W-band and higher-frequency CMOS sources. The proposed VCO can be used in high-efficiency power sources required to realize low-power transceivers.

# **Chapter 4: Efficient mm-Wave Signal Generation on CMOS**

# 4.1 Introduction

Due to availability of wider bandwidth and less populated frequency spectra, the lower frequency bands of the mm-wave spectrum are being explored for a variety of communication, surveillance, and imaging applications [1]. This includes automotive radar at 77-GHz band with a great commercialized viability [83], passive radar imaging in W-band [84], and multi-gigabit wireless communication at 60 GHz [85]. For many of these applications, especially those that are portable, energy efficiency is a critical issue, requiring circuit and system solutions with minimal impact on battery life.

Recent examples of high-efficiency mm-wave oscillators have been demonstrated on CMOS in [86] and [81]–[82] to realize the above-mentioned systems. In [86], the fundamental VCO effectively manipulates the dc current of the drain of the transistors to minimize the time during which the transistor is on while maintaining the same fundamental generated power. The oscillator achieves 6.1% DC-to-RF efficiency in 0.13- $\mu$ m CMOS with a limited tuning range. In [65], the theoretical limits on the drain efficiency for a single transistor are investigated and the major differences between theory and practice are explored to design a high-efficiency mm-wave oscillator utilizing a Class-E oscillator operating at 45 GHz. The design achieves a DC-to-RF efficiency of 15.6%. In this design, the underlying assumption that in an ideal Class-E power amplifier the main switch is opened and closed with zero rise/fall time pulses is revisited. Using a CMOS transistor as a switch, the associated on-resistance and the parasitic drain capacitance are identified as the primary sources that cause departure from the ideal Class-E behavior. Thus, the maximum oscillation frequency ( $f_{osc}$ ) for the minimal efficiency degradation is achieved. This

approach, however, imposes a limit on  $f_{osc}$  in the design. In [88], an 88-GHz fundamental SiGe HBT oscillator is presented by utilizing a device-centric design technique to achieve a high efficiency. The oscillator benefits from a self-feeding oscillator topology, employing the technique presented in [21] to optimize the fundamental power generation and exhibit 19.4% DC-to-RF efficiency. The oscillator, however, does not fully exploit the device capacity as it requires a low supply voltage for the optimum condition.

### 4.2 A High-Efficiency Single-Active-Device-Based Oscillator

At very high frequencies, e.g., mm-wave frequencies, the share of passive devices in power loss become significant due to the limited quality factors (Q) offered by these elements. Any oscillator design at very high frequencies are highly impacted by such detrimental effects. In other words, the maximum achievable quality factor numbers at high frequencies impose a limit on the minimum size of the transistor required for oscillation buildup. This inherently increases the power dissipation by the transistor itself. In fact, the dominance of the parasitics in large transistors restricts power conversion capability in transistors, i.e., conversion of dc supply power to  $P_{out}$ . In CMOS, this forces the transistors to operate in current limited mode with a certain voltage swing on the gate, failing to generate large swing levels at the output. This significantly reduces the transistor efficiency when it is places in a feedback loop to form an oscillator. As such, it is required to lower the supply voltage level to enable efficient active operations when it is undesirable by nature.

To design a high-efficiency oscillator, it is necessary to obtain the power efficiency limits of a single transistor. With the setting in Fig. 4.1, we obtain the maximum power added efficiency (PAE) for a single transistor without employing wave-shaping techniques. To maximize the

efficiency except for the inductive load at the output, the bias voltage of the gate,  $V_B$ , and the voltage amplitude on the gate,  $S_B$ , are the key parameters.



Fig. 4.1 The setting utilized to obtain maximum PAE for a transistor.



Fig 4.2 PAE versus  $V_B$  for several voltage amplitudes on the gate.

The simulations show that in CMOS transistors, PAE increases with the growing voltage amplitude on the gate, as shown in Fig. 4.2. In this figure, PAE is plotted versus the gate bias voltage for a number of swing levels. In Fig. 4.3, the same plot is presented for the power consumption,  $P_{dc}$ . As observed, the peak PAE for larger swings in Fig. 4.2 occurs at  $V_B$  where the  $P_{dc}$  becomes independent of the voltage level on the gate in Fig. 4.3,  $V_B|_{optimum} = V_B = 0.8 V$ . It

is an important finding in oscillator design where providing the start-up condition does not degrade the transistor efficiency. This effectively allows to set the  $V_B$  required to provide the loop transfer when the transistor is placed in an oscillator architecture while accommodating the ideal condition for the peak PAE. In this context,  $V_B$  is a function of voltage swing and is larger for higher  $S_B$ . The inductive load on the drain side is the other parameter which controls  $V_B$ . Once the optimum  $V_B$  is set, the swing must be increased to achieve the best performance according to Fig. 4.2. The simulation shows a voltage swing of 700 mV is required for PAE<sub>max</sub> = 25% (with wave-shaping, PAEs as large as 35% are achievable).



Fig. 4.3  $P_{dc}$  versus  $V_B$  for several swing levels.

The waveforms in Fig. 4.4 explain the presence of such swing-independent point. The dc component  $(I_{DC})$  in the output drain current is a function of voltage swing on the gate and also the bias level. In this context,  $V_B$  is the primary factor in determining  $I_{DC}$  for bias voltages greater than  $V_B|_{optimum}$  while the voltage swing on the gate mainly controls this component for  $V_B < V_B|_{optimum}$ .



time (ps)

Fig. 4.4 Gate voltage and drain current waveforms for different levels of swing and bias voltage.

In any single-transistor oscillator topology the output swing can be sent back to the input with maximum a ratio of 1. With limited maximum swing at very high frequencies, this returning swing is not sufficient to reproduce the ideal efficient condition for the transistor. Furthermore, to implement a single-transistor oscillator (to achieve maximum DC-to-RF efficiency at mm-wave frequencies), the output must be connected to the input such that it provides the required power and phase condition for oscillation.

Colpitts oscillator topology is a candidate where the start-up condition is harder to be provided. Additionally, the resulting voltage swing at the input side due to a capacitive voltage division is even smaller than the output. In self-feeding oscillator topologies, a transmission line is employed with a certain loading condition to sustain oscillation [73]. However, the limited number of degrees of freedom in this configuration and limitations in passive usage do not allow the transistor to capture the ideal condition for a high efficiency at the input and output. Hence, a feedback network with higher number of degrees of freedom is required. As pointed out, higher-order passive feedback circuits are not viable options at frequencies up to 100 GHz. This leads to utilization of active-based feedbacks where the degrees of freedom are numerous while not consisting of passives. In this topology, the transistor in the active feedback network (AFN) does not participate in the oscillation necessarily and, as a result consumes power minimally.

To increase the swing on the gate with  $V_B = V_B|_{optimum}$ , the AFN must source an in-phase current as shown by red arrow in Fig. 4.5. To this end, we inspect the input impedance of the core transistor looking into the gate  $(Z_{in})$ . In Fig. 4.6,  $|Z_{in}|$  is plotted versus frequency. With a peak at the design frequency, it is reliably concluded that the voltage amplitude on the gate is increased (remarks: in CMOS, the input power increases with increasing swing slightly even at higher frequencies and does not impact the efficiency significantly). To tolerate the minimum power consumption in the AFN stage, the size of the transistor, the gate voltage bias, and the inductive load cannot be arbitrarily large and thus, the resulting  $S_G$  on the gate still might not be as large as desired. As  $P_{dc}(total) = P_{dc}(core transistor) + P_{dc}(AFN)$ , it turns out that that  $S_B = S_{ideal}$ is achieved where  $P_{dc}(AFN)$  levels up. This is one of the reasons why  $V_B = V_B|_{optimum}$  so that the efficiency is minimally impacted where  $S_B \neq S_{ideal}$ .

Utilizing the ideal passive components in the proposed structure, we observe that the oscillator has been able to achieve efficiency numbers close to the efficiency limits of the transistors in the process. This well shows the merits of the proposed method. In conventional cross-coupled topologies, the oscillator cannot achieve a DC-to-RF efficiency close to those of the transistors. The comparison has been presented in Fig. 4.7. To enable tuning, we substitute the source capacitors with varactors  $(C_{var})$  [72]. By placing the varactors at the source of the core transistors, the original tank is shielded from the negative impacts of the low quality-factor varactors at such high frequencies of operation.



Fig. 4.5 Pictorial representation to implement the AFN for an increased swing.



Fig. 4.6 Z<sub>in</sub> versus frequency.

The capacitor  $C_{var}$  changes the equivalent inductance of the proposed topology, thereby tuning the frequency of the oscillation. As the optimized value of  $R_L$  is larger than 50  $\Omega$  (based on simulations), a capacitive matching network can perform the load transformation (and no inductors are needed in the matching network).



Fig. 4.7 DC-to-RF efficiency versus frequency for the proposed AFN-based oscillator structure and the conventional cross coupled oscillator.

# 4.3 Circuit Implementation Details

As a proof-of-concept for the presented circuit techniques, we have designed, implemented, and tested a fundamental VCO with the center frequency of 85.75 GHz in a 65-nm CMOS process. Figure 4.8 displays the schematic view of the proposed VCO. The proposed VCO benefits from an included active-feedback single-transistor architecture. The core transistor has  $W = 15 \mu m$  with 10 fingers (of 1.5  $\mu$ m each) and the AFN stage incorporates a smaller transistor with  $W = 8 \mu m$  with 8 fingers. To implement the decoupling capacitors (*C<sub>d</sub>*), the interdigitated capacitors are used

to improve the quality factor and provide a higher intrinsic self-resonance frequency (SRF). The VCO uses a varactor in the source of the core transistor to tune the frequency. To bias the transistors' sources and have freedom in selecting the value of the varactor's size and also lowering the silicon area, an *LC* tank is employed at the source of the transistor instead of an RF-chock. As such, the tuning range can potentially be extended as the value of  $C_{var}$  does not have to be close to  $C_{GS}$  necessarily. To achieve the maximum *Q* at the target frequency, the inductive loads are implemented as CPW transmission-lines with ground shielding.



Figure 4.8 The schematic view of the proposed 85.75 GHz VCO.

The equivalent inductance of the implemented transmission-lines are around 140 pH and 90 pH. The varactor in the core stage provides a reasonable tuning range (TR) and adding another varactor in AFN can extend the tuning range for an additional percent (~0.9%). At the output, the oscillator is connected to a small, low-capacitive RF pad (30  $\mu$ m × 45  $\mu$ m). The capacitance of the pad from the simulations is estimated to be around 10 fF. A small capacitor,  $C_m = 20 fF$ , is placed at the 72 output to form the required matching network with pad to transform the 50- $\alpha$  load to the optimum value of  $R_L$  required for maximum PAE. The inductance of supply  $(V_{DD})$  traces and other interconnects are also carefully modeled and included in the simulation

# 4.4 Measurement Setup and Experimental Results

For the purpose of testing, the chip is mounted on a printed circuit board (PCB) and all the dc biases are wire bonded. Figure 4.9 shows the measurement setup for the VCO to determine the frequency of oscillation, phase noise and output power. A 75-to-110 GHz GSG (ground-signal-ground) Cascade Infinity Probe is used to probe the output of the VCO. The probe is directly connected to a harmonic mixer to mix down the signal. The local oscillator (LO) and intermediate frequency (IF) ports of the mixer are connected to the corresponding outputs on a spectrum analyzer. The output frequency is determined by sweeping LO frequency at various harmonics and measuring IF change. The output power of the oscillator is measured using a VDI Erikson PM4 power meter. Figure 4.10 presents a chip photograph of the fabricated VCO.



Fig. 4.9 The measurement setup for the proposed VCO.



Figure 4.10 The annotated die micrograph of the fabricated fundamental 89.3 GHz VCO.

The measured peak output power and  $f_{osc}$  are plotted in Fig. 4.11. In Fig. 4.12 the simulated peak DC-to-RF efficiency and  $f_{osc}$  are compared with those from measurement. The measurement results are in good agreement with the simulations. The proposed VCO consumes ~8.5 mW of dc power and achieves a measured output power of 1 dBm for the record peak DC-to-RF efficiency of 14.8% when operating at 89.3 GHz. In Fig. 4.12(a), the output frequency can be continuously tuned from 82.2 GHz to 89.3 GHz which is translated to a frequency tuning range of more than 8.2%.

The measurement results are in a good agreement with the simulation results. The proposed VCO consumes 8.5 mW of dc power and achieves a measured peak output power of 1 dBm with the record DC-to-RF efficiency of 14.8% when operating at 89.3 GHz (the high end of tuning range).



Fig. 4.11 The measured oscillation frequency and output-power versus V<sub>tune</sub>.



Figure 4.12 (a) The measured oscillation frequency and DC-to-RF efficiency vs.  $V_{tune}$  (b) The simulated oscillation frequency and efficiency vs. control voltage.

As seen in Fig. 4.12(a), the output frequency can be continuously tuned from 82.2 GHz to 89.3 GHz which is translated to a frequency tuning range of more than 8.2%. Figure 4.13 also includes the measured PN of the VCO at 100–kHz, 1–MHz, and 10–MHz offsets from an 89.3-GHz carrier which features a phase noise of better than –91.8 dBc/Hz at 1–MHz offset.



Figure 4.13: The measured phase noise at 89.3 GHz.

Ref.	F Tech. [0 0.13-μm	Freq.	FTR	P <sub>DC</sub>	Pout	DC-to-RF	PN [dBc/Hz]	V <sub>DD</sub>	<sup>(3)</sup> FOM*	(4) FOM <sub>T</sub> *
		[GHz]	[%]	[mW]	[dBm]	[%]		[V]	[dBc/Hz]	[dBc/Hz]
		91	0.5	46	45	61	-86 2 @ 1 MHz	1.8	-168.8 @	-142.8 @
[72]	CMOS	71	0.5	-10	т.9	0.1	00.2 @ 1 WIIIZ	1.0	1 MHz	1 MHz
	65-nm	105	0.5	(1) 54	4.5	(1)5 2	02 8 @ 1 MIL-	1.2	-175.9 @	-175.5 @
	CMOS	105	9.5	(*) 54	4.5	(5).2	-92.8 @ 1 MHZ	1.2	1 MHz	1 MHz
	65-nm	102.2	5.2	21	2	2	-112.1 @	1.2	-179.2 @	-175.5 @
	CMOS	5.2	21	-2	5	10 MHz	1.2	10 MHz	10 MHz	
[71]	65-nm	98	8.7	21.6	-1	3.7	-90 @ 1 MHz	1.8	-176.5 @	-175.3 @
	CMOS							1.0	1 MHZ	1 MHz
	0.13-µm	88	1 1					16	-191.1 @	
٢٥٥]	BiCMOS	00	1.1	23	1.5	19.4	-107 @ 1 MHZ	1.0	1 MHz	-
This	65-nm	83	85	1	14.8	_01 8 @ 1 MHz	1.2	-181.5 @	-179.8 @	
Work	CMOS	09.5	0.2	0.5	1	14.0	->1.o @ 1 MHZ	1.2	1 MHz	1 MHz

<sup>(1)</sup> Buffer's P<sub>dc</sub> not included. <sup>(3)</sup>  $FOM = PN(f_{offset}) - 20log\left(\frac{f_{osc}}{f_{offset}}\right) + 10log(\frac{P_{DC}}{1mW}).$ <sup>(4)</sup>  $FOM_T = PN(f_{offset}) - 20log\left(\frac{f_{osc}}{f_{offset}}, \frac{\Delta f}{10}\right) + 10log(\frac{P_{DC}}{1mW})$ 

Table 4.1 Performance comparison of W-band VCOs with this work.

The VCO attains the figure of merit (FoM) of -181.5 dBc/Hz and FoM<sub>T</sub> of -179.8 dBc/Hz at 1-MHz offset frequency. The performance of the chip is summarized in Table 4.1.

# 4.5 Conclusion

In this work, we propose an active-feedback-based single-transistor oscillator architecture to demonstrate a high-efficiency 85.75-GHz VCO that is implemented in a 65-nm CMOS process. A single-transistor active feedback network is used to provide the optimum signal swing condition at the gate of the core transistor for maximum PAE. This transistor offers additional degrees of freedom to realize the oscillator with the maximum PAE with minimal power overhead. The fabricated VCO achieves a record peak DC-to-RF efficiency of 14.8%, which to the best of authors' knowledge is the highest reported for mm-wave CMOS oscillators to date. The DC-to-RF efficiency number reported in this work is close to those from compound-semiconductor oscillators [89].

# Chapter 5: A mm-Wave VCO with Switchable Triple-Coupled Transformer

### 5.1 Introduction

The availability of a wide bandwidth in the mm-wave frequency spectrum, in general, and the 60-GHz band, in particular, makes it attractive for implementation of high-data-rate wireless communication standards such as IEEE 802.15.3c and IEEE 802.11ad. Such frequency band allocations enable multi-gigabit wireless communication services, including wireless personal area networks (WPANs) and wireless local area networks (WLANs) at the 60 GHz frequency band. Moreover, the mm-wave bands above 60 GHz support services such as 77-GHz automotive radars and passive radar imaging. In this context, 5G wireless systems are also expanding their operation frequency to mm-wave bands where an increased signal bandwidth is available [90].

VCOs are amongst the essential building blocks required for operation of integrated mm-wave transceivers. One of the main challenges in designing such transceivers is to synthesize an on-chip local oscillator (LO) with PN and wide tuning range (TR). Design of LO-generation circuits at high frequencies requires special considerations due to the following reasons. First, the low Q of passive elements and the lossy silicon substrate largely limit the tank's equivalent Q which, in turn, degrades the overall noise performance. Secondly, at very high frequencies a large gm-cell is required to establish the oscillation as the frequency of operation becomes close to the transistors'  $f_{max}$  (maximum oscillation frequency). Finally, the parasitics contributed by the gm-cell and interconnects significantly restrain the TR.

In general, capacitive and inductive tuning are the two methods to realize the tuning range in VCO topologies [91]–[93]. Capacitive tuning techniques, however, are not commonly employed at mm-wave frequencies due to the degraded Q of required capacitive components [91]. In contrast,

several number of mm-wave VCOs have been introduced based on the inductive tuning methods thus far [91]–[93]. In these works, two variable inductor-based VCOs are introduced to achieve a wide TR. In these structures, the effective inductance is varied by either a resistor or varactor in the secondary coil which are controlled by an analog input signal. However, the designs consume a large amount of power which results in a degraded figure-of-merit (FOM). An inductor-loaded transformer is proposed in [91] with switches placed on the secondary coil at different locations, creating an increased number of sub-bands to extend the tuning range. The relatively large number of switches required for the symmetric implementation in this work increases the total load capacitance reflected back to the primary coil. In [92], a coupled VCO with two core blocks is designed to realize a dual-mode operation at the cost of a higher power consumption and a larger occupied area. In [93], switched inductors are utilized to change the effective inductance of the tank, thereby extending the tuning range at the cost of a degraded noise performance. The designed VCO also occupies a large area and has a high power-consumption. In another example, Reference [94] benefits from two active-mode switching blocks where by turning these blocks on and off, the overall VCO achieves a wide tuning range and a high output power while the PN and area are compromised. In short, a mm-wave VCO which can achieve a decent performance (i.e., with an extended TR and improved PN performance) is highly desirable.

In this work, we introduce a design guideline for a transformer-based mm-wave VCO architecture to extend the TR. The analysis for the proposed transformer, presented in section 2, shows that the TR can be extended while the PN degradation is minimal. Based on the analysis, differential switches are utilized in the secondary and tertiary coils to reduce the off-state parasitic capacitance. This improves the effective quality factor seen by the primary coil as well as the TR. Employing a three-dimensional (3D) implementation of the transformer (i.e., each coil is laid out

in a different metal layer), the mutual coupling factors between each pair of the coupling coils are improved, leading to a lowered effective on-state resistance (due to the switches) at the primary coil. In overall, the proposed VCO offers a compact size, wide tuning range, and low phase noise. An NMOS Class-C VCO is used as the core oscillator as this structure typically exhibits a better performance in terms of power consumption and phase noise as compared to traditional crosscoupled LC-VCOs.

## 5.2 The Coarse Tuning Approach

An inductive tuning method based on mutual coupling between the coils is proposed. In this method, the coarse tuning is realized by switching the loads at the secondary and tertiary coils in and out of the equivalent circuit seen by the primary coil. As such, the magnetic flux passing through the primary coil is varied which in turn changes the effective inductance seen at the primary side. The proposed transformer structure includes three distinct magnetically coupled one-turn spiral inductors where the secondary and tertiary coils each is terminated by a switch. In this setting, four distinct frequency sub-bands are formed by virtue of different possible states of the switched coils. To fine tune the circuit, varactors are utilized to continuously cover the entire frequency tuning range. In the following, we present a design guideline for achieving a wide tuning range and a low phase noise VCO with a compact area, using the switchable transformer-based tank described in [95]. The major issues arise in the design of the proposed transformer-based VCO and various design trade-offs to be made are discussed in this section. Furthermore, analytical expressions are derived for the proposed transformer to highlight the effects of non-idealities introduced by the switches.

# 5.2.1 The Proposed Triple-Coupled Transformer Model

Unlike capacitors, developing an accurate model for on-chip inductors is a challenging task in general. In an on-chip inductor, the terminal characteristics is strongly correlated to its geometric layout dependent due to the skin effects, eddy currents, and displacement currents emerging at very high frequencies. A simple and relatively accurate model for an integrated inductor is presented in Fig. 5.1. Shown in this figure, in this model, L, R, C<sub>X</sub>, C<sub>P</sub>, R<sub>Sub</sub>, and C<sub>Sub</sub> represent the self-inductance, series resistance of the inductor, total parallel stray capacitance, parasitic capacitance, and the substrate resistance and capacitance, respectively. To evaluate and compare the performance of different inductor geometries, these parameters should be modelled with an acceptable precision. In the proposed triple-coupled transformer we incorporate the presented simple model in Fig. 5.1 for each individual coil to model the whole transformer. The resulted model is shown in Fig. 5.2. In this figure,  $k_{ij}$  and  $R_{si}$  represent the coupling factor between the corresponding  $i^{th}$  and  $j^{th}$  coils and the total loss of the  $i^{th}$  inductor, respectively. The transformer model shown in Fig. 5.2 complicates the analysis. To gain a better understanding of the behavior of the transformer, thus, we propose to simplify the circuit to arrive at the model in Fig. 3 that includes the switches. This model is more practical and relatively accurate due to following reasons.

First, as the frequency of operation is significantly lower than the self-resonance frequency (SRF), one can ignore the effect of the parasitic capacitance  $C_X$  which affects the SRF. Also, the substrate loss and the parasitic capacitance  $C_P$  affect the quality factor of the inductor without losing the generality of the proposed technique.



Figure 5.1 A high-frequency circuit model of an integrated inductor.



Figure 5.2 The high-frequency circuit model of the proposed triple-coupled transformer.

This effect is considered in the simulations later. Note that presented simplified model well provides the insights and intuitions required to understand the proposed technique.



Figure 5.3 A simplified circuit model of the proposed switchable triple-coupled transformer.

To analytically examine the circuit shown in Fig. 5.3, without loss of generality, the following simplifying assumptions are considered; (i) the off-state resistance and on-state parasitic capacitance of the switches are negligible. (ii) The loss of the secondary and tertiary coils are assumed to be the same ( $R_{52} = R_{53} = R_5$ ), and (iii) the mutual inductance between the secondary and primary coils is the same as that of the individual auxiliary and the primary one ( $M_{12} = M_{13} = M$ ). The assumption of  $M_{12}=M_{13}=M$  was made as it greatly simplifies the derivations. In the implemented prototype, the values of  $M_{12}$  and  $M_{13}$  are indeed close. However, this is a design choice and, in general,  $M_{12}$  and  $M_{13}$  can be different. According to Fig. 5.3, we can write the following equations;

$$\begin{pmatrix} V_1 \\ V_2 \\ V_3 \end{pmatrix} = \begin{pmatrix} j\omega L_1 & j\omega M & j\omega M \\ j\omega M & j\omega L_2 & j\omega M_{23} \\ j\omega M & j\omega M_{23} & j\omega L_3 \end{pmatrix} \times \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix}$$
(5.1)

$$\begin{pmatrix} V_2 \\ V_3 \end{pmatrix} = \begin{pmatrix} -Z_{L2} & 0 \\ 0 & -Z_{L3} \end{pmatrix} \times \begin{pmatrix} I_2 \\ I_3 \end{pmatrix}$$
(5.2)

where  $Z_{L2}$  and  $Z_{L3}$  include the equivalent impedance of the switch and the loss of the corresponding coil as shown in Fig. 5.3. From (5.1) and (5.2),  $I_3$  can be found as;

$$I_{3} = I_{2} \left[ \frac{j\omega L_{2} + Z_{L2} - j\omega M_{23}}{j\omega L_{3} + Z_{L3} - j\omega M_{23}} \right] = T_{1} I_{2}$$
(5.3)

By substituting (5.3) in (5.1),  $I_2$  can be written as;

$$I_{2} = \left[\frac{-j\omega M}{j\omega L_{2} + Z_{L2} + j\omega M_{23}T_{1}}\right]I_{1} = T_{2}I_{1}$$
(5.4)

where  $T_1$  and  $T_2$  are defined as;

$$T_{1} = \frac{j\omega L_{2} + Z_{L2} - j\omega M_{23}}{j\omega L_{3} + Z_{L3} - j\omega M_{23}}$$
(5.5)

$$T_2 = \frac{-j\omega M}{j\omega L_2 + Z_{L2} + j\omega M_{23}T_1}$$
(5.6)

By substituting (5.3) and (5.4) in (5.1), the input impedance of  $Z_{in}$  can be found as;

$$Z_{in} = j\omega L_1 + j\omega M T_2 (1+T_1) \Longrightarrow$$

$$(5.7)$$

$$Z_{in} = j\omega L_{1} + \frac{\omega^{2} M^{2} (j\omega L_{2} + j\omega L_{3} + Z_{L2} + Z_{L3} - j2\omega M_{23})}{\omega^{2} M_{23}^{2} - \omega^{2} L_{2} L_{3} + Z_{L2} Z_{L3} + j\omega L_{2} Z_{L3} + j\omega L_{3} Z_{L2}}$$
(5.8)

Note that (5.8) is symmetric with respect to  $L_2$  and  $L_3$ . This holds for all the possible switch states. In the following, we derive the input impedance at the primary coil, specifically for each switch state.

## a) State 1: (on-on)

In this state both switches are on and  $Z_{L2}$  and  $Z_{L3}$  are equal to  $(R_S+R_{ON})$ . The value of  $(R_S+R_{ON})$  is relatively small and is negligible as compared to other terms in both the numerator and denominator of (5.5), and thus T<sub>1</sub> simplifies as;

$$T_1 \approx \frac{L_2 - M_{23}}{L_3 - M_{23}} \tag{5.9}$$

The input impedance in (5.8) is simplified as;

$$Z_{in} \approx j\omega L_1 + \frac{\omega^2 M^2}{j\omega L_2 + Z_L + j\omega M_{23}T_1} (1 + T_1)$$
(5.10)

Next, defining A and B as;

$$A = \frac{[L_2 + L_3 - 2M_{23}]}{L_3 - M_{23}} = 1 + T_1$$
(5.11)

$$B = M_{23} \times \frac{L_2 - M_{23}}{L_3 - M_{23}} = M_{23}T_1$$
(5.12)

Then, *Z*<sub>*in*</sub> can be written as;

$$Z_{in} \approx j\omega L_1 + \frac{A.\omega^2 . M^2}{(R_s + R_{on}) + j\omega(L_2 + B)}$$
(5.13)

Equation (5.13) is not symmetric with respect to  $L_2$  and  $L_3$  (it can be shown that it has a symmetric imaginary part while its real part is asymmetric), however, in the frequency band of interest (5.9) presents a good approximation of the original equation which is symmetric. As such, the equivalent inductance and resistance seen by the primary coil can be found as;

$$\frac{L_{eq} \approx L_{1} - \frac{\omega^{2}M^{2} \cdot (1 + \frac{L_{2} - M_{23}}{L_{3} - M_{23}}) \cdot (L_{2} + M_{23} \cdot \frac{L_{2} - M_{23}}{L_{3} - M_{23}})}{(R_{s} + R_{on})^{2} + \omega^{2}(L_{2} + M_{23} \cdot \frac{L_{2} - M_{23}}{L_{3} - M_{23}})^{2}} \approx L_{1} - C_{2}$$
(5.14)

$$\frac{R_{eq} \approx}{\frac{\omega^2 . M^2 . (1 + \frac{L_2 - M_{23}}{L_3 - M_{23}}) . (R_s + R_{on})}{(R_s + R_{on})^2 + \omega^2 (L_2 + M_{23} \frac{L_2 - M_{23}}{L_3 - M_{23}})^2} \approx (R_s + R_{on}) . C_1}$$
(5.15)

where  $C_1$  and  $C_2$  are defined as;

$$C_{1} = \frac{M^{2} \cdot (1 + \frac{L_{2} - M_{23}}{L_{3} - M_{23}})}{(L_{2} + M_{23} \cdot \frac{L_{2} - M_{23}}{L_{3} - M_{23}})^{2}}$$
(5.16)

$$C_{2} = \frac{M^{2} \cdot (1 + \frac{L_{2} - M_{23}}{L_{3} - M_{23}})}{L_{2} + M_{23} \cdot \frac{L_{2} - M_{23}}{L_{3} - M_{23}}}$$
(5.17)

The approximations in (5.14) and (5.15) hold if  $(R_{on} + R_S)$  is small compared to the second term in the denominator which is a reasonable assumption in the frequency band of interest. According to (5.16) and (5.17),  $C_1$  and  $C_2$  have the same nominator while the denominator of  $C_1$  is the square of that of  $C_2$ . Thus, from (5.14) and (5.15), one can manipulate  $C_1$  and  $C_2$  to reduce  $R_{eq}$  at the cost of a small increase in  $L_{eff}$ . For a fair comparison of the PN performance, the effective quality factor can be written as;

$$Q_{eff} \approx \frac{(L_1 - C_2)\omega}{(R_s + R_{on})C_1}$$
(5.18)

86

From (5.16) to (5.18) and considering that  $M_{23} = k_{23}\sqrt{L_2L_3}$ ,  $Q_{eff}$  is a strong function of the coupling factor between the secondary and tertiary coils  $(k_{23})$  and in turn mutual inductance  $(M_{23})$ . This also implies that an increase in the value of  $k_{23}$  drops the equivalent resistance while slightly increasing the equivalent inductance, thereby improving the quality factor at the cost of a slightly degraded TR. In this context, if  $k_{23}$  increases from 0.3 to 0.5, the equivalent resistance reduces by ~20% while the corresponding inductance increases by only ~7%. These changes result in an overall improvement of ~34% in the tank quality factor. To assess the coarse-tuning, we define a parameter called "effective inductance variation", EIV, as  $EIV = \frac{\Delta L_{eff}f}{L_1} \times 100 = \frac{C_2}{L_1} \times 100$ . This parameter approximately predicts the tuning range behavior. Fig. 5.4 shows  $Q_{eff}$  and EIV versus  $k_{23}$  for the design parameters used in the setting. The plots confirms the above observations.



Figure 5.4 Qeff and C2 versus k23 for given design parameters.

To intuitively explain this observation, we consider two ideal cases shown in Fig. 5 where  $k_{23} = 0$  and  $k_{23} = 1$ , respectively. In Fig. 5.5(a), assuming  $L_2 = L_3 = L$  and 87

 $Z_{L2} = Z_{L3} \approx R_{on}$ , we analyze the case where both switches are either on or off (that is,  $Z_{L2}=Z_{L3}$ ). For simplicity, we also assume that  $L_2=L_3=L$  and  $R_{S2}=R_{S3}=R_S$ . Thus, we have  $V_2=V_3$  and  $I_2=I_3=I$ . If the transformer is designed such that  $M_{12}\approx M_{13}\approx M$ , applying the aforementioned assumption in (1), we arrive at;

$$\begin{pmatrix} V_1 \\ V_2 \\ V_3 \end{pmatrix} = \begin{pmatrix} j\omega L_1 & j\omega M & j\omega M \\ j\omega M & j\omega L & j\omega M_{23} \\ j\omega M & j\omega M_{23} & j\omega L \end{pmatrix} \times \begin{pmatrix} I_1 \\ I \\ I \end{pmatrix}$$
(5.19)

If  $k_{23}=0$ , we have;



Figure 5.5 Simplified  $\pi$ -model of the proposed transformer when; (a)  $k_{23} = 0$ . (b)  $k_{23} = 1$ .

If the three coils structure is modeled as in Fig 5.6,



Figure 5.6 Simplified  $\pi$ -model of the proposed triple-coupled transformer.

 $V_1$  and  $V_2$  can be rewritten as,

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} j\omega L_1 & j\omega M \\ j\omega M & j\omega L_{eq} \end{pmatrix} \times \begin{pmatrix} i_1 \\ i_2 \end{pmatrix}$$
(5.21)

Comparing (5.20) and (5.21), the right-hand side of (5.20) should be equal to that of (5.21). This gives  $L_{eq}=0.5L$  and  $i_2=2I=2I_2$ , and thus,  $\frac{V_2}{I_2}=\frac{V_2}{I}=R_{ON}$  and  $R_{eq}=\frac{V_2}{i_2}$ . Given  $i_2=2I=2I_2$ , we can conclude that  $R_{eq}=0.5R_{on}$ . For the other case where  $k_{23}=1$ ,  $V_1$  and  $V_2$  are derived as;

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} j\omega L_1 & j\omega M \\ j\omega M & j\omega L \end{pmatrix} \times \begin{pmatrix} 2I \\ 2I \end{pmatrix}$$
(5.22)

Comparing (5.22) with (5.21), again the right-hand side of (5.22) should be equal to that of (5.20), resulting in  $L_{eq}=0.5L$  and  $i_2=2I=2I_2$  and consequently  $\frac{V_2}{I_2}=\frac{V_2}{I}=R_{ON}$  and  $R_{eq}=\frac{V_2}{I_2}$ . With  $i_2=2I=2I_2$ , we have  $R_{eq}=0.5R_{on}$ . Hence, he equivalent inductance and resistance at the primary coil are derived as  $L_{eq} = 0.5L - M$  and  $R_{eq} = 0.5R_{on}$  in Fig. 5.5(a) while the corresponding parameters are obtained as  $L_{eq} = L - M$  and  $R_{eq} = 0.5R_{on}$  in Fig. 5.5(b) where  $k_{23} = 1$ . Thus, the quality factor at the primary side is doubled if the two loads are perfectly coupled. This is in agreement with our previous observation that increasing  $k_{23}$  improves the effective quality factor at the primary side. As can be seen from (5.14) to (5.18), the effective inductance and quality factor at the primary side are also dependent on the values of  $L_2$  and  $L_3$  (remember  $M_{23} = k_{23}\sqrt{L_2L_3}$ ). As discussed earlier,  $k_{23}$  can be maximized in order to achieve the maximum quality factor. Then, for a given  $k_{23}$ ,  $L_2$  and  $L_3$  can be optimized, based on (5.18), to achieve a better phase noise performance. As the coupling factor is geometry dependent, the values of  $L_2$  and  $L_3$  can be varied by changing the width of the coils with only a small change in the geometry which in turn approximately keeps the corresponding coupling factor  $(k_{23})$  constant. Simulation results show that  $k_{23}$  remains almost intact by preserving the original structure of the coils even if  $L_2$  and  $L_3$  vary. 89

Figures 5.7 and 5.8 show the plot of the effective quality factor and inductance variation for different values of  $L_2$  and  $L_3$ . As observed,  $L_2$  and  $L_3$  should increase to improve  $Q_{\text{eff}}$ . In other words, for a given  $k_{23}$ , larger coupled inductors enhance  $Q_{eq}$ . However, we cannot increase  $L_2$  and  $L_3$  arbitrarily as it leads to a larger occupied area, higher parasitic capacitance, and larger loss with only a marginal improvement in  $Q_{eq}$ . Therefore,  $L_2$  and  $L_3$  are optimized to achieve a low phase noise while having minimal adverse effect on the tuning range.



Figure 5.7 Effective quality factor versus L2 with a constant L3.

#### b) State 2: (off-off)

If both switches are off,  $Z_{L2}$  and  $Z_{L3}$  are equal to  $(R_s - \frac{j}{\omega C_{off}})$ . Having a high Q for each

inductor,  $R_s$  can be neglected as compared to other terms in (23). Thus, the simplified  $T_1$  becomes;

$$T_{1} = \frac{R_{s} + j\omega(L_{2} - (\omega^{2}C_{off})^{-1} - M_{23})}{R_{s} + j\omega(L_{3} - (\omega^{2}C_{off})^{-1} - M_{23})}$$

$$\approx \frac{L_{2} - (\omega^{2}C_{off})^{-1} - M_{23}}{L_{3} - (\omega^{2}C_{off})^{-1} - M_{23}}$$
(5.23)

90


Figure 5.8 Effective inductance variation versus L<sub>2</sub> with a constant L<sub>3</sub>.

The input impedance in this case can be calculated as;

$$\frac{Z_{in} \approx j\omega L_{1} + \frac{\omega^{2}M^{2}}{j\omega L_{2} - j(\omega C_{off})^{-1} + R_{s} + j\omega M_{23}T_{1}} (1 + T_{1})$$
(5.24)

Again, defining *A* and *B* as;

$$A = \frac{[L_2 + L_3 - 2(\omega^2 C_{off})^{-1} - 2M_{23}]}{L_3 - (\omega^2 C_{off})^{-1} - M_{23}}$$
(5.25)

$$B = M_{23} \times \frac{L_2 - (\omega^2 C_{off})^{-1} - M_{23}}{L_3 - (\omega^2 C_{off})^{-1} - M_{23}}$$
(5.26)

In this case, the input impedance can be re-written as;

$$Z_{in} \approx j\omega L_1 + \frac{A.\omega^2.M^2}{j\omega (L_2 + B) - j(\omega C_{off})^{-1} + R_s}$$
(5.27)

91

Thus, the effective inductance and resistance can be found as;

$$L_{eq} \approx L_1 - \frac{A.\omega.M^2}{\omega(L_2 + B) - (\omega C_{off})^{-1}}$$
(5.28)

$$R_{eq} = \frac{\omega^2 M^2}{\left[\omega (L_2 + B) - (\omega C_{off})^{-1}\right]^2} R_s$$
(5.29)

As can be observed, although both switches are off, they still adversely affect the performance of the VCO by reducing the effective inductance and in turn  $Q_{eff}$ . To minimize the adverse effect of  $C_{off}$  on the performance, this capacitance must be as small as possible, dictating a small size for the switch. On the other hand, a small size switch increases the on-state resistance which deteriorates the phase noise performance. To lower  $C_{off}$  without compromising  $R_{on}$ , a differential switch is utilized, shown in Fig. 5.9. In this configuration, the on-state resistance of M<sub>3</sub> is halved as compared to a single-ended switch with the same size. Thus, a halved size switch is required in the differential implementation to have the same on-state resistance. This means that the off-state capacitance is halved which increases  $L_{eff}$  and subsequently  $Q_{eff}$ . Figure 5.10 illustrates the variations of the effective inductance versus the off-state parasitic capacitance. By reducing  $C_{off}$ from 60 fF to 30 fF, effective inductance increases by ~44%.



Figure 5.9 Circuit implementation of the differential switch.



Figure 5.10 Effective inductance variation versus the off-state parasitic capacitance of the switch.

### c) State 3: (on-off)

In Fig. 3, if SW<sub>1</sub> is on and SW<sub>2</sub> is off, we have  $Z_{L2} = R_{on} + R_S$  and  $Z_{L3} = \frac{1}{j\omega C_{off}} + R_S$ .

Thus, the input impedance (refer to (5.7)) can be found as;

$$Z_{in} = j\omega L_1 + \frac{\omega^2 M^2}{j\omega L_2 + (R_{on} + R_S) + j\omega M_{23}T_1} (1 + T_1)$$
(5.30)

where T<sub>1</sub> is modified as;

$$T_1 \approx \frac{L_2 - M_{23}}{L_3 - (\omega^2 C_{off})^{-1} - M_{23}}$$
(5.31)

Thus, the equivalent inductance and resistance seen by the primary coil can be found as;

$$L_{eq} \approx L_{1} - \frac{\omega M^{2}}{\omega L_{2} + \frac{\omega^{2} M_{23} (L_{2} - M_{23})}{\omega L_{3} - (\omega C_{off})^{-1} - \omega M_{23}}}.$$

$$\frac{L_{2} + L_{3} - 2M_{23} - (\omega^{2} C_{off})^{-1}}{L_{3} - (\omega^{2} C_{off})^{-1} - M_{23}}$$

$$R_{eq} \approx \frac{\omega^{2} M^{2}}{(\omega L_{2} + \frac{\omega^{2} M_{23} (L_{2} - M_{23})}{\omega L_{3} - (\omega C_{off})^{-1} - \omega M_{23}})^{2}}.$$

$$\frac{L_{2} + L_{3} - 2M_{23} - (\omega^{2} C_{off})^{-1}}{L_{3} - (\omega^{2} C_{off})^{-1} - M_{23}}.$$
(5.33)

Equations (5.32) and (5.33) indicate that the existence of the tertiary coil still impacts the equivalent inductance and resistance at the primary even when the switch is off and the tertiary coil is open. The design parameters  $C_{off}$  and  $k_{23}$  (and in turn  $M_{23}$ ) significantly affect the performance. Figures 5.11 and 5.12 show the effective quality factor and effective inductance variation versus  $C_{off}$  and  $k_{23}$ , respectively. As seen, the PN and TR performance behavior of this sub-band are the same as those of state-1 and state-2 which were discussed earlier.



Figure 5.11 Effective quality factor and effective inductance variation versus the off-state parasitic capacitance of the switch.



Figure 5.12 Effective quality factor and effective inductance variation versus the coupling factor between the loads.

# d) State 4: (off-on)

In the last state, the secondary coil is open and the tertiary one is shorted. Thus, the equivalent inductance and resistance can be found as;

$$L_{eq} \approx L_{1} - \frac{\omega M^{2}}{\omega L_{3} + \frac{\omega^{2} M_{23} (L_{3} - M_{23})}{\omega L_{2} - (\omega C_{off})^{-1} - \omega M_{23}}}.$$

$$\frac{L_{2} + L_{3} - 2M_{23} - (\omega^{2} C_{off})^{-1}}{L_{2} - (\omega^{2} C_{off})^{-1} - M_{23}}$$
(5.34)

$$R_{eq} \approx \frac{\omega^2 M^2}{(\omega L_3 + \frac{\omega^2 M_{23} (L_3 - M_{23})}{\omega L_2 - (\omega C_{off})^{-1} - \omega M_{23}})^2}.$$

$$\frac{L_2 + L_3 - 2M_{23} - (\omega^2 C_{off})^{-1}}{L_2 - (\omega^2 C_{off})^{-1} - M_{23}}.(R_S + R_{on})$$
(5.35)

By comparing (5.34) and (5.32) and considering that  $L_2$  is not equal to  $L_3$ , we can conclude that the center frequency of the second and third sub-bands are different. Thus, two distinct sub-bands can be achieved where just one of the switches are on. The behavior of this sub-band is the same as that of the previous state.

#### 5.2.2 Verification of The Proposed Model

To validate the simplified model of the proposed transformer, we compare simulation results for the cases in which inductors are modelled in Keysight's advanced design system (ADS) Momentum on the layout of the proposed transformer and the case employing the simplified *R* and *L* model with constant mutual coupling. In the latter case, the equivalent *L* and *R* of each inductor are extracted from electromagnetic (EM) simulation of the corresponding individual inductor. Figure 5.13 compares the simulated quality factor seen at the primary coil in two cases when  $k_{23} =$ 0 and  $k_{23}=0.5$ , respectively.

The quality factor at the primary coil is improved as the coupling factor between the loads increases (in this case from 0 to 0.5). Both the EM simulations on the layout and simulations on the simplified schematic view show the same trend. Thus, although the quality factor extracted from the simplified model is different from that of the EM layout, the predicted trends derived in equations from the simplified circuit are still valid and the effective quality factor is improved as the coupling factor increases. Same conclusion can be drawn for the effective inductance. This confirms the usefulness of the proposed model (Fig. 5.3) in predicting the behavior of the circuit and its design.



Figure 5.13 Quality factor comparison between the proposed model and simulation of the transformer.

Fig. 5.14 shows the flowchart for the step-by-step guideline to design the wide tuning range VCO.

## 5.3 VCO Implementation and Design

An NMOS class-C topology with a better performance in terms of PN and power consumption is utilized to design the core oscillator in this work. The schematic view of the proposed VCO is shown in Fig. 5.15. In this figure, M<sub>4</sub> and M<sub>5</sub> are the cross-coupled transistors configuring the  $g_m$ cell of the VCO.  $C_{d1}$  and  $C_{d2}$  are the decoupling capacitors that are chosen to be large so as to lower the loop gain degradation. The combination of control voltage and digital inputs allows for having fine and coarse control over the entire frequency tuning range.

Two bits are considered to control the two NMOS switches which are placed in the secondary and tertiary coils. The size of the switches are optimized to achieve a reasonable compromise between the on-state resistance and the off-state parasitic capacitance. Thus, the width of  $M_{1,2}$  and  $M_3$  are selected to be 5 and 25 µm, respectively, with the minimum length of 60 nm.



Figure 5.14 Flowchart of the summarized design procedures.

This would result in  $R_{on}$  and  $C_{off}$  of 7 ohms and 28 fF, correspondingly. The tank quality factor is mainly determined by the loss of the inductor, metal-insulator-metal (MIM) capacitors, and MOS varactors. At mm-wave frequencies, the quality factor of MOS varactors limits the tank quality factor. Therefore, the varactors are designed large enough to cover the minimum required overlap between the adjacent sub-bands. The length and width of the varactors are set to be 600 nm with a finger number of 4, respectively.

To begin the design, it is necessary to know the initial values of the quality factor and the effective inductance for various switch states. Figures 5.16 and 5.17 present and compare the EM simulation results of  $Q_{\text{eff}}$  and  $L_{\text{eff}}$  over different switch states.



Figure 5.15 Schematic of the proposed VCO.

As can be seen, when both switches are off, the auxiliary coils have minimum effect on the primary coil due to the relatively high-impedance of off-state parasitic capacitance of switches, and thus, the quality factor and effective inductance in the off-off state are the highest amongst all the possible states. When  $SW_1$  is off and  $SW_2$  is on, the tertiary coil is shorted while the secondary one is open. This configuration achieves the second highest effective inductance and quality factor due to the lower coupling factor among the corresponding coil owing to larger distance to the primary coil. When  $SW_1$  is on and  $SW_2$  is off, due to the higher coupling factor which is the corollary of a shorter distance to the primary coil, the effective inductance of this sub-band is lower than that of the previous sub-band. However, the quality factor is the worst compared to the other states due to the lower effective inductance as well as larger switch loading of this sub-band. Finally, for the state when both switches are on, the effect of both auxiliary coils are added to further reduce the effective inductance. The quality factor of this setting is better than that of the previous sub-band. Switches are on, the effect of both auxiliary coils are added to further reduce the effective inductance. The quality factor of this setting is better than that of the previous has been both switches are on the effect of both auxiliary coils are added to further reduce the effective inductance.

previous sub-band (SW<sub>1</sub> on and SW<sub>2</sub> off). Although the effective inductance is reduced in this state, the loading effect is further reduced (we can intuitively say that the loads are in parallel which in turn results in a lower loss), which reduces the Q-factor degradation. Thus, the on-on state achieves a lower effective inductance along with a larger Q compared to the on-off state.



Figure 5.16 The simulated Q-factor over all switch states.



Figure 5.17 Simulated effective inductance over the entire frequency tuning range.

As discussed earlier, increasing the coupling factor between the coils can improve the phase noise by reducing the loading effect of the switches. To improve the coupling factor while occupying a compact area, the triple-coupled coils are implemented in the topmost three thick metal layers in the 65 nm CMOS technology which are metal 9 (M<sub>9</sub>), metal 8 (M<sub>8</sub>), and aluminum layer (AP). In this setting, the primary, secondary, and tertiary coils are implemented in M<sub>9</sub>, M<sub>8</sub>, and AP, respectively. As such, to increase the coupling factors, especially  $k_{23}$ , we implement the auxiliary coils in  $M_8$  and AP layers instead of utilizing the same  $M_8$  layer. There are challenges associated with a planar implementation of the auxiliary coils. This includes a larger silicon area, a smaller coupling factor, proximity effect due to the adjacent traces, and a higher parasitic capacitance between adjacent traces which can be alleviated by the proposed implementation. In this configuration, the minimum distance between two adjacent traces with the same metal layer, enforced by the design rules, is irrelevant as the coils are sitting on the different metal layers. Therefore, the coupling factor is improved due to the smaller distance between the implemented coils. Thus, the proposed three-dimensional transformer achieves a higher coupling factor as compared to the simple planar implementation. This would also minimize the occupied silicon area.

The quality factor differences among the four states would substantially degrade the performance. Figure 5.18 shows the frequency allocation based on the achieved quality factor of all the possible switch states. The  $3^{rd}$  and  $4^{th}$  sub-bands (on-off state and on-on state, respectively) have the worst quality factor (refer to Fig. 5.16) due to a larger loading effect and also lower effective inductance (refer to Fig. 5.17). In addition, the PN performance is further degraded in such sub-bands as the oscillation frequency is higher. The proposed transformer can ameliorate the aforementioned issue. The M<sub>8</sub> layer is utilized to implement the secondary coil since a larger

inductance with a smaller inner radius is required due to having a higher inductance per unit length as compared to AP. On the other hand, AP layer is an appropriate choice for the implementation of the tertiary coil since the outermost coil has a lower coupling factor to the primary coil. Thus, a ticker metal layer is required to compensate for the lower coupling factor. Note that AP layer can achieves a higher quality factor compared to  $M_8$  thanks to its larger thickness and skin depth.



Tertiary coil is closed.

Figure 5.18 Frequency allocation of the four switch states based on the effective quality factor.

According to the discussion in section 2, for a given coupling factor and switch size, the effective quality factor can be enhanced by enlarging the auxiliary coupled inductors in the proposed transformer at the cost of a slightly degraded tuning range. Therefore, it is recommended to design a larger coupled inductor to improve the PN performance. In this regard, there are two options for the implementation of the auxiliary coils; placing the tertiary coil either inside or outside of the secondary coil. Figures 5.19 (a) and (b) show the simulated coupling factor and Q for various possible configurations. In these configurations, the secondary coil is implemented in M<sub>8</sub> layer and the tertiary coil can be either implemented using M<sub>8</sub> (planar structure) or AP (the proposed structure). The tertiary coil can also be placed either outside or inside of the secondary coil. For a fair comparison, the spacing between the secondary and tertiary coils is chosen to be

the minimum distance allowed by the design rules for the planar implementation. As observed, implementing the tertiary coil using AP layer outside of the secondary coil achieves the highest coupling and quality factors amongst the possible configurations.



Figure 5.19 Simulation results of the proposed geometry for different configurations (a) k; (b) Q.

The cross-sectional view and the top view of the proposed three-dimensional transformer used in the tank which consists of three mutually coupled one-turn inductors shown in Figs. 5.20 and 5.21, respectively. The design parameters r,  $W_i$ , and  $S_i$  are the inner radius, width of  $i^{th}$  inductor, and the spacing between adjacent traces, respectively. The inner radius of the primary, secondary, and tertiary coils is 23, 50, and 62 µm. The width of the inductors is optimized to be 12 µm in order to have a large inductance while keeping the Q<sub>eff</sub> high.  $S_2$  and  $S_3$  are chosen to be zero in order to maximize the coupling factor without increasing the inter-layer parasitic capacitance. S<sub>1</sub> is chosen to be 3 µm to maximize the quality factor of the overall tank at the target frequency.

Based on the EM simulation results, the self-inductance of the primary, secondary, and tertiary coils are 137, 168, and 192 pH, respectively, and the resulted coupling factor between the adjacent 103

coils is about 0.55 at 60 GHz. The self-resonant frequency of the proposed triple-coupled transformer including the switches is 85 GHz.



Figure 5.20 The cross-section view of the proposed switchable transformer.

Typically, the quality factor of a spiral inductor is degraded at mm-wave frequencies due to the loss originating from the substrate and parasitic capacitances. Some layout-related techniques are utilized in this work to lower the associated high-frequency loss with the spiral inductors. A solution to reduce the substrate loss is to lower the number of turns as well as the width of the inductor in order to lower the effective substrate area under the inductor.

This would also increase the FTR and SRF of the tank. Moreover, a striped inductor structure is used in the tank to mitigate the loss at high frequencies [96]. Furthermore, a grounded shield ring using  $M_9$  is designed to enhance the quality factor. This structure functions similar to a coplanar waveguide in which the space between the ground and signal planes is optimized to achieve the maximum possible quality factor at the frequency band of interest. The optimized distance (*r*) in our design is 25 µm.



Figure 5.21 The top view of the proposed switchable transformer.

#### 5.4 Measurement Results

To confirm the effectiveness of the proposed architecture, a proof-of-concept prototype chip is designed and fabricated in a 65-nm CMOS process. Figure 5.22 shows the die micrograph of the fabricated VCO which is packaged in a ceramic flat package (CFP) with 24 pins (CFP-24) to facilitate DC biasing of the chip using bond wires. Excluding the pads, the active area of the circuit is 260  $\mu$ m × 250  $\mu$ m. Two open-drain output buffers are designed at the output of the VCO to drive a 50  $\Omega$  load. The gate of the buffers are connected directly to the output of the VCO to avoid using decoupling capacitors which limit the tuning range. The drain of the buffers are biased using external bias-tees. These buffers collectively draw 8 mA from a 1-V supply. The VCO output is measured using on-wafer probing while the DC biasing of the circuit is supplied by bond wires. Two Zprobe Z067V3N-GSG-100 are used as the output probes. The VCO output is measured

single-ended while the other output is terminated by a 50  $\Omega$  load. The output of the VCO is measured using a Rohde & Schwartz (R&S) FSW 26.5 GHz spectrum analyzer with an external R&S Z75 harmonic mixer to extend the operation frequency to V-band.

The loss of the cable, probe, and the harmonic mixer are estimated to be 5.5 and 28 dB, respectively, which are calibrated by the spectrum analyzer. At the center frequency, the fabricated VCO achieves a phase noise of -89.9 and -111.72 dBc/Hz at 1 and 10-MHz offset, respectively. The measured PN at 10-MHz offset from the carrier frequency is plotted for the entire frequency range in Fig. 5.23. In this figure, the best, average, and the worst-case phase noise are -114.2, -111.9, and -109.9 dBc/Hz, respectively. Also, the best PN performance is achieved in the first frequency band because; first, in this sub-band both switches are off which minimizes the loading effect of the switches, and secondly, the operation frequency is lower than that of other sub-bands which entails a better PN due to a larger signal period. The measured calibrated output power over the entire frequency tuning range is also shown in Fig. 5.24. As seen, the first sub-band achieves the largest output power as the transformer load is lower, compared to those of other sub-bands. Furthermore, based on Fig. 5.25, over different states of the switches, the tuning range including four frequency bands is over 18%, from 50.1 to 59.8 GHz, resulting in a FOM and FOM<sub>T</sub> of -178.9 and -184 dBc/Hz at 10-MHz offset, respectively.

Table 5.1 summarizes the performance summary of the proposed VCO and compares this work with the state-of-the-art. The performance of the VCO, particularly in terms of FOM and  $FOM_{T,}$  compares favorably with other designs.



Figure 5.22 Micrograph of the fabricated mm-wave VCO.



Figure 5.23 The measured PN at 10-MHz offset over the entire frequency tuning range.



Figure 5.24 The measured output power over the entire frequency tuning range.



Figure 5.25 Measured tuning range versus control voltage.

Reference	[91]	[97]	[98]*	[99]	This work
Freq. (GHz)	61	77	63	66.1	55
Ave. PN at 10 MHz (dBc/Hz)	-108.3	-108.4	-92.2	-103	-111.9
TR (%)	14.2	15.8	24.6	27.9	18
Power(mW)	6	14.3	21.5	13	6.2
FOM @ 10 MHz (dBc/Hz)	-176.2	-174.5	-174.7	-168.3	-178.9
FOMT @ 10 MHz (dBc/Hz)	-179.3	-178.5	-182.6	-177.2	-184

Table 5.1:Performance summary of 60-GHz VCOs in 65-nm CMOS

\* Measured at 1 MHz offset

$$FOM = PN(f_{offset}) - 20log\left(\frac{f_{osc}}{f_{offset}}\right) + 10log(\frac{P_{DC}}{1mW})$$

$$^{(4)}FOM_T = PN(f_{offset}) - 20log\left(\frac{f_{osc}}{f_{offset}}, \frac{\Delta f}{10}\right) + 10log(\frac{P_{DC}}{1mW})$$

### 5.5 Conclusion

A mm-wave triple-coil transformer-based VCO with wide tuning range and low phase noise is presented. It is proposed that maximizing the coupling factor between the secondary and tertiary coils as well as optimizing their individual values improves the phase noise substantially as compared to the conventional transformer-based designs. By switching in and/or out the secondary and tertiary coils of the transformer, 4 overlapping frequency sub-bands are resulted which collectively enable designing a wide tuning range. The design techniques that are introduced to 109

implement the high-Q tunable inductor for the tank allow the VCO to achieve a competitive phase noise and FOM. As a proof of concept, a 55-GHz VCO is implemented in a 65-nm CMOS process. Table 1 summarizes the performance of the proposed design and compares it with that of the stateof-the-art.

# **Chapter 6: Conclusion**

This thesis outlines the efforts on the road to touch the limits of silicon-based technologies for efficient signal generation at mm-wave frequencies. It goes beyond the current limits of electronic design in CMOS/BiCMOS technologies to achieve high speed, efficient, high performance circuits and systems overall. The proposed design procedures are employed to develop circuit topologies and system architectures offering the promise to realize applications in imaging, radar, sensing and communications for medical, security and internet systems within mm-wave band and beyond. This includes the circuits, systems, and the links for IoT as well. The major resulting achievements are summarized as below.

### 6.1 Achievements

### 6.1.1 A SiGe HBT mm-Wave Source for MRFM Imaging

As discussed in Chapter 2, the ultra-sensitive mechanical detection of electron spin resonance is opening up exciting possibilities for determining the three-dimensional structure of individual biomolecules and biological assemblies using nanoscale magnetic resonance imaging. In this work, we report on the design of chip-scale microwave sources, along with the required specifications, that are intended for use in a microwave-enhanced nanoscale magnetic resonance imaging experiment carried out at cryogenic temperatures. A methodology is proposed to systematically design, without the need for low-temperature transistor models, a VCO optimized for operation at low temperatures. Utilizing the proposed approach, a 40-GHz HBT VCO is designed and implemented in a 0.13-µm BiCMOS process. The VCO operates at temperatures ranging from 300 K to 12 K, efficiently furnishing the minimum output power required for a nanoscale magnetic resonance imaging experiment. The cryogenic VCO meets the technical requirements needed to implement the dynamic nuclear polarization effect in microcantileverbased cryogenic nanoscale magnetic resonance imaging experiments. In brief, the fundamentals of signal generation at very low temperatures is outlined.

#### 6.1.2 Close-to-fmax Efficient Signal Generation

As discussed in Chapter 3, in this work, we present a systematic method for designing highefficiency, high-power mm-wave oscillators. This method effectively manipulates the DC current of the drain of the core transistors to minimize the time during which the transistor is on. Furthermore, an additional capacitor at the source of the transistor assists lowering the power consumption while maintaining the same fundamental generated power. To show the feasibility of this method, a class-E-type oscillator architecture is implemented in a 0.13- $\mu$ m CMOS technology ( $f_{max} = 116$  GHz). The proposed VCO achieves the record DC-to-RF efficiency of 6.1% at a center frequency of 91 GHz (~ 0.8× $f_{max}$ ) on CMOS. The measured peak output power is 4.5 dBm while the VCO consumes 46 mW of DC power and features a phase noise figure of merit (FoM) of -169.6 dBc/Hz at 1–MHz offset frequency. The fabricated VCO occupies 0.51 mm<sup>2</sup> of silicon area including the pads.

### 6.1.3 Super-Efficient mm-Wave Signal Generation on CMOS

As discussed in Chapter 4, in this work, we present circuit techniques to design fundamental mm-wave high-efficiency VCOs. To increase the efficiency, we propose an oscillator topology that employs an active feedback network around the core transistor to amplify the voltage swing on the gate, thereby enhancing the generated fundamental current by the transistor and, in turn, the

DC-to-RF efficiency. Utilizing the proposed method, an 85.75-GHz VCO is designed and implemented in a 65-nm CMOS process. In this VCO, the tradeoffs among efficiency and tuning range are also considered and the VCO achieves a peak DC-to-RF efficiency of 14.8% at 89.3 GHz and a wide tuning range of more than 8.2% while consuming only 8.5 mW of dc power from a 1.2-V supply. Compared to the prior art, the fundamental VCO exhibits more than twice improvement in DC-to-RF efficiency. The oscillatory circuit occupies a silicon area of less than 0.25 mm<sup>2</sup> including the pads.

#### 6.1.4 A mm-Wave VCO with Switchable Triple-Coupled Transformer

As discussed in Chapter 5, in this work, we present a circuit technique to design a wide tuning range VCO. The employed technique extends the TR with a minimal adverse effect on the PN. In this context, a switchable triple-coupled transformer, implemented utilizing the topmost three metal layers of the process, is proposed to achieve a wide TR and a low PN. In the proposed configuration, the mutual coupling between the loads is increased while canceling the negative effects of the switches on the transformer performance. Depending on the possible states of the switches used in the secondary and tertiary coils, four overlapping frequency sub-bands are introduced. An extensive set of analysis are provided to support the idea. As a proof-of-concept, a 55-GHz VCO is implemented in a 65-nm CMOS process. Based on the measurement results, the VCO achieves an average PN of -111.9 dBc/Hz at 10-MHz offset over the entire frequency range and a TR of ~18%, from 50.1 to 59.8 GHz, exhibiting a figure of merit incorporating the tuning range (FOM<sub>T</sub>) of -184 dBc/Hz at 10-MHz. The VCO core consumes 6.2 mW from a 1-V supply and, excluding the pads, occupies a compact silicon area of 0.06 mm<sup>2</sup>.

# References

- [1] J. Laskar, S. Pinel, D. Dawn, S. Sarkar, B. Perumana, and P. Sen, "The next wireless wave is millimeter wave," *Microw. J.*, vol. 8, iss. 8, pp. 22-36, Aug. 2007.
- [2] J. Lee, Y. Li, M. Hung, and S. Huang, "A fully-integrated 77-GHz FMCW radar transceiver in 65-nm CMOS technology," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 12, pp. 2746 –2756, Dec. 2010.
- [3] A. Tomkins, P. Garcia, and S.P. Voinigescu, "A passive W-band imaging receiver in 65nm bulk CMOS," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 10, pp. 1981 –1991, Oct. 2010.
- [4] S. Emami, R. F. Wiser, E. Ali, M. G. Forbes, M. Q. Gordon, X. Guan, S. Lo, P. T. McElwee, J. Parker, J. R. Tani, J. M. Gilbert, and C. H. Doan, "A 60 GHz CMOS phased-array transceiver pair for multi-Gb/s wireless communications," in *IEEE Int. Solid-State Circuits Conf. Digest of Technical Papers*, San Francisco, CA, USA, Feb. 2011, pp. 164–166.
- [5] R. Wu, S. Kawai, Y. Seo, N. Fajri, K. Kimura, S. Sato, S. Kondo, T. Ueno, T. Siriburanon,
  S. Maki, B. Liu, Y. Wang, N. Nagashima, M. Miyahara, K. Okada, and A. Matsuzada,
  "13.6 A 42Gb/s 60GHz CMOS transceiver for IEEE 802.11ay," in *IEEE Int. Solid-State*Circuits Conf. Digest of Technical Papers, San Francisco, CA, USA, Feb. 2016, pp. 248-249.

- [6] C. H. Doan, S. Emami, D. A. Sobel, A. M. Niknejad, and R. W. Brodersen, "Design considerations for 60 GHz CMOS radios," *IEEE Commun. Mag.*, vol. 42, no. 12, pp. 132-140, Dec. 2004.
- [7] A. Valdes-Garcia, S. T. Nicolson, J.-W. Lai, A. Natarajan, P.-Y. Chen, S. K. Reynolds, J.-H. Zhan, D. G. Kam, D. Liu, and B. Floyd, "A fully integrated 16-element phased- array transmitter in SiGe BiCMOS for 60-GHz communications," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 12, pp. 2757–2773, 2010.
- [8] A. Babakhani, X. Guan, A. Komijani, A. Natarajan, and A. Hajimiri, "A 77-GHz phasedarray transceiver with on-chip antennas in silicon: Receiver and antennas," *IEEE Journal* of Solid-State Circuits, vol. 41, no. 12, pp. 2795–2806, 2006.
- [9] MC Ziskin, "Physiological mechanisms underlying millimeter wave therapy," pp. 241– 251. Springer Pub., 2005.
- [10] I.V. Okhryamkina E.A. Palamarchuk V.P. Tondy, L.D. Skresanov, V.N. Kas, "Complex treatment cardiovascular disease with a low power millimeter-wave radiation," *Physics* and Engineering of Millimeter and Sub-Millimeter Waves, 2001. The Fourth International Kharkov Symposium on, 2001, vol. 2, pp. 939–940.
- [11] M. Gupta, "Power gain in feedback amplifiers a classic revisited", *IEEE Trans. Microw. Theory Tech.*, vol. 40, no. 5, pp. 864-879, May 1992.
- [12] V. Radisic, D. Sawdai, D. Scott, W. Deal, L. Dang, D. Li, J. Chen, A. Fung, L. Samoska, T. Gaier, R. Lai, "Demonstration of a 311-GHz fundamental oscillator using InP HBT technology," *IEEE Trans. Microw. Theory Tech.*, vol. 55, no. 11, pp. 2329-2335, Nov. 2007.

- Z. Lao, J. Jensen, K. Guinn, M. Sokolich, "80-GHz differential VCO in InP SHBTs," *IEEE Microw*. Wireless Compon. Lett., vol. 14, no. 9, pp. 407-409, Sep. 2004.
- [14] M. Hossain, N. Weimann, V. Krozer, W. Heinrich, "A 100 GHz fundamental oscillator with 25% efficiency based on transferred-substrate InP-DHBT technology," *Microwave Conference (EuMC) 2016 46th European*, pp. 497-500, 2016.
- [15] B. Heydari, M. Bohsali, E. Adabi, A. Niknejad, "Millimeter-wave devices and circuit blocks up to 104 GHz in 90 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 42, no. 12, pp. 2893-2903, Dec. 2007.
- [16] E. Laskin, P. Chevalier, A. Chantre, B. Sautreuil, S. Voinigescu, "165-GHz transceiver in SiGe technology," *IEEE J. Solid-State Circuits*, vol. 43, no. 5, pp. 1087-1100, May 2008.
- [17] S. Nicolson, K. Yau, P. Chevalier, A. Chantre, B. Sautreuil, K. Tang, S. Voinigescu,
  "Design and scaling of W-band SiGe BiCMOS VCOs," *IEEE J. Solid-State Circuits*, vol. 42, no. 9, pp. 1821-1833, Sep. 2007.
- [18] R. Wanner, R. Lachner, G. Olbrich, "A monolithically integrated 190-GHz SiGe push-push oscillator," *IEEE Microw. Wireless Compon. Lett.*, vol. 15, no. 12, pp. 862-864, Dec. 2005.
- B. Razavi, "A 300-GHz fundamental oscillator in 65-nm CMOS technology," *IEEE Symp. VLSI Circuits Dig.*, pp. 113-114, 2010-Jun.
- [20] E. Seok, K. K. O, "A 410 GHz CMOS push-push oscillator with an on-chip patch antenna," 2008 IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, pp. 472-473, 2008-Feb.

- [21] O. Momeni, E. Afshari, "High power terahertz and millimeter-wave oscillator design: a systematic approach", *IEEE J. Solid-State Circuits*, vol. 46, no. 3, pp. 583-597, 2011.
- [22] R. Spence, Linear Active Networks, New York: Wiley-Interscience, 1970.
- [23] D. Ha, J. Paulsen, N. Sun, Y.-Q. Song, and D. Ham, "Scalable NMR Spectroscopy with Semiconductor Chips," *Proc. Natl. Acad. Sci.* vol. 111, pp. 11955-11960, Aug 2014.
- [24] J. Anders, J. Handwerker, M. Ortmanns, and G. Boero. "A Fully-integrated Detector for NMR Microscopy in 0.13 μm CMOS," 2013 IEEE Asian Solid-State Circuits Conference (A-SSCC), Singapore, 2013, pp. 437-440.
- [25] J. Anders, J. Handwerker, M. Ortmanns, and G. Boero, "A Low-power High-sensitivity Single-chip Receiver for NMR Microscopy," J. Magn. Reson. Vol. 266, pp. 41-50, May 2016.
- [26] T. Yalcin and G. Boero, "Single-chip Detector for Electron Spin Resonance Spectroscopy," *Rev. Sci. Instrum.* vol. 79, 094105 (2008).
- [27] J. Anders, A. Angerhofer, and G. Boero, "K-band Single-chip Electron Spin Resonance Detector," J. Magn. Reson. vol. 217, pp. 19-26, April 2012.
- [28] A. V. Matheoud, G. Gualco, M. Jeong, I. Zivkovic, J. Brugger, H. Ronnow, J. Anders, and G. Boero, "Single-chip Electron Spin Resonance Detectors Operating at 50 GHz, 92 GHz, and 146 GHz," *J. Magn. Reson.* vol. 278, pp. 113-121, May 2017.
- [29] G. Gualco, J. Anders, A. Sienkiewicz, S. Alberti, L. Forr, and G. Boero, "Cryogenic Singlechip Electron Spin Resonance Detector," J. Magn. Reson. vol. 247, pp. 96-103, Oct 2014.

- [30] A.V. Matheoud et al., "Single-chip electron spin resonance detectors operating at 50 GHz,
  92 GHz, and 146 GHz," *J. Magn. Reson.* 278 (2017) 113–121.
- [31] J. Handwerker, B. Schlecker, U. Wachter, P. Radermacher, M. Ortmanns, and J. Anders, "28.2 A 14GHz Battery-operated Point-of-care ESR Spectrometer Based on a 0.13 μm CMOS ASIC," 2016 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, 2016, pp. 476-477.
- B. Patra, R. Incandela, J. P. G. van Dijk, H. A. R. Homulle, L. Song, M. Shahmohammadi,
   R. B. Staszewski, A. Veladimirescu, M. Babaie, F. Sebastiano, E. Charbon, "Cryo-CMOS Circuits and Systems for Quantum Computing Applications," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 1, pp. 309-321, Jan. 2018.
- [33] J. D. Cressler, "Silicon-Germanium as an Enabling Technology for Extreme Environment Electronics," *IEEE Transactions on Device and Materials Reliability*, vol. 10, no. 4, pp. 437-448, Dec. 2010.
- [34] T. K. Thrivikraman, J. Yuan, J. C. Bardin, H. Mani, S. D. Phillips, W. L. Kuo, J. D. Cressler, S. Weinreb, "SiGe HBT X-Band LNAs for Ultra-Low-Noise Cryogenic Receivers," *IEEE Microwave and Wireless Components Letters*, vol. 18, no. 7, pp. 476-478, July 2008.
- [35] J. Grzyb, K. Statnikov, N. Sarmah, B. Heinemann, U. R. Pfeiffer, "A 210–270-GHz circularly polarized FMCW radar with a single-lens-coupled SiGe HBT chip," *IEEE Transactions on Terahertz Science and Technology*, vol. 6, pp. 771-783, 2016.

- [36] A. Nikpaik, A. Masnadi Shirazi, A. Nabavi, S. Mirabbasi, and S. Shekhar, "A 218-to-231GHz Frequency-Multiplier-Based VCO with ~3% peak DC-to-RF Efficiency in 65nm CMOS," *IEEE Journal of Solid-States Circuits (JSSC)*, vol. 53, pp. 389-403, 2017.
- [37] S. Rong, H.C. Luong, "Design and analysis of varactor-less interpolative-phase-tuning millimeter-wave LC oscillators with multiphase outputs," *IEEE J. Solid-State Circuits*, vol. 46, no. 8, pp. 1810-1819, 2011.
- [38] B. Banerjee, S. Venkataraman, Y. Lu, Q. Liang, C. Lee, S. Nuttinck, D. Heo, Y. Emery Chen, J. D. Cressler, J. Laskar, G. Freeman, and D. C. Ahlgren, "Cryogenic Operation of Third-Generation, 200-GHz Peak-fT, Silicon-Germanium Heterojunction Bipolar Transistor," *IEEE Transaction on Electron Devices*, vol. 52, no. 4, pp. 585-593, April 2005.
- [39] C. Degen, M. Poggio, H. Mamin, C. Rettner, and D. Rugar, "Nanoscale magnetic resonance imaging," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106, pp 1313-1317, Feb 2009.
- [40] E. W. Moore, S. Lee, S. A. Hickman, S. J. Wright, L. E. Harrell, P. P. Borbat, J. H. Freed, and J. A. Marohn, "Scanned-Probe Detection of Electron Spin Resonance from a Nitroxide Spin Probe," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106(52), pp. 22251 – 22256, Nov 2009.
- [41] C. E. Isaac C. M. Gleave, P. T. Nasr, H. L. Nguyen, E. A. Curley, J. L. Yoder, E. W. Moore, L. Chena, and J. A. Marohn, "Dynamic Nuclear Polarization in a Magnetic Resonance Force Microscope Experiment," *Phys. Chem. Chem. Phys.*, vol. 18, pp. 8806 8819, April 2016.

- [42] E. Ohmichi, T. Okamoto, M. Mitani, H. Takahashi, and H. Ohta, "High-frequency Electron Paramagnetic Resonance of Metal-containing Porphyrin Compounds Using a Microcantilever," *J. Inorg. Biochem.*, vol. 162, pp. 190-193, Sep 2016.
- [43] K. R. Thurber, W.-M. Yau, and R. Tycko, "Low-temperature Dynamic Nuclear Polarization at 9.4 T with a 30 mW Microwave Source," *J. Magn. Res.* vol. 204, pp. 303-313, June 2010.
- [44] T. A. Siaw, S. A. Walker, B. D. Armstrong, and S.-I. Han, "Inductively Coupled NMR Probe for Versatile Dynamic Nuclear Polarization Operation at 7 T: Observation of 61±2% 1H Polarization at 4 K," *J. Magn. Res.* vol. 221, pp. 5-10, Aug 2012.
- [45] P. A. S. Cruickshank, and G. M. Smith, "Force Detected Electron Spin Resonance at 94
   GHz," *Rev. Sci. Instrum.* vol. 78, pp. 015101, Jan 2007.
- [46] D. Ha, J. Paulsen, N. Sun, Y.-Q. Song, and D. Ham, "Scalable NMR Spectroscopy With Semiconductor Chips," *Proc. Natl. Acad. Sci.* vol. 111, pp. 11955-11960, Aug 2014.
- [47] J. Anders, J. Handwerker, M. Ortmanns, and G. Boero. "A Fully-integrated Detector for NMR Microscopy in 0.13 μm CMOS," 2013 IEEE Asian Solid-State Circuits Conference (A-SSCC), Singapore, 2013, pp. 437-440.
- [48] J. Anders, J. Handwerker, M. Ortmanns, and G. Boero, "A Low-power High-sensitivity Single-chip Receiver for NMR Microscopy," J. Magn. Reson. Vol. 266, pp. 41-50, May 2016.
- [49] T. Yalcin and G. Boero, "Single-chip Detector for Electron Spin Resonance Spectroscopy," *Rev. Sci. Instrum.* vol. 79, 094105 (2008).

- [50] J. Anders, A. Angerhofer, and G. Boero, "K-band Single-chip Electron Spin Resonance Detector," J. Magn. Reson. vol. 217, pp. 19-26, April 2012.
- [51] A. V. Matheoud, G. Gualco, M. Jeong, I. Zivkovic, J. Brugger, H. Ronnow, J. Anders, and G. Boero, "Single-chip Electron Spin Resonance Detectors Operating at 50 GHz, 92 GHz, and 146 GHz," *J. Magn. Reson.* vol. 278, pp. 113-121, May 2017.
- [52] G. Gualco, J. Anders, A. Sienkiewicz, S. Alberti, L. Forr, and G. Boero, "Cryogenic Singlechip Electron Spin Resonance Detector," *J. Magn. Reson.* vol. 247, pp. 96-103, Oct 2014.
- [53] J. Handwerker, B. Schlecker, U. Wachter, P. Radermacher, M. Ortmanns, and J. Anders, "28.2 A 14GHz Battery-operated Point-of-care ESR Spectrometer Based on a 0.13 μm CMOS ASIC," 2016 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, 2016, pp. 476-477.
- [54] J. Yuan, J. D. Cressler, C. Zhu, Y. Cui, G. Niu, Q. Liang, A. J. Joseph, "An Investigation of Negative Differential Resistance and Novel Collector–Current Kink Effects in SiGe HBTs Operating at Cryogenic Temperatures," *in IEEE Transactions on Electron Devices*, vol. 54, no. 3, pp. 504-516, March 2007.
- [55] J. Yuan, J.D. Cressler, R. Krithivasan, T. Thrivikraman, M.H. Khater, D.C. Ahlgren, A.J. Joseph, and J.-S. Rieh, "On the Performance Limits of Cryogenically-operated SiGe HBTs and Its Relation to Scaling for TeraHertz Speeds," *IEEE Transactions on Electron Devices*, vol. 56, no. 5, pp. 1007-1019, May 2009.
- [56] B. Patra, R. Incandela, J. P. G. van Dijk, H. A. R. Homulle, L. Song, M. Shahmohammadi,R. B. Staszewski, A. Veladimirescu, M. Babaie, F. Sebastiano, E. Charbon, "Cryo-CMOS

Circuits and Systems for Quantum Computing Applications," *in IEEE Journal of Solid-State Circuits*, vol. 53, no. 1, pp. 309-321, Jan. 2018.

- [57] R. Krithivasan, Y. Lu, L. Najafizadeh, C. Zhu, J. D. Cressler, S. Chen, C. Ulaganathan, B. Blalock, "A High-slew Rate SiGe BiCMOS Operational Amplifier for Operation down to Deep Cryogenic Temperatures," *Proc. IEEE BCTM*, pp. 72-75, 2006.
- [58] D. Ma, X. Geng, F. Dai, J. D. Cressler, "A 6th Order Butterworth SC Low Pass Filter for Cryogenic Applications -180 °C to 120 °C," *Proc. IEEE Aerosp.* Conf., pp. 1-8, 2009.
- [61] P. Chevalier, N. Zerounian, B. Barbalat, F. Aniel, and A. Chantre, "On the Use of Cryogenic Measurements to Investigate the Potential of Si/SiGe:C HBTs for Terahertz Operation," 2007 IEEE Bipolar/BiCMOS Circuits and Technology Meeting, Boston, MA, 2007, pp. 26-29.
- [62] P. Chevalier, G. Avenier, E. Canderle, A. Montagné, G. Ribes and V. T. Vu, "Nanoscale SiGe BiCMOS Technologies: From 55 nm Reality to 14 nm Opportunities and Challenges," 2015 IEEE Bipolar/BiCMOS Circuits and Technology Meeting - BCTM, Boston, MA, 2015, pp. 80-87.
- [63] H. Khatibi, S. Khiyabani and E. Afshari, "An Efficient High-Power Fundamental Oscillator Above f<sub>max</sub>/2: A Systematic Design," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 11, pp. 4176-4189, Nov. 2017.
- [64] Jae-Sung Rieh, D. Greenberg, A. Stricker and G. Freeman, "Scaling of SiGe Heterojunction Bipolar Transistors," in *Proceedings of the IEEE*, vol. 93, no. 9, pp. 1522-1538, Sept. 2005.

- [65] E. Juntunen, D. Dawn, J. Laskar, and J. Papapolymerou, "High-Power, High-Efficiency CMOS Millimeter-Wave Oscillators", *IET Microw. Antennas Propag.*, vol. 6 iss. 10, pp 1158-1163, 2012.
- [66] Che-Chen Lee, Shu-Van Huang, and Hong-Y eh Chang, "A 44-49 GHz Low Phase Noise CMOS Voltage-Controlled Oscillator with 10-dBm Output Power and 16.1 % Efficiency," *Microwave Symposium (IMS), 2014 IEEE MTT-S International*, pp. 1-4, Jun. 2014.
- [67] K.-J. Kim, T.-H. Lim, H.-C. Park, and K.H. Ahn, "mm-Wave CMOS Colpitts VCO and frequency divider for the 60GHz WPAN," *IEEE Int. Conf on Communications*, Dec. 2009, pp. 358-36.
- [68] B. A. Floyd, S. K. Reynolds, U. R. Pfeiffer, T. Zwick, T. Beukema and B. Gaucher, "SiGe bipolar transceiver circuits operating at 60 GHz," *in IEEE Journal of Solid-State Circuits*, vol. 40, no. 1, pp. 156-167, Jan. 2005.
- [69] R. Groves, D. L. Harame and D. Jadus, "Temperature Dependence of Q and Inductance in Spiral Inductors Fabricated in a Silicon-Germanium/BiCMOS Technology," *in IEEE Journal of Solid-State Circuits*, vol. 32, no. 9, pp. 1455-1459, Sep 1997.
- [70] C. H. Doan, S. Emami, A. M. Niknejad and R. W. Brodersen, "Millimeter-wave CMOS Design," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 1, pp. 144-155, Jan. 2005.
- [71] S. Jameson and E. Socher, "A 93.9 102.5 GHz Colpitts VCO Utilizing Magnetic Coupling Band Switching in 65nm CMOS," *IEEE International Conference on Microwaves, Communications, Antennas and Electronic Systems (COMCAS)*, Tel-Aviv, 2015.

- [72] M. Adnan and E. Afshari, "A 105-GHz VCO With 9.5% Tuning Range and 2.8-mW Peak Output Power in a 65-nm Bulk CMOS Process," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 4, pp. 753-762, April 2014.
- [73] R. Han and E. Afshari, "A CMOS High-Power Broadband 260-GHz Radiator Array for Spectroscopy," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 12, pp. 3090-3104, Dec. 2013.
- [74] D. J. Sharma and H. Krishnaswamy, "216- and 316-GHz 45-nm SOI CMOS Signal Sources Based on a Maximum-Gain Ring Oscillator Topology," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 1, pp. 492–504, Jan. 2013.
- [75] M. Kraemer, D. Dragomirescu and R. Plana, "A High Efficiency Differential 60 GHz VCO in a 65 nm CMOS Technology for WSN Applications," *IEEE Microwave and Wireless Components Letters*, vol. 21, no. 6, pp. 314-316, June 2011.
- [76] N. O. Sokal and A. D. Sokal, "Class E-A New Class of High-efficiency Tuned Singleended Switching Power Amplifiers," *IEEE Journal of Solid-State Circuits*, vol. 10, no. 3, pp. 168-176, June 1975.
- [77] A. Chakrabarti and H. Krishnaswamy, "High Power, High Efficiency Stacked mmWave Class-E-like Power Amplifiers in 45nm SOI CMOS," *Proceedings of the IEEE Custom Integrated Circuits Conference*, San Jose, CA, 2012.
- [78] M. Garampazzi, P. M. Mendes, N. Codega, D. Manstretta, and R. Castello, "Analysis and Design of a 195.6 dBc/Hz Peak FoM p-n Class B Oscillator with Transformer-Based Tail Filtering," *in IEEE J. Solid-State Circuits*, vol. 50, no. 7, pp. 1657-1668, 2015.

- [79] A. Mazzanti and P. Andreani, "Class-C Harmonic CMOS VCOs, with a General Result on Phase Noise," *in IEEE J. Solid-State Circuits*, vol. 43, no. 12, pp. 2716-2729, 2008.
- [80] C. B. Soltanian and P. Kinget, "Tail Current-Shaping to Improve Phase Noise in LC Voltage-Controlled Oscillators," *IEEE J. Solid-State Circuits*, vol. 41, no. 8, pp. 1792-1802, 2006.
- [81] E. Socher, S. Jameson, "Wide Tuning Range W-band Colpitts VCO in 90 nm CMOS," *Electronics Letters*, vol. 47, no. 22, pp. 1227-1229, October 2011.
- [82] N. Zhang and K. K. O, "94 GHz Voltage Controlled Oscillator with 5.8% Tuning Range in Bulk CMOS," *IEEE Microwave and Wireless Components Letter*, vol. 18, no. 8, pp. 548-550, Aug. 2008.
- [83] J. Lee, et al., "A fully-integrated 77-GHz FMCW radar transceiver in 65-nm CMOS technology," JSSC, vol. 45, pp. 2746 –2756, Dec. 2010.
- [84] A. Tomkins, et al., "A passive W-band imaging receiver in 65-nm bulk CMOS," JSSC, vol. 45, no. 10, pp. 1981 –1991, Oct. 2010.
- [85] S. Emami, et al., "A 60 GHz CMOS phased-array transceiver pair for multi-Gb/s wireless communications," *ISSCC*, Feb. 2011, pp. 164–166.
- [86] A. Tarkeshdouz, et al., "A 91-GHz fundamental VCO with 6.1% DC-to-RF efficiency and
  4.5 dBm output power in 0.13-μm CMOS," *SSC-L*, vol. 1, no. 4, pp. 102-105, April 2018.
- [87] S. Kang, et al., "A 100GHz active-varactor VCO and a bi-directionally injection-locked loop in 65nm CMOS," *RFIC*, pp. 231-234, June 2013.

- [88] H. Aghasi, et al., "An 88-GHz Compact Fundamental Oscillator With 19.4% DC-to-RF Efficiency and 7.5-dBm Output Power in 130-nm SiGe BiCMOS," SSC-L, vol. 1, no. 5, pp. 106-109, May 2018.
- [89] R. Weber, et al., "A 92 GHz GaN HEMT voltage-controlled oscillator MMIC," *IMS*, pp. 1-4, June 2014.
- [90] P. Agarwal et al., "Switched substrate-shield-based low-loss CMOS inductors for wide tuning range VCOs," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 8, pp. 2964–2976, Aug. 2017.
- [91] W. Fei, H. Yu, H. Fu, J. Ren, and K. S. Yeo, "Design and analysis of wide frequencytuning-range CMOS 60 GHz VCO by switching inductor loaded transformer," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 3, pp. 699–711, Mar. 2014.
- [92] Q. Zou, K. Ma, and K. S. Yeo, "A low phase noise and wide tuning range millimeter-wave VCO using switchable coupled VCO-cores," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 62, no. 2, pp. 554–563, Feb. 2015.
- [93] J. Zhang, N. Sharma, and K. K. O, "21.5-to-33.4 GHz voltage-controlled oscillator using NMOS switched inductors in CMOS," *IEEE Microw. Compon. Lett.*, vol. 24, no. 7, pp. 478–480, Jul. 2014.
- [94] M. Demirkan, S. P. Bruss, and R. R. Spencer, "Design of wide tuning-range CMOS VCOs using switched coupled-inductors," *IEEE J. Solid-State Circuits*, vol. 43, no. 5, pp. 1156– 1163, May 2008.
- [95] M. Haghi Kashani, A. Tarkeshdouz, R. Molavi, E. Afshari and S. Mirabbasi, "A Wide-Tuning-Range Low-Phase-Noise mm-Wave CMOS VCO With Switchable Transformer-Based Tank," in *IEEE Solid-State Circuits Letters*, vol. 1, no. 4, pp. 82-85, April 2018.
- [96] H. Tsuji et al., "A study of flicker noise suppression of K-band VCO using striped inductor," *in Proc. Asia–Pac. Microw. Conf. (APMC)*, Nanjing, China, Nov. 2015, pp. 1–3.
- [97] G. Liu, R. Berenguer, and Y. Xu, "A MM-Wave configurable VCO using MCPW-based tunable inductor in 65-nm CMOS," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 58, no. 12, pp. 842–846, Dec. 2011.
- [98] N. Yanay and E. Socher, "Wide tuning-range mm-wave voltage-controlled oscillator employing an artificial magnetic transmission line," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 4, pp. 1342–1352, Apr. 2015.
- [99] C. H. Hung and R. Gharpurey, "A 57-to-75 GHz dual-mode wide-band reconfigurable oscillator in 65nm CMOS," *in Proc. IEEE Radio Freq. Integr. Circuits Symp.*, Tampa, FL, USA, Jun. 2014, pp. 261–264.
- [100] J. D. Cressler et al., "On the profile design and optimization of epitaxial Si- and SiGe-base bipolar technology for 77 K applications. I. Transistor DC design considerations," *in IEEE Transactions on Electron Devices*, vol. 40, no. 3, pp. 525-541, March 1993.
- [101] S. K. Mandal, G. K. Marskole, K. S. Chari and C. K. Maiti, "Transit time components of a SiGe-HBT at low temperature," 2004 24th International Conference on Microelectronics (IEEE Cat. No.04TH8716), Nis, Serbia, 2004, pp. 315-318 vol.1.

- [102] J. D. Cressler and G. Niu, *Silicon-Germanium Heterojunction Bipolar Transistors*. Boston, MA: Artech House, 2003.
- [103] M. Rodwell et al., "On the Feasibility of few-THz Bipolar Transistors," 2007 IEEE Bipolar/BiCMOS Circuits and Technology Meeting, Boston, MA, 2007, pp. 17-21.
- [104] E. A. Guttierez, M. J. Dean, C. L. Claeys, Low Temperature Electronics: Physics, Devices, Circuits, and Applications. London, U.K.: Academic, 2001.
- [105] D. L. Harame *et al.*, "Si/SiGe epitaxial-base transistors. I. Materials, physics, and circuits," in *IEEE Transactions on Electron Devices*, vol. 42, no. 3, pp. 455-468, March 1995.
- [106] P. S. Chakraborty et al., "A 0.8 THz f<sub>max</sub> SiGe HBT Operating at 4.3 K," in *IEEE Electron Device Letters*, vol. 35, no. 2, pp. 151-153, Feb. 2014.
- [107] R. Krithivasan et al., "Half-terahertz operation of SiGe HBTs," in IEEE Electron Device Letters, vol. 27, no. 7, pp. 567-569, July 2006.
- [108] Banerjee et al., "Cryogenic performance of a 200 GHz SiGe HBT technology," 2003 Proceedings of the Bipolar/BiCMOS Circuits and Technology Meeting (IEEE Cat. No.03CH37440), Toulouse, France, 2003, pp. 171-173.
- [109] B. Geynet *et al.*, "SiGe HBTs featuring f<sub>T</sub>≫400GHz at room temperature," 2008 IEEE Bipolar/BiCMOS Circuits and Technology Meeting, Monteray, CA, 2008, pp. 121-124.
- [110] H. Kroemer, "Two integral relations pertaining to the electron transport through a bipolar transistor with a nonuniform energy gap in the base region," *Solid State Electron*, vol.28, pp.1101-1103, Nov. 1985.

## Appendix

## A.1 Magnetic Resonance Force Microscopy Imaging (MRFM)

The ability to characterize organic nanostructures could have a dramatic impact on basic research in materials design, biomolecular power and energy generation, biosensing, virology, and regenerative medicine. It is required to develop a procedure to image individual molecules and biomolecules such as proteins complexes at nanometer or sub-nanometer resolutions. For instance, proteins are the worker molecules that are necessary for every activity in the body. Hence, studying the shapes of such structures reveals the proteins functionality in our bodies and helps one understand diseases by observing abnormality in proteins' shapes. Magnetic resonance imaging (MRI) is a medical imaging technique used in radiology to image the anatomy and physiological processes of the human body at nanometer scales. There are biological molecules imaging techniques such as x-ray crystallography, nuclear magnetic resonance (NMR), cryo-electron microscopy (cryo-EM), electron spin resonance (ESR), and magnetic resonance force microscopy (MRFM) among which the last method (i.e., MRFM) is the most promising technique for threedimensional (3D) of disordered molecules in situ. Bing capable to yield resolutions as low as 500 protons compared to an improved-MRI imaging method proves the superiority of such an imaging method. This resolution is translated to  $10^{10}$  to  $10^{17}$  improvement in sensitivity. The MRFM imaging combines the ideas of MRI and atomic force microscopy (AFM), using cantilever-tipped with a ferromagnetic particle to directly detect a modulated spin gradient force between sample spins and the tip. As the cantilever tip is brought close to the sample, the atoms' nuclear spins response and generate a small force on the cantilever. The spins are then repeatedly flipped,

causing the cantilever to gently move back and forth in a synchronous motion. The displacement then is measured to generate images.

To help attain a further enhanced sensitivity, the dynamic nuclear polarization effect (DNP) is introduced. Dynamic nuclear polarization (DNP) is a widely-used approach for increasing the signal-to-noise ratio in conventional magnetic resonance experiments. DNP achieves an improved signal-to-noise ratio by increasing the polarization — e.g., the net alignment — of the nuclear spins in the sample. In this method, the microwave irradiation of free-radical polarizing agents is utilized to create hyper-polarized nuclear spins in a cryogenic MRFM experiment. DNP results from transferring spin polarization from electrons to nuclei, thereby aligning the nuclear spins to the extent that electron spins are aligned.



Figure A.1 The details and the primary principals of an MRFM imaging experiment

## A.2. A Semi-Quantitative Mapping Between Cooling and Device Scaling

Cryogenic SiGe HBTs are intriguing for a number of cryogenic applications as mentioned earlier. This includes modern space systems (e.g., the equipment needed for earth orbit or outer planets and their moons) where the extreme cold ambient temperatures are encountered. Low temperature condition can be highly detrimental to conventional electronics. A wide range of space electronics are currently being deployed to operate at very low temperature ambient environments utilizing bulky and power-hungry heating units. It has been demonstrated that operation at cold environments is a natural fit for SiGe transistors since the performance is improved with cooling. This bandgap-engineered platform enables cryogenic electronics tailored for space missions without the required heating tools. Pioneering works on cryogenic SiGe HBTs demonstrate an enhanced performance at low temperatures [33]. The level of cooling-induced performance improvements, nonetheless, were not remarkable in first-generation SiGe HBTs [100]. In contrast, in scaled modern SiGe HBTs, with relatively generous bandgap grading in advanced generations and increased level of base doping, the improved performance level observed is tremendous [33]. In this context, the effect of cooling on SiGe HBTs in many aspects is similar to the combined effect of vertical and lateral device scaling [55]. Moreover, cooling helps identify the key optimization issues required for the new scaling rules and roadmap [55].

It is well known that a tuned reduction in both vertical and lateral dimensions of the device is a key for an optimized scaling in SiGe HBTs. Vertical scaling is accomplished by reducing the base width (W<sub>B</sub>) of the transistor while base and collector doping (N<sub>B</sub>, N<sub>C</sub>) are increased to lower the base and collector transit times ( $\tau_b$ ,  $\tau_c$ , and  $\tau_{bc}$ ). As a result of vertical scaling, the transistor's current gain ( $\beta$ ) is augmented. Vertical scaling, also, influences  $f_T$  if an HBT by effecting the components forming the total emitter-to-collector transit time,  $\tau_{ec}$ . These components are shown in Fig. A.2.

In emitter-to-collector transit time calculations, it is costumery to drop  $\tau_{eb}$  as it is very hard to decouple this component from  $\tau_b$  and  $\tau_e$  [55]. In this scheme, thus,  $f_T$  of a SiGe HBT is primarily governed by the base minority carrier storage time ( $\tau_b$ ) [101].

131



Figure A.2 Transit time components versus the vertical profile of a bipolar transistor.

The emitter minority carrier storage charge and the base-collector space charge region (SCR) transit times are the next two dominant components. On the downside, in vertical scaling, the collector-emitter breakdown voltage ( $BV_{CEO}$ ) is degraded as a result of increasing N<sub>C</sub> [102].

In addition, vertical scaling introduces large device parasitics such as base resistance and BC junction capacitance, if it is not associated with lateral scaling. Even though vertical scaling does not deteriorate  $f_T$  necessarily, it directly degrades  $f_{max}$ . From a circuit perspective, however, it is ideal to have a more balanced set of  $f_T$  and  $f_{max}$  for transistors for mm-wave amplification [103]. In this context, lateral scaling is prescribed to compensate for the degradations resulting from vertical scaling to lower parasitics. Lateral scaling per se is associated with additional struggles to reduce  $R_E$  and  $R_C$ . This entails a fundamental trade-off which obliges one to employ a more balanced approach for vertical and lateral scaling. Recently, raised extrinsic base technique and the selectivity implanted collector method have been introduced to alleviate the trade-off. This has resulted in more balanced  $f_T$  and  $f_{max}$  in advanced modern SiGe HBT technologies.

In SiGe HBTs, the presence of Ge grading in the base introduces the similar desired effects of vertical profile scaling. In these devices, as shown in Fig. A.3, Ge lowers the conduction-band barrier by  $\Delta E_{g,Ge}(0)$  at the emitter side and enhances the electron injection toward the collector.

In the context of SiGe platform, one needs to manipulate  $\Delta E_{g,Ge}(0)/kT$  and  $\Delta E_{g,Ge}(grade)/kT$  to improve ac and dc performance characteristics of the device, as will be shown later [55].



Fig. A.3 Si and SiGe HBT schematic band diagram in the presence of Ge grading.

As opposed to scaling, cooling effectively reduces the primary performance-limiting parasities of SiGe HBTs without having to physically scaling the lateral dimensions. This is accomplished by introducing modifications and improvements to the base-emitter transit components such as  $\tau_{bc}$ . The mechanisms by which cooling reduces the charging times and the parasities can be explained from the fundamentals of physics. For instance, the reduction in C<sub>BC</sub> is resulted from the following mechanism. The fermi level of the n-type collector region moves closer to the conduction band due to cooling [104]. This introduces an additional potential across the p-n junction and widens the space charge region with a fixed bias and accordingly the carrier movements are accelerated. As for the base, the increased base doping level can be associated with a higher chance of freeze-out. However, increasing the doping level beyond the Mott transition to degenerate the semiconductor guarantees that the dopants remain active even at deep cryogenic temperatures. This, indeed, suggest that R<sub>B</sub> is reduced with decreasing the temperature. In addition,  $\tau_{be}$  is

reduced through the changes introduced by cooling in space charge region in BE junction. This is achieved as follow; in advanced generation SiGe HBTs, the semiconductors are heavily degenerated and hence, the fermi levels are close to the conduction band in the emitter and the valence band in the base. At deep cryogenic temperatures, a higher  $V_{BE}$  required for operation effectively reduces the BE junction width. Therefore,  $C_{BE}$  is increased and the electron transit time through the BE SCR is decreased.

In general, cooling bolds the favorable effects of bandgap engineering [55]. This simply comes from the positive role of kT in the device performance and the increase in carrier mobility at low temperatures [33], [105]. As pointed out, the consequences of cooling on SiGe HBTs in many ways match that of the resulting from a coordinated lateral and vertical scaling. There are extensive numbers of works that characterize the ac and dc performance metrics of SiGe HBTs across a wide range of temperatures to explore similarities between cooling and scaling [106–109]. For instance, in [55],  $f_T$  and  $f_{max}$  of a second-generation SiGe HBT is presented at 300 K and 93 K. The  $f_T/f_{max}$  of a cooled device increase from 117/124 GHz at 300 K to 182/295 GHz at 93 K which is comparable to the room-temperature performance of a scaled third-generation SiGe HBT with  $f_T/f_{max}$  of 207/285 GHz at 300 K. Likewise, a  $f_T/f_{max}$  of 260/310 GHz is achieved by cooling to 85 K which is close to the performance of a room-temperature fourth generation SiGe HBT. Hence, it is logical to establish a semi-quantitative mapping between cooling and scaling by further investigations. In [62], the evolution of the  $f_T$  is presented as a function of the transistor collector current density for a number of high-speed SiGe technologies. If  $f_T$  is plotted versus  $1/J_C$  the graphs in Fig. A.4 are obtained.



Fig. A.4 (a) Evolution of the current gain cut-off frequency vs. the inverse of collector current density in STMicroelectronics SiGe BiCMOS technologies. The plot is reproduced from the data in [14]. (b) The peak  $f_T$  values of the plot in (a) versus the minimum emitter width of the transistors.

Fig. A.4 (a) shows the high-frequency performance of SiGe HBTs versus 1/J<sub>C</sub> values. The dotted line in this figure indicates the trend by which the peak values of  $f_T$  ascend as the technology nodes evolve. The trajectory of peak  $f_T$  well scales with 1/W<sub>E</sub>. The similar trend is reported in [64] for the current density at peak  $f_T$  over the emitter width with a scaling factor of  $\lambda^{0.97}$ . Fig. A.4(b) plots the peak  $f_T$  for STMicroelectronics high-speed SiGe BiCMOS technologies versus the minimum corresponding emitter widths. This exhibits that peak  $f_T$  varies approximately with  $\frac{c}{W_E}$  (which is similar to  $\lambda^{0.97}$ ) where c is a constant and the upper bound and lower bound are  $c'(\frac{1}{W_E} - \frac{1}{W_E^2})$  and  $c''(\frac{1}{W_E^2} - \frac{1}{W_E^3})$ , respectively. The trends is almost analogous to the high-frequency performance characteristics of CMOS transistors as a function of gate length (L<sub>G</sub>) [62]. To obtain an expression for the collector current density in a SiGe HBT as a function of temperature, using the generalized Moll-Ross relation in [110], we write  $J_{c,SiGe}$  as in [105],

$$J_{c,SiGe} = \frac{q n_{io}^2 D_{nb}}{W_b N_b} \left( e^{-\frac{\Delta E_{gb} a^{app}}{kT}} \cdot e^{\frac{q V_{BE}}{kT}} - 1 \right)$$

$$\times \left[ \gamma \eta \frac{\Delta E_{g,Ge}(grade)}{kT} \frac{e^{\Delta E_{g,Ge}(0)/kT}}{1 - e^{-\Delta E_{g,Ge}(grade)/kT}} \right]$$
(A.1)

where  $\Delta E_{g,Ge}(0)$  and  $\Delta E_{g,Ge}(grade)$  are shown in Fig. A.3. In (A.1),  $W_b$  is the quasi-neutral base width at  $V_{BE}$ ,  $D_{nb}$  is the minority electron diffusivity in the base,  $n_{io}^2$  is the effective intrinsic carrier density,  $N_b$  is a constant base doping profile,  $\gamma$  accounts for the reduction in the effective densityof-states with increasing Ge content,  $\eta$  is the strain-enhancement of the minority electron mobility, and k is the Boltzmann constant. Concerning the improvement in the ac performance of SiGe HBTs at cryogenic conditions, we investigate the variations of transit frequency as a function of temperature. A simplistic equation for  $f_T$  is presented below,

$$f_T = \frac{1}{2\pi} \left\{ \frac{kT}{qI_c} (C_{EB} + C_{CB}) + \tau_b + \tau_e + \tau_{bc} \right\}^{-1}$$
(A.2)

In the above equation,  $\tau_b$ ,  $\tau_e$ , and  $\tau_c$  are the base, emitter, and collector transit time, respectively, and other parameters are usual HBT parasitics. In (2), term  $\frac{kT}{qI_c}(C_{EB} + C_{CB})$  encapsulate the effect of  $\tau_{bc} + \tau_{ec}$ . As outlined earlier, we expect the base transit time to be the dominant components in  $f_T$  equation. This plus the summation of  $\tau_{bc}$  and  $\tau_{ec}$  would be well enough to explore how  $f_T$ changes with variations in ambient temperature to the first-order. Theoretically, the base transit time for a SiGe HBT with a constant base doping written as,

$$\tau_{b,SiGe} = \frac{W_b^2}{D_{nb}} \frac{kT}{\Delta E_{g,Ge}(grade)} \times \left[ 1 - \frac{kT}{\Delta E_{g,Ge}(grade)} (1 - e^{-\Delta E_{g,Ge}(grade)/kT}) \right]$$
(A.3)

At low temperatures, (A.3) becomes,

136

$$\tau_{b,SiGe} \approx \frac{W_b^2}{D_{nb}} \frac{kT}{\Delta E_{g,Ge}(grade)}$$
(A.4)

In (A.4), the minority electron diffusivity in the base is proportional to  $aT^b$  where b is a number between -1 and -2. Thus,  $\tau_{b,SiGe}$  is proportional to temperature as follow,

$$\tau_{b,SiGe} \propto k' T^{b+1} \tag{A.5}$$

Assuming a typical b = -1.6 gives  $\tau_{b,SiGe} \propto k'T^{-0.8}$ .

On the other hand,  $\frac{kT}{qI_c}(C_{EB} + C_{CB})$  can be approximated by  $\frac{g_m}{c_D}$  where  $g_m$  is the transistor transconductance and  $C_D$  is  $\sim C_{EB} + C_{CB}$ . For an HBT, if we assume that that modulations do not impact the width of the base significantly,

$$g_{m,SiGe} = \frac{\partial J_c}{\partial V_{BE}}|_{V_{BE}} = \left(\frac{\partial J_c}{\partial W_b}\right) \left(\frac{\partial W_b}{\partial V_{BE}}\right)|_{V_{BE}} \approx \alpha \frac{\partial J_c}{\partial V_{BE}}$$
(A.6)

In (A.1), assuming that  $n_{io}^2 D_{nb}$  does not change by temperature, we can conclude that,

$$J_{c,SiGe} \propto \frac{e^{\frac{-1}{kT}}}{kT}$$

Using the above relation, we conclude that  $g_{m,SiGe}$  is proportional to  $e^{\frac{-1}{kT}}/(kT)^2$ . If we assume that changes in  $C_D$  is not extensive due to cooling,

$$\frac{g_m}{C_D} = \frac{J_C}{V_T \cdot C_D} \propto \frac{e^{\frac{-1}{kT}}}{(kT)^2} \tag{A.7}$$

Hence,  $f_T$  in (A.2) becomes,

$$f_T \approx \frac{1}{2\pi} \left\{ \frac{kT}{qI_c} (C_{EB} + C_{CB}) + \tau_b \right\}^{-1} \propto k' T^{-0.8} + \frac{e^{\frac{-1}{kT}}}{(kT)^2}$$
(A.8)

137

The term on the right-hand side in (A.8) can be written as,

$$\approx \frac{e^{\frac{-1}{T}}}{T^2} = \frac{1}{T} \left( 1 - \frac{1}{T} + \cdots \right) \approx \frac{1}{T} - \frac{1}{T^2}$$

The above trend exactly follows that of the upper-limit presented in Fig. A.4(b). That is, if T is substituted by  $W_E$  with proper corresponding coefficients, it approaches the limit in the plot. The entire process shows how the above equations can be exploited, at least to the first-order, to predict and obtain the ac and dc performance characteristics of SiGe HBTs at low temperatures.

Nevertheless, there are metrics that their variations and values at low temperatures cannot be well approximated even to the first-order using the above equations. For instance,

$$f_{max} = \left\{ \frac{f_T}{8\pi C_C R_b} \right\}^{1/2}$$
(A.9)

where  $C_c$  is approximately equal to  $C_{CB}$  and  $R_b$  is the base resistance. In (A.9),  $R_b$  benefits from cooling, for instance. However, the extent of variations cannot be well formulated as a function of temperature. Figure A.5 is reproduced using the published  $f_{max}$  data for a large set of SiGe HBTs with the corresponding cooled  $f_{max}$  values. As observed, the general trend for  $f_{max}$  analogously agrees with the ones discovered above. Therefore, given the mapping between scaling and cooling developed so far and the resembling matched trends, we need to resort to scaling onward to proceed with the circuit design.



Fig A.5 Published  $f_{max}$  of SiGe HBTs with the corresponding cooled prototypes versus temperature.

## A.3. Semi-Quantitative Temperature Prediction

In all reported cooled SiGe HBT characterizations to date, the improvements introduced in device metrics from a technology node to the next (~40%) are equivalent to that of induced by cooling from room temperature, i.e., 330 K, up down to almost liquid nitrogen temperature, i.e., ~77 K [33], [55], [106–109]. Any further reduction in temperature cannot be captured by scaling from a technology node to the next more advanced one and requires further scaling. Extensive prior studies show cooling down to temperatures as low as ~4 K is equivalent to two generation scaling pertaining to the SiGe HBTs device metrics [106-109]. Hence, with only two technology nodes at hand, any further cooling beyond almost 77 K necessitates employing the first-order SiGe HBTs model to compensate for the lack of more advanced technology nodes. Thermo-dynamical computations show that ambient temperature over the area very close to the IC in this work is higher than that of actually are being measured at the cryogenic chunk (the cryogenic measurement

procedure is explained in the subsequent sections). With the chunk's temperature reading T=12 K, the I<sub>C</sub> is anticipated to be operating at ~45 K.

The commercial transistor models incorporated in the process design kits are customarily assembled to be valid for a temperature range where most practical applications are functional. In regard to low-temperature applications, the models are valid within the range of -45 °C < T(°C) < 80 °C corresponding to 228 K < T(K) < 353 K with T=298 K being the room temperature. As said earlier, any further reduction in temperature beyond 77 K cannot be captured by the advanced technology node in use. Therefore, using the proportionalities in (A.7) and (A.8), for instance, we can, with the a good approximation, compute the relative changes in ac parameters from 75 K to 45 K (the operating temperature of the IC), e.g.,  $f_T$  and  $g_m$ , utilizing the corresponding changes resulted by cooling the transistor form room temperature, 298 K, to the temperature where the same relative changes in the parameters of interest are achieved. In such a case, if *T* happens to be greater than 228 K, then the PDK models can well simulate for the additional temperature decrement down to 45 K. If the first term in (A.8) is ignored for simplicity,

$$\left(\frac{e^{\frac{-1}{228}}}{228^2} - \frac{e^{\frac{-1}{298}}}{298^2}\right) / \left(\frac{e^{\frac{-1}{45}}}{45^2} - \frac{e^{\frac{-1}{75}}}{75^2}\right) \approx 0.6$$
(A.10)

In (A.10), any ratio smaller than 1 suggests that the low temperature models provided by the PDK are not sufficient. The above equation indicates that a change in temperature ( $\Delta T$ ) of ~140° is required (from room temperature) to fully compensate for the lack of transistor models and the employed scaled technology at *T*=45 K. Differently stated, setting the simulation temperature to be at -45 °C, we can confidently predict and simulate the performance parameters, i.e., the circuit

block operation, up down to 55 K. The improvement trend in ac and dc parameters onward is predicted to adhere to the corresponding above equations yet.

Remarks: to be more rigorous,  $\Delta T$  should be estimated for every one of the device attributes as the proportionality relations are different according to (A.5) to (A.8). For example, a  $\Delta T$ =140° may model the performance up down to 55 K in terms of  $f_T$  while it does so down to 50 K with respect to  $g_m$ . In the above estimation in (A.10), the changes are calculated for the worst case of  $f_T$ ,  $f_{max}$ , and  $g_m$  (i.e., for the parameter which requires the highest reduction in temperature).