

Analog Circuits and Signal Processing

Series Editors

Mohammed Ismail, Khalifa University, Dublin, OH, USA

Mohamad Sawan, 18, Shilongshan Road, School of Engineering, Westlake University, Hangzhou, Zhejiang, China

The *Analog Circuits and Signal Processing* book series, formerly known as the *Kluwer International Series in Engineering and Computer Science*, is a high level academic and professional series publishing research on the design and applications of analog integrated circuits and signal processing circuits and systems. Typically per year we publish between 5-15 research monographs, professional books, handbooks, and edited volumes with worldwide distribution to engineers, researchers, educators, and libraries. The book series promotes and expedites the dissemination of new research results and tutorial views in the analog field. There is an exciting and large volume of research activity in the field worldwide. Researchers are striving to bridge the gap between classical analog work and recent advances in very large scale integration (VLSI) technologies with improved analog capabilities. Analog VLSI has been recognized as a major technology for future information processing. Analog work is showing signs of dramatic changes with emphasis on interdisciplinary research efforts combining device/circuit/technology issues. Consequently, new design concepts, strategies and design tools are being unveiled. Topics of interest include: Analog Interface Circuits and Systems; Data converters; Active-RC, switched-capacitor and continuous-time integrated filters; Mixed analog/digital VLSI; Simulation and modeling, mixed-mode simulation; Analog nonlinear and computational circuits and signal processing; Analog Artificial Neural Networks/Artificial Intelligence; Current-mode Signal Processing; Computer-Aided Design (CAD) tools; Analog Design in emerging technologies (Scalable CMOS, BiCMOS, GaAs, heterojunction and floating gate technologies, etc.); Analog Design for Test; Integrated sensors and actuators; Analog Design Automation/Knowledge-based Systems; Analog VLSI cell libraries; Analog product development; RF Front ends, Wireless communications and Microwave Circuits; Analog behavioral modeling, Analog HDL.

Bart Philippe • Patrick Reynaert

Mm-wave Circuit Design in 16nm FinFET for 6G Applications



Springer

Bart Philippe
ESAT-MICAS
KU Leuven
Leuven, Belgium

Patrick Reynaert
ESAT-MICAS
KU Leuven
Leuven, Belgium

ISSN 1872-082X ISSN 2197-1854 (electronic)
Analog Circuits and Signal Processing
ISBN 978-3-031-11223-2 ISBN 978-3-031-11224-9 (eBook)
<https://doi.org/10.1007/978-3-031-11224-9>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Our online activity is growing at an unprecedented rate. New online applications change the way we communicate, work, play, watch entertainment and so much more. The requirements put on the network to enable these applications are already stringent and congest the available bandwidth. Future applications such as virtual or extended reality will put even more stringent requirements in place. For example, to provide a high response time and increased video streaming quality, the latency needs to decrease and data rates need to improve up to a 50-fold increase. The sixth-generation (6G) mobile networks will address these challenges by making extensive use of the large available bandwidth at mm-wave frequencies.

CMOS is the technology of choice for mass-produced circuits. It provides high yields at low costs, while the continued downscaling has led to fast transistors allowing the implementation of mm-wave circuits. The possibility to integrate mm-wave and low-power analog building blocks with digital signal processing has made CMOS a key technology for mass adoption. Deeply scaled CMOS has, therefore, gained growing attention for future 6G mm-wave fronts from both industry and research institutes. Unfortunately, aggressive technology scaling does not only provide benefits but also comes with its own challenges. Firstly, the smaller supply voltage limits the maximal voltage swing, limiting linearity, SNR, and output power. Secondly, metal interconnects get thinner and closer to the substrate resulting in larger resistances, limiting the achievable mm-wave performance. The advent of FinFET technology to continue digital density scaling by using the vertical space instead of smaller planar transistors has resulted in temporary stagnation in the mm-wave performance. The new 3D transistor topology introduces new challenges and changes the design trade-offs in the mm-wave designs compared to planar technologies.

This book looks at the impact on mm-wave circuit design, for going from a planar CMOS technology toward a FinFET technology. It introduces techniques and practices to realize high-performance mm-wave circuits operating between 90 GHz and 170 GHz. First, a more in-depth look at the elementary active and passive components is taken. The FinFET transistor is analyzed and optimized for mm-wave performance. The effects of nanoscale technology scaling on widely used passive

components and the corresponding impact on mm-wave circuits are discussed. At the end of the chapter, the active and passive components are combined to form a fundamental mm-wave amplifier block. Secondly, the challenges of frequency generation in a CMOS FinFET technology are discussed. The challenges are approached from two viewpoints, which lead to two design examples; a fundamental and a harmonic oscillator. The third chapter discusses the basics of mm-wave power amplifiers in a FinFET technology. The maximal power limits are discussed and solutions to these challenges are proposed. This results in the design example of a power combining power amplifier implemented in a 16 nm FinFET technology. Finally, the issue of the low efficiency of mm-wave power amplifiers above 100 GHz is discussed. This leads to a novel power amplifier topology with improved back-off efficiency and linearity. In a design example, the power amplifier is implemented in a direct conversion transmitter, in 16 nm FinFET technology, demonstrating data transmission up to 44 Gbps.

Leuven, Belgium
Leuven, Belgium
April 2022

Bart Philippe
Patrick Reynaert

Contents

1	Introduction	1
1.1	Toward the Sixth-Generation (6G) Mobile Networks	1
1.1.1	Millimeter Wave to Increase Data Rates	1
1.2	Millimeter Wave in FinFET CMOS, the Next Step	3
1.3	Book Outline	5
	References	5
2	Basic Components in mm-Wave Design	7
2.1	Actives	7
2.1.1	f_T , f_{max} a FoM for mm-Wave Transistors	8
2.1.2	The Effect of Scaling; mm-Wave in FinFET	10
2.1.3	Transistor Layout and mm-Wave Performance	12
2.2	Passives	17
2.2.1	Capacitors	18
2.2.2	Inductors	20
2.2.3	Transformers	22
2.2.4	Transmission Lines	25
2.3	Basic Design of mm-Wave Circuits	26
2.3.1	A Capacitive Neutralized mm-Wave Amplifier	26
2.3.2	Design of mm-Wave Interconnects	28
2.4	Conclusion	33
	References	34
3	Frequency Generation	37
3.1	VCO Basics	37
3.1.1	Brief on Phase Noise	38
3.1.2	Frequency Tuning	40
3.2	LO Architectures and Trends at mm-Wave	42
3.3	Challenges of a Deeply Scaled Technology at mm-Wave	44
3.4	A 4th Order Transformer-Based Resonator	47
3.5	Design Example: A Fundamental Oscillator	50
3.5.1	The Cross-Coupled Pair	50

3.5.2	The Resonator Design	52
3.5.3	Measurement Results	54
3.6	Design Example: A Harmonic Oscillator	55
3.6.1	Basic Principle	55
3.6.2	Proposed Design	57
3.6.3	Output Buffers	58
3.6.4	Measurement Results	61
3.7	Conclusion	63
	References	66
4	Power Amplification	69
4.1	Introduction	69
4.2	Power Amplifier Basics	69
4.2.1	Gain and Power	69
4.2.2	Efficiency.....	71
4.2.3	PA Linearity.....	72
4.2.4	PA Classes and Biasing.....	73
4.2.5	Challenge at mm-Wave in a Deeply Scaled Technology.....	78
4.3	Design Example: A Class-A Two-Way Power Combining D-Band PA in 16 nm FinFET	83
4.3.1	Design of the 16 nm FinFET Transistor	83
4.3.2	Build-up of the Power Amplifier.....	85
4.3.3	Measurement Results	88
4.4	Conclusion	93
	References	94
5	A D-band Direct-Conversion Transmitter with Enhanced PA	97
5.1	Efficiency Enhancement Techniques	97
5.1.1	Common Efficiency Enhancement Techniques.....	97
5.1.2	A Sequential Power Amplifier	101
5.1.3	Conclusion	105
5.2	Linearization Techniques by Dynamic Bias	105
5.2.1	AM–AM Compensation	105
5.2.2	AM–PM Compensation	106
5.3	Design Example: A Direct-Conversion TX with Enhanced PA	110
5.3.1	The Dynamic Bias Enhanced PA	110
5.3.2	The I/Q Modulator and LO Generation.....	114
5.4	Measurements and Discussion	118
5.4.1	Power Amplifier Measurements	119
5.4.2	Transmitter Measurements	123
5.5	Conclusion	131
	References	132
Index		135

Acronyms and Symbols

AMAM	Amplitude modulation to amplitude modulation
AMPM	Amplitude modulation to phase modulation
AWG	Arbitrary waveform generator
BEOL	Back end of line
BER	Bit error rate
BiCMOS	Bipolar CMOS
BW	Bandwidth
CPW	Coplanar waveguide
DAC	Digital to analog converter
DC	Direct current
DE	Drain efficiency
DRC	Design rule check
DWG	Dielectric waveguide
EIRP	Equivalent isotropically radiated power
EVM	Error vector magnitude
FFT	Fast fourier transform
FinFET	Fin field-effect transistor
FoM	Figure of merit
FoMt	Figure of merit tuning (used for oscillators)
FSPL	Free space path loss
FSU	Spectrum analyzer from Rohde & Schwarz
FSW	Spectrum analyzer from Rohde & Schwarz
GSG	Ground-signal-ground
IF	Intermediate frequency
IFFT	Inverse fast fourier transform
IL	Insertion loss
IM	Intermodulation
IMRR	Image rejection ratio
IR	Infrared
ISI	Intersymbol interference
LNA	Low noise amplifier

LO	Local oscillator
LOFT	Local oscillator feedthrough
MIM-capacitor	Metal-insulator-metal capacitor
MOM-capacitor	Metal-oxide-metal capacitor
MOS	complementary Metal oxide semiconductor
MOS	Metal oxide semiconductor
MUT	Material under test
MUX	Multiplexer
NF	Noise figure
NMOS	N-type metal oxide semiconductor
OP1dB	Output referred 1dB compression point
PA	Power amplifier
PAE	Power added efficiency
PAPR	Peak to average power ratio
PBO	Power back-off
PCB	Printed circuit board
PDK	Process design kit
PLL	Phase-locked loop
PM5	Erickson power meter from Virginia diodes
PMOS	P-type metal oxide semiconductor
PN	Phase noise
PSD	Power spectral density
PSG	Signal generator from Keysight
PVT	Process variation and temperature
QAM	Quadrature amplitude modulation
R&S	Rohde & Schwarz
RBW	Resolution bandwidth
RDL	Redistribution layer
RF	Radio frequency
SiGe	Silicon-germanium
SNR	Signal to noise ratio
SR	Sample rate
TIA	Transimpedance amplifier
TL	Transmission line
TSMC	Taiwanese semiconductor manufacturing company
VDI	Virginia Diodes Inc.
VNA	Vector network analyzer
<i>T</i>	Temperature
<i>W</i>	Bandwidth
<i>C_{channel}</i>	Channel capacity
<i>S/N</i>	Signal to noise ratio
<i>P</i>	Power
<i>f</i>	Frequency
<i>d</i>	Distance
λ	Wave length

Q	Quality factor
g_m	Transconductance
r_G	Gate resistance
f_T	Transition frequency
f_{max}	Maximum oscillation frequency
G	Gain
W	Width
N_f	Number of fingers
n_{fin}	Number of fins
ω	Angular frequency
k_m	Mutual coupling
α	Attenuation constant
β	Phase constant
γ	Propagation constant
v	Wave velocity
Γ	Reflection coefficient
S	Constellation symbol
dB	Decibel
dBm	Decibels relative to 1 mW
dBc	Decibels relative to the carrier
Y	Admittance
Z	Impedance
C	Capacitance
L	Inductance
R	Resistance
V	Volt

Chapter 1

Introduction



1.1 Toward the Sixth-Generation (6G) Mobile Networks

The sixth-generation (6G) wireless network will provide a multifold increase in performance compared to the fifth-generation (5G) network, being rolled out at the moment of writing. One of the major improvement goals for 6G is to increase the peak data rate by a factor of 50, from 20 Gbps to 1000 Gbps with an average user experience data rate of 1 Gbps [1–3]. The increased data rates are required to support advanced multimedia applications. For example, virtual reality (VR) or extended reality (XR) requires high-resolution video streaming to a single user or between multiple users. A second requirement is a low latency of 100 μ s for the over-the-air latency and 1 ms for the end-to-end latency [1]. Furthermore, the 6G systems would require improvements on reduced power consumption and increased reliability for industry V4.0 but also enhanced mobility and extreme coverage. Of course all these performance metrics cannot be met at the same time so a whole range of technical solutions will need to be applied to realize the 6G networks of the future.

1.1.1 Millimeter Wave to Increase Data Rates

To find an answer to the question on how the data rate can be improved, and thereby facilitate the 6G requirements, we first need to look at the communication basics. In 1948, Claude Shannon already gave an upper limit for the capacity of a data channel in [4] as formulated below:

$$C_{channel} = BW \log_2(1 + SNR). \quad (1.1)$$

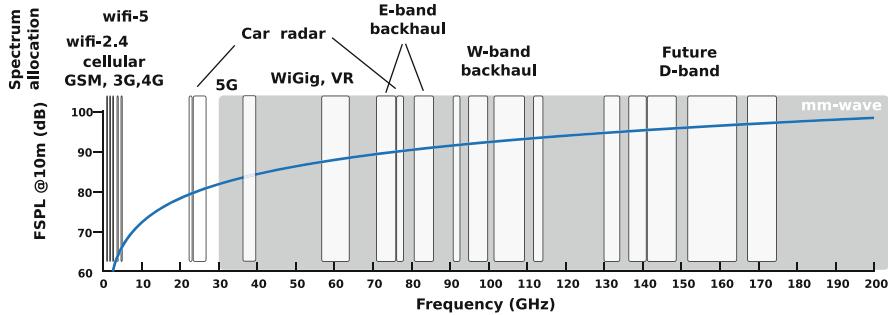


Fig. 1.1 Some major allocation of the electromagnetic spectrum and the FSPL at 10m at each frequency

The total channel capacity ($C_{channel}$) is directly proportional to the used bandwidth (BW) and can be further improved by the signal-to-noise power ratio (SNR). The signal-to-noise ratio can only be improved so far, and the benefits diminish rapidly due to its logarithmic contribution. Furthermore, to exploit the high SNR in a practical application, high symbol schemes and therefore complex modulation schemes are required such as 4096-QAM or higher. These complex modulation rates put very stringent linearity requirements on the circuits and system. The only viable method to further increase the channel capacity is to increase the bandwidth. Figure 1.1 shows the electromagnetic spectrum up to 200 GHz and some of the widely used regulated frequency bands for communication and sensing. The low frequencies of the current cellular and Wi-Fi systems hold fairly small bandwidths. For example, a Wi-Fi 802.11 user only has a 22 MHz bandwidth available. Furthermore, the small available bandwidth needs to be spatially shared with a lot of users. For the deployment of 5G, the operating frequency is already increased up to 39 GHz to increase the bandwidth and meet the corresponding data rate requirements. For 6G, the research community is looking to further increase the available bandwidth by increasing the operating frequency to 100 GHz and above. Recently, the European Conference of Postal and Telecommunications administrations (CEPT) have released its recommendations for both the W-band ((18)02) and the D-band ((18)01); the U.S. Federal Communications Commission (FCC) (2019) and the International Telecommunication Regulations (ITR) have released its radio regulations for these bands (2016) as indicated in Fig. 1.1. The total bandwidths made available in the W-band and D-band are 17.85 GHz and 31.8 GHz, respectively, or in total about 50 GHz could become available for use.

Increasing the operating frequency (f_r) not only unlocks more spectral bandwidth, but also provides a higher bandwidth for a given resonator quality factor Q as given below:

$$BW = \frac{f_r}{Q}. \quad (1.2)$$

The property of bandwidth to scale proportionally with frequency has led the communication industry to higher frequencies. Increasing the operating frequency is not for free and comes with a large apparent wireless transmission loss. This apparent wireless transmission loss is captured in the free space path loss (FSPL) given below:

$$FSPL = \frac{P_r}{P_t} = \left(\frac{\lambda}{4\pi d} \right)^2. \quad (1.3)$$

The FSPL is the ratio between the received power (P_r) and the transmitted power (P_t). An isotropic transmitter will transmit its power in all directions, where the power is evenly distributed across the surface of a sphere centered around the transmitter. The power density at the surface of the transmitted sphere thus diminishes quadratically with the distance (d) from the transmitter. On the receiver side, only a small part of the transmitted sphere surface is captured by the receiver antenna. Furthermore, the receiver antenna area is directly proportional with the square of the wavelength (λ). A higher frequency will result in a diminished received power [5]. The FSPL over a distance of 10 m expressed in dB is shown on top of the frequency spectrum in Fig. 1.1. This increased path loss of a mm-wave system puts a constraint on the system's link power budget and limits its communication distance. Fortunately, by utilizing high gain antennas in the transmitter (G_T), power can be directed more efficiently toward the receiver, and at the receiver side, a high gain antenna (G_R) can increase its capture area and receive more power. This combines into the modern Friis transmission equation as follows [5]:

$$\frac{P_r}{P_t} = G_T G_R \left(\frac{\lambda}{4\pi d} \right)^2. \quad (1.4)$$

A common method that is already used in 5G networks to achieve a high gain antenna is a phased array system. In a phased array, multiple antenna elements are combined to form a single large array. The constructive behavior of the transmitted signal of each element results in a narrow beam with a high total antenna gain. Furthermore, by controlling the phase of each element in the array, the direction of the beam can be electrically controlled. The electrical beam control allows for an improved coverage with a high gain antenna and provides spatial selectivity.

1.2 Millimeter Wave in FinFET CMOS, the Next Step

The fin field-effect transistor (FinFET) logic process technology was introduced in 2011 [6, 7]. The traditional planar transistor was replaced by a three-dimensional geometry, a silicon fin going upward from the substrate. Figure 1.2 shows a side-by-side comparison of a traditional planar transistor and a FinFET. The active area

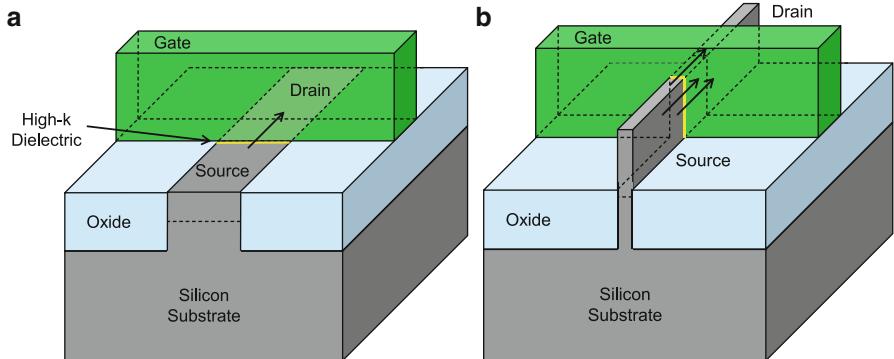


Fig. 1.2 (a) A traditional planar transistor and (b) a 3D FinFET [8]

of the transistor is now not limited to the surface of the silicon wafer but runs along the side and top of the fin. The use of vertical active area results in an increased active density and allows Moore's law to continue. Besides the density benefits, the channel formed inside the narrow silicon fin is under better electrostatic control by the surrounding gate. The improved electrostatic control results in a steeper sub-threshold slope leading to decreased leakage current [7].

In a traditional planar technology, increasing the density came from scaling down transistors; hence, each node was named after the smallest dimension, the gate length. For the analog designer, the shrinking gate lengths resulted in faster transistors. The transconductance g_m scales with $1/L$, and the parasitic gate capacitance also reduces for smaller channel lengths resulting in a higher unity gain frequency, f_T . The step toward FinFET technology has no reduction in gate length and will therefore stop improving the high-frequency performance for the analog and mm-wave designer. The Fin geometry results in changed transistor parasitics, crucial for mm-wave design, such as the gate resistance. The increased vertical component in the gate resistance will lead to a reduced gate resistance and a lower maximal oscillation frequency (f_{max}). Fortunately, better channel control improves the drain-induced barrier lowering and will result in an increased output impedance, leading to an increased voltage gain.

When the mm-wave designer should not expect any improved mm-wave performance and will come across new challenges, why move to a FinFET technology? The goal of fully integrating a 6G mm-wave phased array does not only require excellent mm-wave circuits attached to each antenna but also requires significant analog and digital parts [9]. It is in these non-mm-wave circuits that FinFET can add a lot of benefits. The complexity of controlling all the antennas of a 6G beam-former requires significant amounts of digital circuits [10]. The large 6G bandwidths need to be supported by the baseband circuits and require high-speed analog circuits that can benefit from the FinFET process.

In this book, the challenges and limitations of designing millimeter-wave circuits in a FinFET process are tackled. The FinFET process is analyzed, and design

examples for oscillators, power amplifiers, and a transmitter are given. A more detailed outline is given below.

1.3 Book Outline

- Chapter 2 introduces the basic circuit components in mm-wave design and how they are used to create the more complex circuits in the following chapters. The challenges and limits of using bulk CMOS transistors at mm-wave frequencies are discussed. The effects and challenges of technology scaling on the performance of the transistors as we move from a 40 nm bulk CMOS to a 16 nm FinFET technology are presented. A similar discussion is performed on the most common passive devices: capacitors, inductors, transmission lines, and transformers. The design of a basic differential amplifier from these components is discussed and how a transformer is used to interconnect multiple active devices in the design.
- Chapter 3 gives the basis for mm-wave frequency generation with an overview of the expected performance from several topologies at mm-wave frequencies. The effect of scaling on the performance of oscillators is discussed. Finally, two design examples with layout and measurements in 16 nm FinFET technology are given: a fundamental oscillator and a harmonic oscillator.
- Chapter 4 is dedicated to the most power-hungry building block of the transmitter: the power amplifier. Power amplifiers' basic parameters and design metrics are given and how the design at mm-wave frequencies in a deeply scaled technology impacts them. The chapter is concluded with the design of a two-way power combining PA in 16 nm FinFET, the design and its limits are discussed in detail, and measurements of the fabricated die are shown.
- Chapter 5 extends on Chap. 4 by looking at the challenges of improving the efficiency and linearity of mm-wave PAs and more specifically of PAs above 100 GHz. This has led to the design of a power combining PA with a dynamically biased PA branch. The dynamic bias is used to improve the back-off efficiency as well as both amplitude and phase linearity. The PA design is further combined with a modulator to form a direct conversion transmitter at 125 GHz. The design is discussed, and both the PA and TX are fabricated in a 16nm FinFET CMOS process. The characterization and modulated measurements are shown and discussed.

References

1. Samsung Research, *6G the next hyper-connected experience for all* (2022). <https://research.samsung.com/next-generation-communications>. Accessed: 2022-03-05
2. Y. Wu, S. Singh, T. Taleb, A. Roy, H.S. Dhillon, M.R. Kanagarathinam, A. De, *6G Mobile Wireless Networks* (Springer, Berlin, 2021)

3. M. Polese, J.M. Jornet, T. Melodia, M. Zorzi, Toward end-to-end, full-stack 6G terahertz networks. *IEEE Commun. Mag.* **58**(11), 48–54 (2020)
4. C.E. Shannon, A mathematical theory of communication. *The Bell System Technical Journal* **27**(3), 379–423 (1948)
5. D. Pozar, *Microwave Engineering Fourth Edition* (2005)
6. C. Auth, C. Allen, A. Blattner, D. Bergstrom, M. Brazier, M. Bost, M. Buehler, V. Chikarmane, T. Ghani, T. Glassman, R. Grover, W. Han, D. Hanken, M. Hattendorf, P. Hentges, R. Heussner, J. Hicks, D. Ingerly, P. Jain, S. Jaloviar, R. James, D. Jones, J. Jopling, S. Joshi, C. Kenyon, H. Liu, R. McFadden, B. McIntyre, J. Neirynck, C. Parker, L. Pipes, I. Post, S. Pradhan, M. Prince, S. Ramey, T. Reynolds, J. Roesler, J. Sandford, J. Seiple, P. Smith, C. Thomas, D. Towner, T. Troeger, C. Weber, P. Yashar, K. Zawadzki, K. Mistry, A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors, in *2012 Symposium on VLSI Technology (VLSIT)* (2012), pp. 131–132
7. M.T. Bohr, I.A. Young, CMOS scaling trends and beyond. *IEEE Micro* **37**(6), 20–29 (2017)
8. Intel, Intel's revolutionary 22 nm transistor technology (2011). http://download.intel.com/newsroom/kits/22nm/pdfs/22nm-Details_Presentation.pdf. Copyright© Intel Corporation, Accessed: 2020-06-15
9. P. Skrimponis, S. Dutta, M. Mezzavilla, S. Rangan, S.H. Mirfarshbafan, C. Studer, J. Buckwalter, M. Rodwell, Power consumption analysis for mobile mmWave and sub-THz receivers, in Proceedings of the 2020 2nd 6G Wireless Summit (6G SUMMIT) (IEEE, New York, 2020), pp. 1–5
10. B. Sadhu, A. Paidimarri, W. Lee, M. Yeck, C. Ozdag, Y. Tojo, J.-O. Plouchart, X. Gu, Y. Uemichi, S. Chakraborty, Y. Yamaguchi, N. Guan, A. Valdes-Garcia, A 24-to-30GHZ 256-element dual-polarized 5G phased array with fast beam-switching support for >30, 000 beams, in *Proceedings of the 2022 IEEE International Solid-State Circuits Conference (ISSCC)*, vol. 65 (2022), pp. 436–438

Chapter 2

Basic Components in mm-Wave Design



2.1 Actives

The main active device in integrated circuits is the transistor. A MOS transistor available in a CMOS process consists in its most basic form, as the name explains, of a metal on top of a semiconductor with a very thin piece of oxide sandwiched in between. A cross section can be seen in Fig. 2.1. In its most basic operation, a conductive channel between the drain (D) and Source (S) of the transistor can be created, by applying a voltage at the gate (G). This channel will allow a current to flow between the drain and the source (I_{DS}). It is the modulation of this current by the gate-to-source voltage that will perform all the magic and make analog, RF, and mm-wave design possible.

The design of the MOS-transistor device is a complex work in its own right and is handled by the process engineers of the foundries. However, it is still interesting to look at some of the technology and process parameters that will affect our design process. The derivation and insights of these technology parameters are beyond the scope of this work and can be found in [1]. The cross section in Fig. 2.1 shows a NMOS transistor in saturation ($V_{GS} > V_t$ and $V_{DS} > V_{GS} - V_t$ with V_t the threshold voltage). Both the oxide and the depletion layer have a thickness t_{ox} and t_{si} with dielectric constants ϵ_{ox} and ϵ_{si} , respectively. These lead to the oxide capacitor density $C_{ox} = \epsilon_{ox}/t_{ox}$ and the depletion capacitor density $C_D = \epsilon_{si}/t_{si}$. Furthermore, the technology parameter $K' = (\mu_n C_{ox})/(2n)$ with $n = C_D/C_{ox} + 1$, where μ_n is the electron mobility, and the velocity saturation: v_{sat} , where all the electrons reach their maximum speed, are defined.

For the basic operation, a transistor can be represented by a small-signal equivalent model as shown in Fig. 2.2. At low frequencies, the capacitors can be viewed as open circuits, and the gate resistance (r_G) can thus be ignored. The voltage-dependent channel is modeled as a transconductance (g_m) by taking the

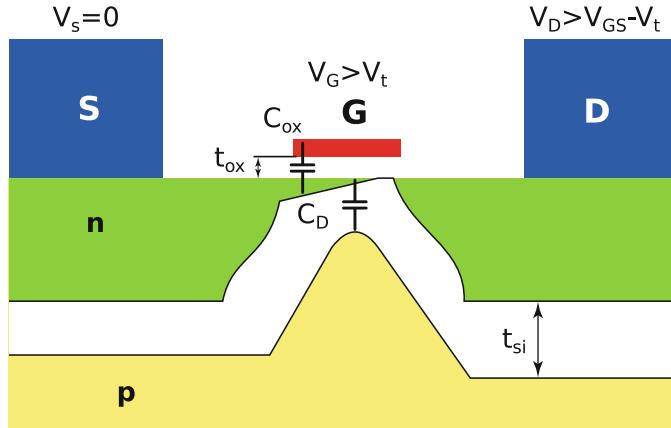


Fig. 2.1 Cross section of a planar (N)MOS transistor in saturation [1]

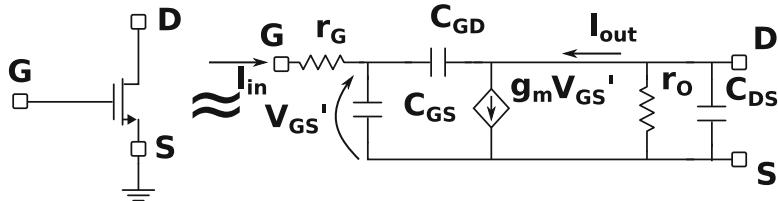


Fig. 2.2 NMOS transistor and its simplified small-signal model

derivative of I_{DS} with respect to the gate–source voltage. The transconductance in strong inversion (SI) and velocity saturation (VS) region is given as follows:

$$g_{m,SI} = K' \frac{W}{L} 2(V_{GS} - V_t) \quad (2.1)$$

$$g_{m,VS} = W C_{ox} v_{sat}, \quad (2.2)$$

with W and L the transistors' width and length, respectively. The intrinsic small-signal voltage gain is defined as $A_v = g_m \cdot r_O$ with r_O the output resistance of the transistor itself. The intrinsic analog bandwidth, and thus speed, at the input is limited by C_{GS} and at the output by C_{DS} .

2.1.1 f_T , f_{max} a FOM for mm-Wave Transistors

The high-frequency potential of the transistor is often expressed by two factors: f_T , the transition frequency, and f_{max} the maximum oscillation frequency. Both parameters act as a figure of merit (FOM) for high-frequency performance. Whereas

f_T is looked at by analog designers, f_{max} is used by mm-Wave designers as this incorporates some of the penalties paid by the practical layout of the transistor .

The transition frequency f_T is defined as: the frequency for which the current gain of the transistor, as shown in Fig. 2.2, is equal to 1, when the drain is shorted to the source (2.3). From this, the f_T is calculated:

$$|H_{21}(f)| = \frac{I_{out}(f)}{I_{in}(f)} = 1 \quad (2.3)$$

$$f_T \approx \frac{g_m}{2\pi(C_{GS} + C_{DS})}. \quad (2.4)$$

If we substitute g_m for the technology version for SI and VS from Eqs. (2.1) and (2.2), the f_T becomes the following:

$$f_T \approx \frac{3\mu}{4\pi L^2}(V_{GS} - V_t) \quad (2.5)$$

$$f_T \approx \frac{v_{sat}}{2\pi L}. \quad (2.6)$$

In both cases, the f_T increases for smaller transistor lengths L, clearly demonstrating the benefit of technology scaling for high-frequency circuits. The f_T relies on current gain; however, mm-wave design is dominated by power and power waves. It is, therefore, more interesting to look at the power gain of the transistor. To overcome the frequency limitations, caused by the parasitic capacitances, matching techniques are applied at higher frequencies. These techniques resonate out the capacitors at the operating frequency, avoiding the otherwise low impedance caused by the capacitor at high frequencies. Matching techniques will be discussed later in the chapter. The (unilateral) power gain is defined as the ration of output power over input power, in their respective conjugate load. Which can be further reduced assuming $C_{GD} = 0$ for the small-signal model in Fig. 2.2 as follows:

$$|G_U(f)| = \frac{P_{out}(f)}{P_{in}(f)} \quad (2.7)$$

$$|G_U(f)| = \frac{R_{DS}}{4 \cdot r_G} \cdot \left(\frac{g_m}{2\pi f C_{GS}} \right)^2 = \frac{r_O}{4 \cdot r_G} \cdot \left(\frac{f_T}{f} \right)^2. \quad (2.8)$$

The maximum power gain benefits from a high f_T or a small technology node and requires a low gate resistance and large output resistance. Despite both becoming difficult at higher frequencies, the gain is inherently dependent on the frequency.

From the definition of power gain, one can also define the frequency where the gain is equal to 1, as with f_T . This is the maximum oscillation frequency f_{max} and can be found as follows:

$$|G_{max}(f)| = \frac{P_{out}(f)}{P_{in}(f)} = 1 \quad (2.9)$$

$$f_{max} \approx \frac{g_m}{2\pi C_{GS}} \sqrt{\frac{r_O}{r_G}} \approx \frac{f_T}{2} \sqrt{\frac{r_O}{r_G}}. \quad (2.10)$$

From Eq. (2.10), it is clear that f_{max} takes f_T and takes into account the parasitic effects of the gate resistance due to layout. The layout of mm-wave transistors will be discussed in detail later in the chapter.

2.1.2 The Effect of Scaling; mm-Wave in FinFET

From the previous section, it is clear that downscaling the dimensions of a transistor benefits the mm-wave designer. However, the drive for smaller transistors is done for the digital designer. In 1965, Gordon Moore noticed that every 2 years the number of transistors integrated on a single die doubles. He then further predicted that this trend would continue for at least the coming 10 years [2]. This trend is called Moore's law and has been used as a goal to follow by process designers. Figure 2.3 shows the total number of transistors in processors coming to market; this clearly shows the trend of density continuing into 2020.

At the time when Gordon Moore made his prediction, an increase in density was achieved by downscaling the transistors resulting in a smaller gate length. This has led to identifying each new technology by its gate length, from 10 μm around 1970 down to 7 nm in 2020, as shown in Fig. 2.4. Unfortunately for the most recent

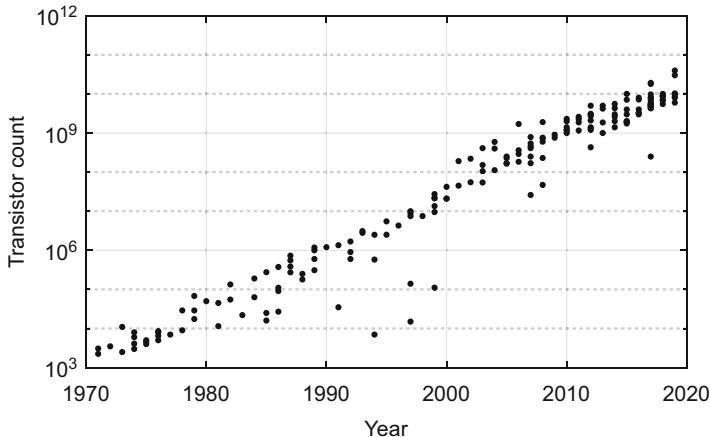


Fig. 2.3 An overview of the number of transistors in a processor over time (data from [3])

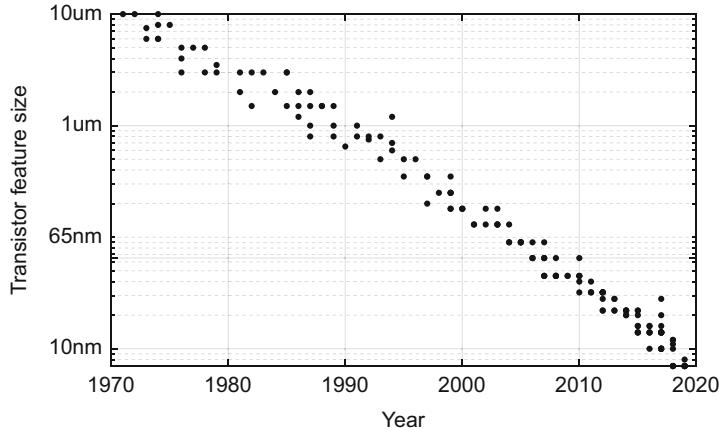


Fig. 2.4 An overview of the transistor feature size of processors over time (data from [3])

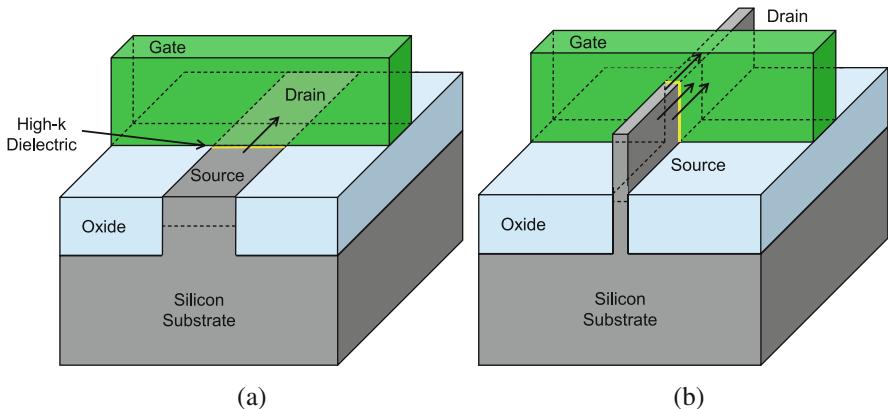


Fig. 2.5 (a) A traditional planar transistor and (b) a 3D FinFET [7]

smallest nodes, the name does not correspond to the real feature size but indicates an improvement in density by applying different techniques [4, 5].

One of these techniques is by going from planar transistors to three-dimensional transistors, first by extending the channel upward as a fin field-effect transistor (FinFET) and in the future with channel all-round transistors [6]. Figure 2.5 shows a FinFET next to a traditional planar transistor.

The active channel is now not only at the planar surface but all around the protruding fin at the interface with the gate metal. This allows to effectively increase the active width of the transistor with twice the height of the fin. The increased width for a smaller area has resulted in the desired density increase for a FinFET technology. Furthermore, this has resulted in a better-controlled channel leading to a reduced leakage current.

Although the name implies a smaller gate length, FinFETs at smaller nodes do not gain a high-frequency performance boost because the gate length has not decreased in size. Furthermore, the new 3D fin structure changes the dynamics of the gate resistance as discussed in the following section. The overall result is a stagnation in the improvement of mm-wave performance. Furthermore, as the technology continues to scale down so does the nominal supply voltage. A reduced supply voltage only penalizes the analog and mm-wave designer, as a lower voltage swing translates to lower linearity for LNAs or lower output powers for PAs.

2.1.3 Transistor Layout and mm-Wave Performance

The transistor is sized depending on the specific requirements: linearity, current, transconductance, etc. It is assumed that scaling the width of a transistor with a factor n will also lead to a linear change of the parasitic capacitance with a factor n and the gate resistance with a factor $1/n$. Under this assumption, the mm-wave performance of the transistor is independent of the size of the transistor. Unfortunately, this is not the case. The gate resistance is determined on one hand by the technology and the other hand by the layout of the transistor. The technology determines the resistivity of the gate material. From 28 nm and smaller, a high- k metal gate is used instead of the older poly gate technology to reduce gate leakage and improve electrostatic control. This, however, increases the intrinsic gate resistance of the transistor [8] and how we design with it. The layout of the transistor can have a large impact on the overall gate resistance, for example, a long gate will show a larger gate resistance compared to multiple parallel short gates. Figure 2.6

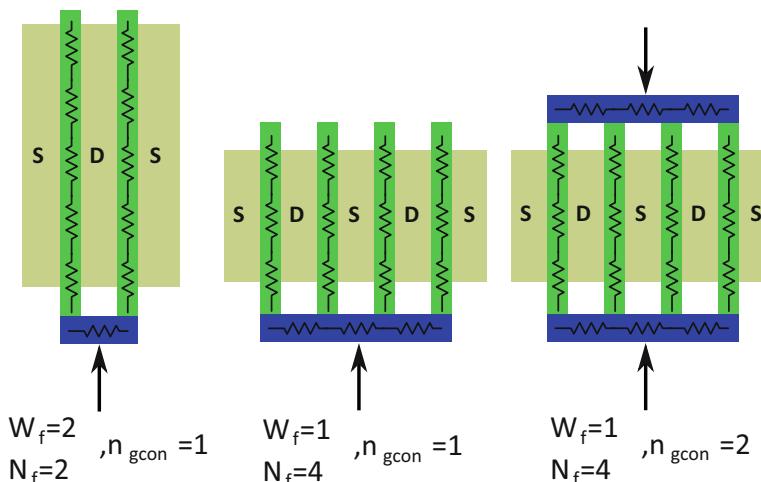


Fig. 2.6 Three different layouts of an equal $4 \mu\text{m}$ width transistor

shows the layout of the same transistor in three different layout configurations: 2 fingers with a large finger width of $2\text{ }\mu\text{m}$ connected from one side, 4 fingers with a small finger width connected from one side, and 4 fingers with a small finger width connected from both sides. The minimal gate resistance for a certain transistor size, therefore, depends on the layout parameters W_f , N_f , and n_{gcon} , and the technology parameters $R_{interconn}$, r_{sh} , and r_v , which represent: finger width, the total number of fingers, the number of gate connections, metal sheet resistance, the gates' horizontal resistance, and the gates' vertical resistance, respectively. To calculate the total gate resistance of a transistor, the interconnect resistance is divided into three pieces. The first is the resistance due to the interconnects between the fingers, $R_{g,int}$. The second part is the horizontal gate resistance over the length of each finger, $R_{g,h}$. The third resistance is the vertical resistance down from the gate conductor to the gate dielectric, $R_{g,v}$. Using this, the total gate resistance can be calculated as follows: [9].

$$R_{g,tot} = R_{g,int} + R_{g,h} + R_{g,v} \quad (2.11)$$

$$R_{g,tot} = \frac{R_{interconn}N_f}{n_{gcon}} + \frac{r_{sh}W_f}{3n_{gcon}^2 \cdot n_f} + \frac{r_v}{n_f W_f}. \quad (2.12)$$

When the technology parameters $R_{interconn}$, r_{sh} , and r_v change between nodes, the optimal finger width and the number of fingers may shift. For example, when a high-k metal gate results in a higher r_{sh} and r_v , smaller finger widths are preferred to reduce the total resistance.

For a FinFET technology, the finger width is discretized to a number of fins n_{fin} , and the total gate resistance becomes as follows:

$$R_{g,tot} = \frac{R_{interconn}N_f}{n_{gcon}} + \frac{r_{sh}n_{fin}}{3n_{gcon}^2 n_f} \left(1 - \frac{1}{4n_{fin}^2}\right) + \frac{r_v}{3n_f \cdot n_{fin}} \left(1 - \frac{1}{2n_{fin}}\right). \quad (2.13)$$

For which, r_v now represents the resistance down to the side of the fins as shown in Fig. 2.7 [10, 11].

Figure 2.8 shows the f_{max} of a transistor for different finger widths for a 40 nm and 16 nm CMOS technology at different sizes. The 40 nm technology shows a peak f_{max} around 290 GHz for the smallest total transistor width of $10\text{ }\mu\text{m}$ and a finger width of $0.8\text{ }\mu\text{m}$. For increasing total widths, the optimal finger width increases to get the optimum interconnect and finger resistance. Although the optimum shifts, the f_{max} does drop with increasing size. By contrast, the 16 nm FinFET technology shows a peak f_{max} of 280 GHz for the smallest transistor of 90 fins and 6 fins per finger. Two differences stand out, the first difference is the optimal finger width that stays constant at 6 fins per finger, there is no trade-off in the interconnect, and the gate metal connection on top of the fins is the dominant resistance. The second

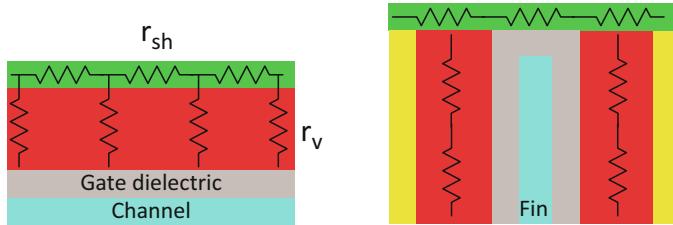


Fig. 2.7 Cross section of a planar transistor and a FinFET, showing the horizontal and vertical gate resistances together with the gate dielectric and the channel

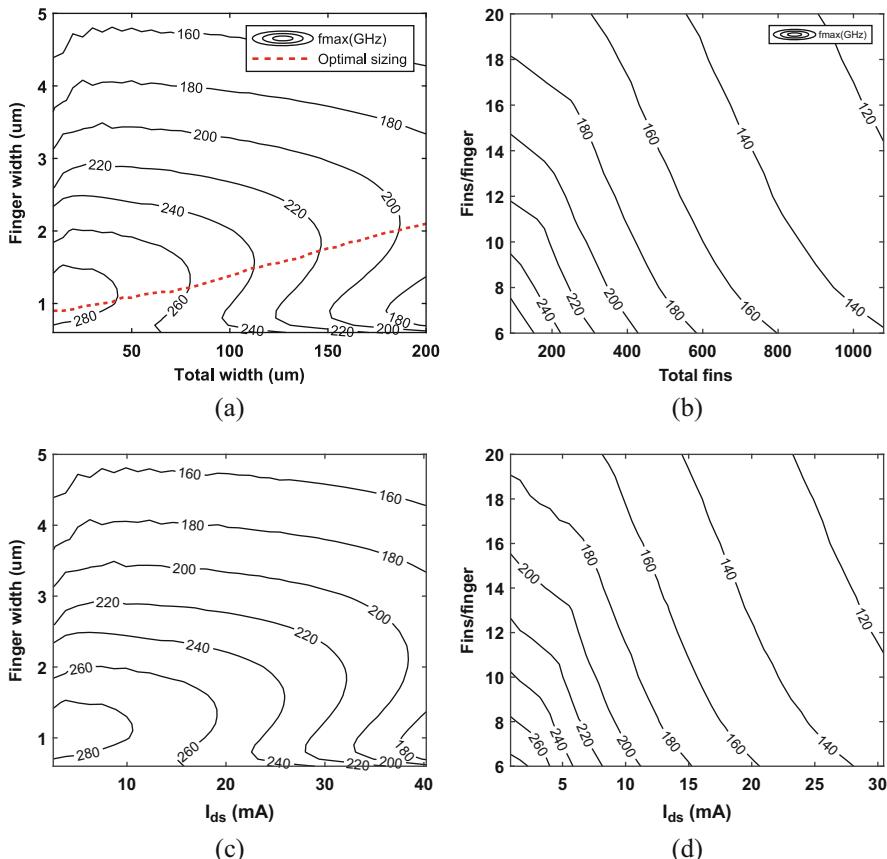


Fig. 2.8 The f_{max} of (a), (c) a 40 nm and (b), (d) 16 nm technology dependence on transistor sizing for total size and total drain–source current for comparison

and most problematic is a steep degradation in mm-wave performance as the total width is increased. As the transistor is sized up from 90 fins to 1080 fins, a 12-fold increases and the f_{max} drops by 50%. For the 40 nm technology, a size up

from $10\text{ }\mu\text{m}$ up to $200\text{ }\mu\text{m}$, the f_{max} only drops by 34%. Furthermore, a 1000 fin device can be compared to a $110\text{ }\mu\text{m}$ 40 nm transistor, for practical purposes, and this makes the sizing problem even worse.

To analyze the f_{max} performance of both technologies a bit better, the necessary parameters from Eq. (2.10) are extracted from the transistors. From the small-signal model in Fig. 2.2, the gate resistance and other values can be derived from the Y-parameters as follows: (as long as $\omega^2(C_{GS} + C_{GD})r_G^2 \ll 1$) [12]

$$r_G = \frac{\Re(Y_{11})}{\Im(Y_{11})^2} \quad (2.14)$$

$$r_O = \left. \frac{1}{\Re(Y_{22})} \right|_{\omega=0} \quad (2.15)$$

$$C_{GD} = \frac{-\Im(Y_{21})}{\omega} \quad (2.16)$$

$$C_{GS} = \frac{\Im(Y_{11}) + \Im(Y_{12})}{\omega}. \quad (2.17)$$

The gate and output resistance and the gate–source capacitance for increasing transistor widths are shown in Fig. 2.9. To compare the planar 40 nm technology, where the total width is expressed in μm , to a 16 nm FinFET technology, where the total width is expressed as a number of fins, their drain–source current is used as a reference for total width. Both 40 nm and 16 nm transistors are biased and sized for optimal f_{max} . Besides the extracted values, the theoretical inversely proportional trend for gate and drain resistance when increasing the size of the transistor is added as a reference. For a certain current, both technologies show the

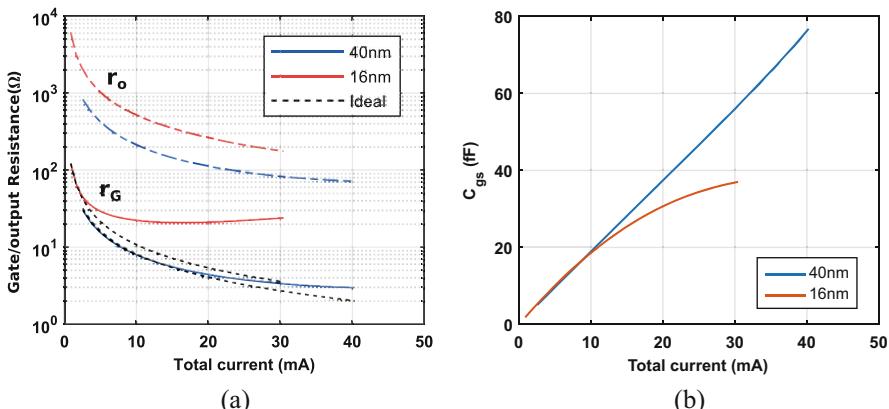


Fig. 2.9 (a) The gate and output resistance and (b) C_{GS} for an optimally sized transistor over drain current for both 40 nm and 16 nm. The dashed line represents the ideal scaling trend of gate resistance

same parasitic capacitance. When scaling, the reduction in the difference between both technologies seems to come from the gate resistance. The gate resistance of the 16 nm transistor is a lot larger than the 40 nm transistor. However, the better channel control of the 16 nm finFET device results in larger output resistance, compared with the planar 40 nm transistor. The increased output resistance partially compensates for the worse gate resistance. When the size of the transistor is increased to handle more current, the gate resistance should drop inversely proportionally with the current, as indicated in Fig. 2.9 by the dashed line. This holds true for both technologies as long as the transistor remains fairly small. The gate resistance of the 16 nm transistor quickly deviates from this trend and shows a larger gate resistance than expected. It is this quick deviation from the trend that resulted in the rapid decrease in f_{max} in Fig. 2.8.

Although a small 16 nm FinFET transistor is comparable in mm-wave performance to a transistor in an older planar 40 nm technology, the size of the transistor cannot be increased without rapid loss of performance. Furthermore, the transistor needs to be connected to the passive devices to make a useful mm-wave circuit. The passive devices, discussed in Sect. 2.2, are designed in the topmost layers where the metals are thicker and have lower resistivity. The transistors interconnect layout will thus incorporate vias from the substrate up to the top metals. The extra connectivity of the transistor needs to be designed with the least amount of resistance and capacitance to minimize its impact on the performance of the transistor. To minimize the gate resistance of large transistors, the 16 nm FinFET transistor will be further discretized into a single unit transistor (UT) of 6 fins and 15 fingers [13]. These unit transistors can be connected in the thicker top metals to mitigate the otherwise high gate resistance. Figure 2.10 shows the layout metalization of a single unit transistor. The layout of transistors interconnect is such that it can easily be tiled together to form larger transistors. The gate of the transistor is connected on both sides and combines to lower thin metal layers to reduce the resistivity. The gate is brought to one side and moved upward toward the thicker metal layers. A low resistive source connection is achieved by combining 3 x-metal layers to make a thicker

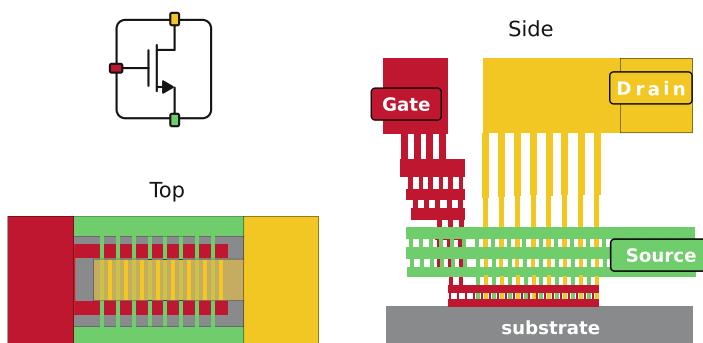


Fig. 2.10 The layout of a unit transistor

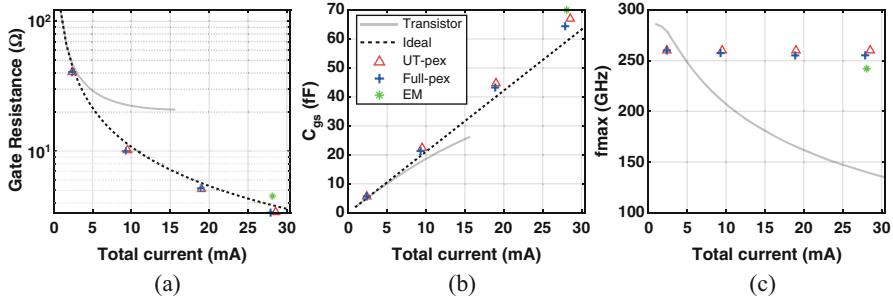


Fig. 2.11 The simulated performance comparison of a 16 nm transistor with and without unit transistors for (a) r_G , (b) C_{GS} , and (c) f_{max}

metal, connected from both sides, going over the metal gate connection. The drain is brought straight up and connected opposite to the gate with thicker metal.

Figure 2.11 compares the performance of creating a larger transistor by increasing the number of fingers, to the discretized approach of combining multiple unit transistors in the top metal layers. The discretized approach shows a reduced gate resistance compared to a single larger transistor and even follows the expected inversely proportional trend. Combining UT transistors will result in a slightly increased C_{GS} , compared to a single large transistor due to the increased area. The difference is minimal and is a fair price to pay for better gate resistance. The extra interconnects drop the f_{max} with 20 GHz, but combining multiple UTs only shows a minor reduction in f_{max} . The only downside of employing UTs is an increased area overhead, compared with a single large transistor.

2.2 Passives

In addition to the active components, a variety of passive components can be created in the BEOL. As the frequency increases, the passives become a more important part of the design and more time is spent designing passives than designing actives. It is, therefore, crucial to have a look at these components and what effect scaling has on them. The evolution of scaling on the BEOL is shown in Fig. 2.12. First of all, a designer has some choice in the desired BEOL. A digital design mainly requires a lot of metal layers, to make it easier for the tools to make all the connections, and a few thicker metals to create the power grid. However, the RF and mm-wave designer is mainly interested in a thick, low loss, metal layer for inductors and transformers far away from the substrate. This desire is fulfilled by the option of an ultra-thick metal (UTM) in some BEOLs. A digital and RF BEOL is shown in Fig. 2.12 for a 28 nm technology. The metal layers in a PDK are often referred to by letters, where x-metals represent the thin metals close to the substrate, y-metals the intermediate thicker metal, and Z the thickest non-ultra-thick metal layer.

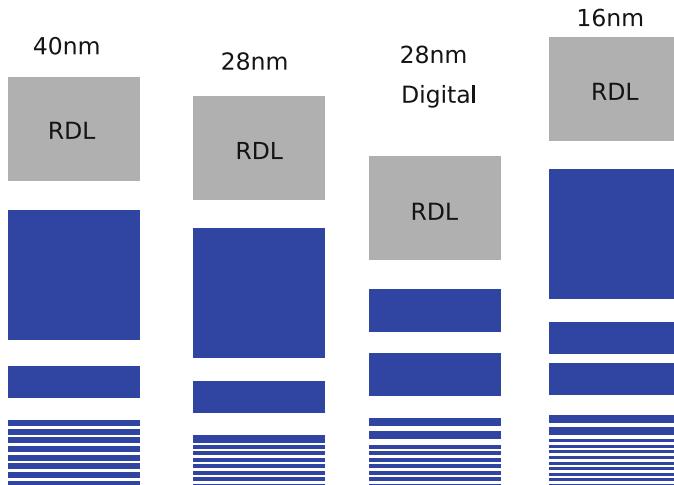


Fig. 2.12 A comparative representation of the scaling on the BEOL, showing a 40 nm, 28 nm RF, 28 nm digital, and 16 nm metal stack

2.2.1 Capacitors

Capacitors in a CMOS process can be divided into two main categories: fixed capacitors and voltage-controlled capacitors. Adjustable capacitors are made using the active layers. Fixed capacitors are created in the BEOL and will be focused on here. There are two types of fixed capacitors available: the metal–oxide–metal capacitor (MOM) and the metal–insulator–metal (MIM) capacitor. The difference is a MIM cap that uses a special ultra-thin high-K dielectric between 2 top metals, to create a high density and low loss capacitor. The downside is that this requires extra process steps and comes with an extra cost. A MOM cap on the other hand uses the available metals and oxides to create a capacitor. The two most widely used MOM capacitors are a multi-finger capacitor and a parallel plate capacitor, as shown in Fig. 2.13. The multi-finger capacitor is a design often provided by the PDK and creates most of its capacitance between two metals on the same layer. In this case, the minimum distance is limited by the lithography process of the BEOL and provides a smaller spacing than between 2 metal layers. The parallel plate capacitor creates the capacitance between the 2 metal layers. Because the spacing between metal layers is larger than the minimum spacing of the lithographic process, this will result in a lower density capacitor. The implemented capacitors are not ideal and show some losses when being charged and discharged. These losses can be represented by a resistor, R_s , in series with the capacitor. The efficiency of the capacitor at a certain frequency is given by its quality factor Q. The capacitor characteristics can be extracted from the simulated Y-parameters as follows:

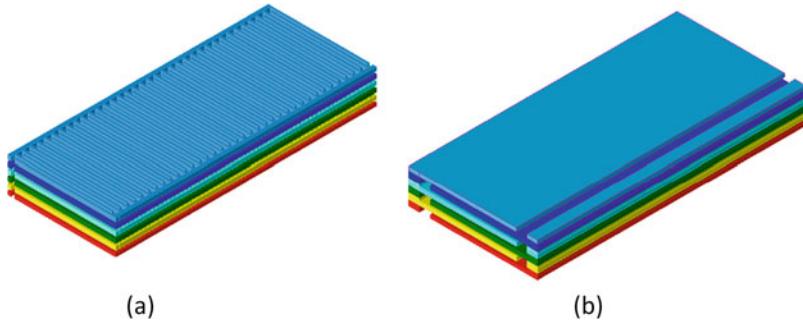


Fig. 2.13 (a) A multi-finger and (b) parallel plate MOM capacitor

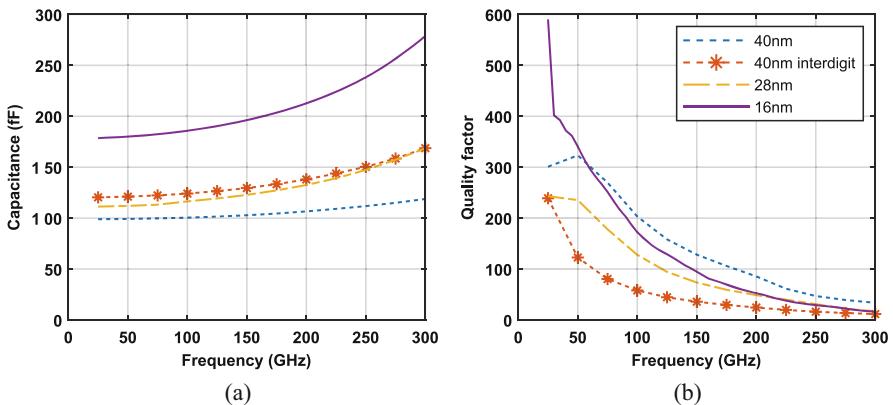


Fig. 2.14 The effect of scaling on an equal area capacitor showing (a) capacitance and (b) quality factor

$$C_{cap} = \frac{\Im(Y_{11})}{\omega} \quad (2.18)$$

$$Q_{cap} = \frac{\Im(Y_{11})}{\Re(Y_{11})} = \frac{1}{\omega C_{cap} R_s}. \quad (2.19)$$

A parallel plate MOM capacitor is designed and simulated in three technologies: 40 nm, 28 nm, and 16 nm, alongside a multi-finger capacitor in 40 nm for the same area. Figure 2.14 shows their capacitance and quality factor over frequency. The quality factor of all capacitors degrades fast when operating at higher frequencies, making a multi-finger capacitor less suited for mm-wave operations. The interdigit capacitor has the largest capacitive density but comes with a severe degradation in quality compared to a parallel plate capacitor. The low-quality factor is due to the increased resistive losses in the long and thin fingers at mm-wave frequencies. Because of the bad quality of interdigit capacitors and neglectable area penalty, only parallel plate capacitors are used in the included mm-wave designs. Between

technologies, the quality drops as the metal layers get thinner, but the density increases as metal layers move closer together. The increased capacitance of the 16 nm technology comes from a larger number of thin closely spaced metal layers (x-metals), as can be seen in Fig. 2.12, available to make the capacitor.

2.2.2 Inductors

Inductors are commonly used in RF circuits to resonate with (parasitic) capacitors. As the frequency of RF circuits increases, inductors could be implemented on-chip [14]. Low-frequency designs require large inductances and are implemented as multi-turn inductors. For higher frequency mm-wave designs, smaller inductances are required, and single-turn inductors become more prevalent. An ideal inductor would return all of the stored energy in the magnetic field without any losses. Unfortunately, all real implemented inductors show losses that can be modeled as a resistor, R_S , in series with the inductor. This resistor represents multiple loss mechanisms such as resistive losses in the metal traces, eddy current in the substrate, and skin effect [14, 15]. The efficiency of an inductor at a certain frequency is quantified by its quality factor Q and defined as follows:

$$Q = \frac{\omega L}{R_S}. \quad (2.20)$$

In most cases, a high-quality inductor is desired as this will result in the lowest losses and improve the efficiency of the circuits designed with it. From formula 2.20, the quality should improve moving to higher frequencies. However, a lot of loss mechanisms are more detrimental at higher frequencies. Due to the skin effect and current crowding, the current is pushed more to the edge of the inductors increasing the effective resistance and increasing the losses. This effect is an argument to not use wide metal traces for mm-wave inductors. Another frequency-dependent loss are the eddy currents in both the metal dummies and the substrate. The three values for a simplified series-inductor model can be extracted from the simulated Z-parameters using the following expressions:

$$L_S = \frac{\Im(Z_{11})}{\omega} \quad (2.21)$$

$$Q = \frac{\Im(Z_{11})}{\Re(Z_{11})} \quad (2.22)$$

$$R_S = \Re(Z_{11}). \quad (2.23)$$

The effect of the BEOL is demonstrated on the same inductor with a diameter of 40 μm and a trace width of 4 μm . The inductance and quality factor of the same inductor at different metal layers are shown in Fig. 2.15. This simulation demon-

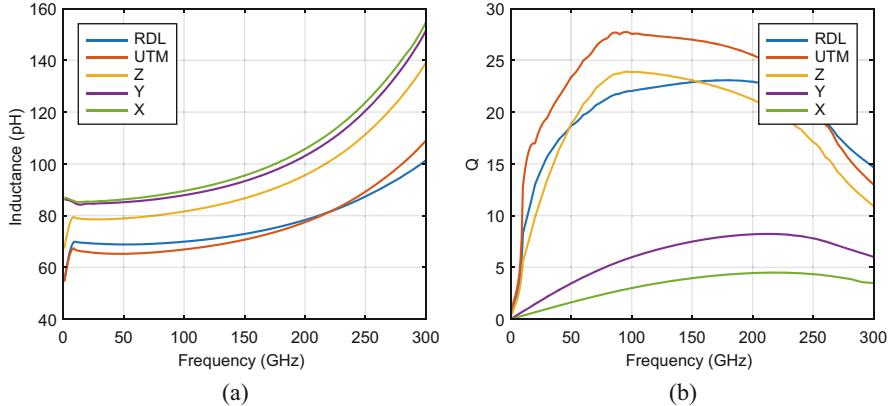


Fig. 2.15 The inductance and quality factor of the same inductor dimension at different metal layers in a 28 nm BEOL

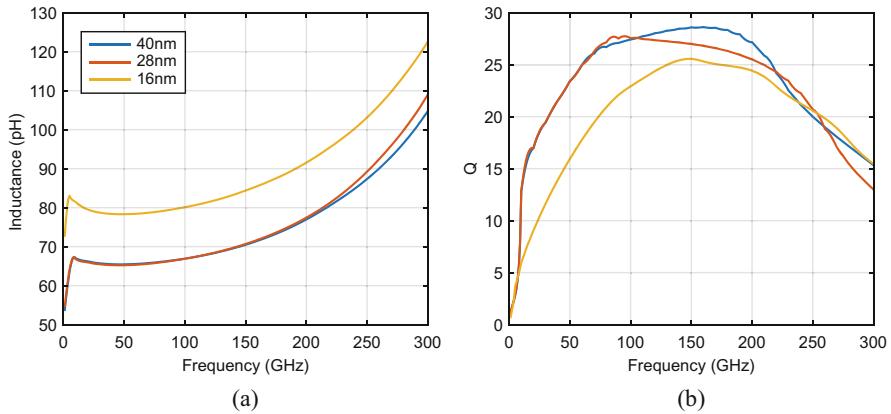


Fig. 2.16 The inductance and quality factor of the same inductor dimension in 3 different BEOLs

strates the benefit of an ultra-thick metal layer in the BEOL, compared with using the Z-layer or RDL for inductors. Another interesting observation is the similar quality of using the non-polished RDL compared to the Z-layer, both showing very similar quality despite one being closer to the substrate. This can be explained by a larger sheet resistance of the UTM due to surface roughness because the layer is unpolished. For capacitors, a clear difference is seen between technologies as the bottom metal layers get closer together. However, both inductance and quality factor stay the same between the planar technologies, and they show only a slight difference with the 16 nm finFET technology as depicted in Fig. 2.16.

2.2.3 Transformers

A transformer is formed when two inductors are brought together and coupled magnetically, allowing energy to be transferred between both coils. A transformer is one of the most used passive components in mm-wave design as it allows to achieve matching between amplifier stages [16]. Furthermore, the transformer provides the designer with additional benefits: DC isolation, a possibility to set the DC without impeding the RF signal, and it can act as a balun. Because the transformer consists of two inductors, the same loss mechanisms and BEOL effects play as seen for inductors. The most simple model represents the transformer with two inductors, with inductances L_p and L_s with their respective losses represented by Q_p and Q_s . The coupling is represented by k_m , which goes from 0 to 1 or from no to full coupling, respectively. From the Z-parameters, the characteristics of the transformer can be calculated using the following equations [15]:

$$L_p = \frac{\Im(Z_{11})}{\omega} \quad (2.24)$$

$$L_s = \frac{\Im(Z_{22})}{\omega} \quad (2.25)$$

$$Q_p = \frac{\Im(Z_{11})}{\Re(Z_{11})} \quad (2.26)$$

$$Q_s = \frac{\Im(Z_{22})}{\Re(Z_{22})} \quad (2.27)$$

$$k_m = \sqrt{\frac{\Im(Z_{12})\Im(Z_{21})}{\Im(Z_{11})\Im(Z_{22})}}. \quad (2.28)$$

2.2.3.1 Transformer Topology

To use transformers freely in any design, we would like to be able to choose the three parameters L_p , L_s , and k_m independently while maximizing the quality factor. Unfortunately, this is not the case as increasing the difference in inductance between primary and secondary will result in a small and large inductor, resulting in a lower coupling factor. Depending on the required inductance ratio and coupling, a different on-chip topology can be used. For an inductance ratio around one, nothing special is necessary, and a n-to-m transformer can be implemented with n and m the same number of turns and spaced according to the required coupling. For mm-wave designs, the number of turns is limited to one or two, as extra turns result in complexity and extra resistive losses, given that the separate coils need to be connected by a bridge. With both coils on top of each other, maximum coupling is achieved. The coupling can be lowered by shifting the two inductors and decreasing their overlapping area.

Table 2.1 Transformer topology comparison for a 2.7 inductance ratio

	One-to-one	One-to-one with leakage	Two-to-one reduced coupling	Two-to-one tightly coupled
L_p (pH)	75.5	76.4	88	91
L_s (pH)	26.8	28	32	31
L_P/L_s	2.82	2.72	2.75	3
k_m	0.2	0.45	0.46	0.58
Q	19	14	13.5	13
IL (dB)	2.2	1.3	1.4	1.1

For inductance ratios of more than one (or equivalently less than one), different topologies can be used. Three transformer topologies are compared trying to achieve the same primary and secondary inductance with a ratio of 2.7, for a high coupling since the low coupling is always possible. The characteristics and layout of the discussed topologies are shown in Table 2.1 and Fig. 2.17, respectively.

The first topology is a classical one-to-one transformer; to get a high turn ratio, a large difference in diameter between the two inductors is necessary. The increased distance between the coils will lead to a reduced coupling as shown in Fig. 2.17. The highest possible coupling for this layout is 0.2, which also results in a high IL of 2.2 dBm despite the good Q.

The second transformer uses a tightly coupled one-to-one transformer and adds leakage inductance on one side, to increase the inductance ratio as shown in Fig. 2.18. The design can be simplified to a regular transformer, where the primary inductance is the sum of the tightly coupled transformer and the leakage inductor $L_p = L_1 + L_2$. However, the added leakage inductor will reduce the coupling. The relation between the new inductance ratio and coupling can be expressed as follows:

$$\frac{L_p}{L_s} = \frac{k_H^2}{k_m^2}. \quad (2.29)$$

This sets the maximum coupling achievable, k_m , by the maximal starting coupling k_H (0.8) and the inductance ratio L_p/L_s . The starting transformer has an inductance of 24 pH with a Q of 16 and a coupling of $k_H = 0.75$. If the ratio is increased to 2.7, an equivalent coupling of 0.46 is calculated, and this is already larger than just using differently sized coils. In the design, two leakage inductors are added; they bring the inductance up to 76 pH and reduce the coupling down to 0.45, which is close to the calculated. This demonstrates a technique to increase and predict the coupling for larger inductance ratios without using double windings as shown next.

The third transformer increases the inductance by employing a two-turn inductor. This increases the inductance and keeps the diameter smaller, which improves the coupling to the other single-turn inductor. This is a great topology for tightly coupled

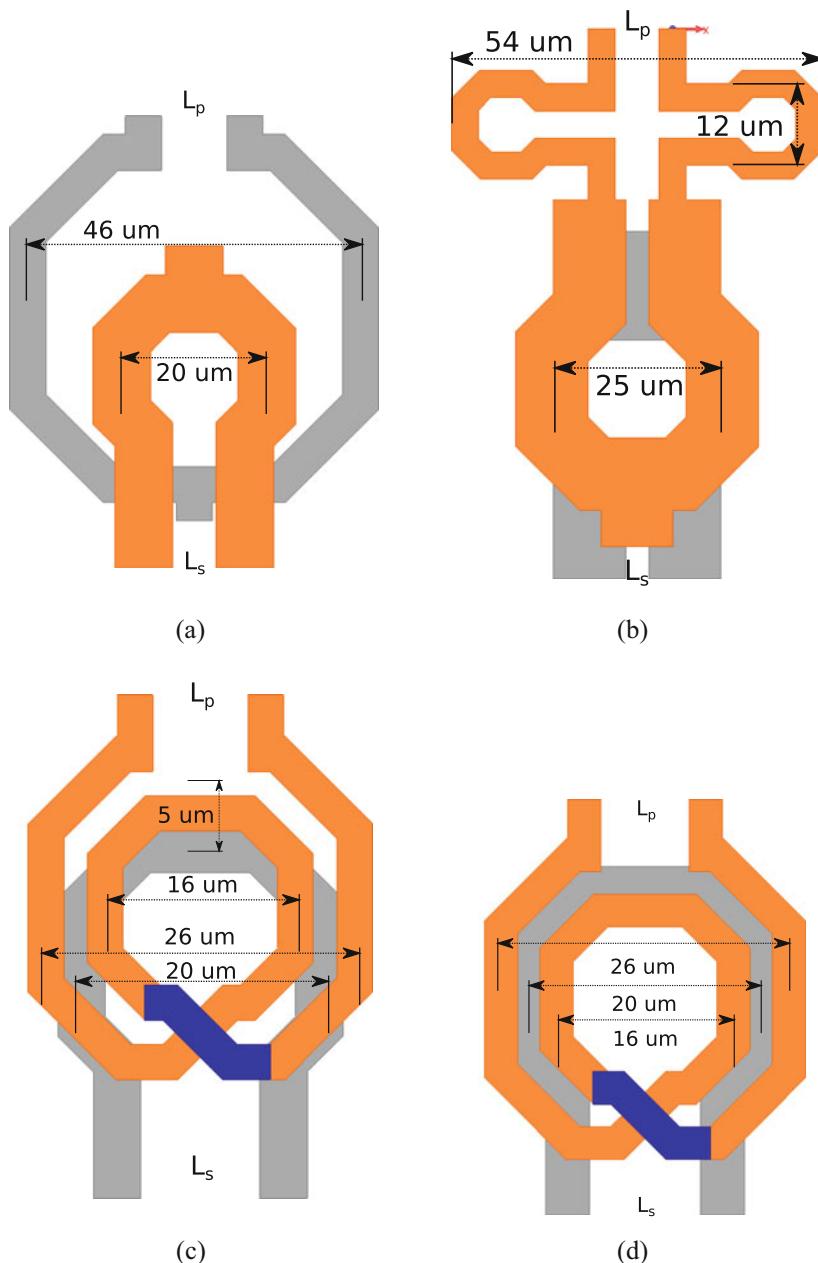


Fig. 2.17 The layout of the 4 transformers in the topology example: **(a)** an one-to-one transformer, **(b)** a tightly coupled one-to-one transformer with leakage inductance in series, **(c)** a two-to-one transformer with lowered coupling by lateral shift, and **(d)** a tightly coupled two-to-one transformer

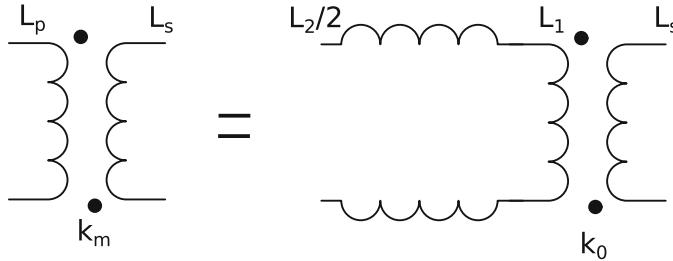


Fig. 2.18 A transformer and its tightly coupled equivalent with series leakage inductance

high inductance ratios. The two turns, however, come with extra complexity and vias that decrease the Q of the inductor. Furthermore, for frequencies above 100 GHz a two-turn inductor can quickly provide too much inductance. For comparison, the coupling is reduced down to the same coupling of the second transformer, by shifting the two inductors further apart as shown in Fig. 2.17.

To summarize: for increasing inductance ratios above one, a classical one-to-one transformer can be used for low coupling. If medium coupling is required, the coupling can be improved by combining a tightly coupled transformer with a series leakage inductance. If this still does not provide the required coupling, multi-turn inductors are necessary to get the high coupling and a high inductance ratio.

2.2.4 Transmission Lines

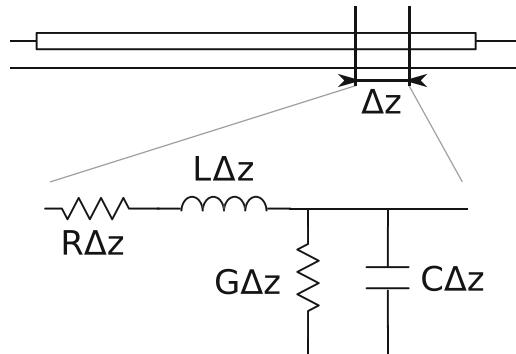
With the decreased wavelength at mm-wave frequencies, the lengths of interconnects are in the same order of magnitude as the wavelength. The interconnects can therefore not be represented as simple RC interconnects, but need better modeling. The transmission line(TL) is a high-frequency interconnect model using a lumped-element LC model, as shown in Fig. 2.19. The TL can now be characterized by: the characteristic impedance Z_0 , the wave velocity v, and the propagation constant γ . The propagation constant consists of a real part, α , the attenuation constant, and an imaginary part, β , the phase constant. These parameters are given as follows [17], where simplifications are applied for a lossless TL with $R=0$ and $G=0$:

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \approx \sqrt{\frac{L}{C}} \quad (2.30)$$

$$v = \lambda f \approx \frac{1}{\sqrt{LC}} \quad (2.31)$$

$$\gamma = \alpha + j\beta \quad (2.32)$$

Fig. 2.19 A transmission line and its distributed lumped equivalent circuit



$$\beta = \frac{2\pi f}{v} = \omega \sqrt{LC}. \quad (2.33)$$

The transmission lines are made in the top metals, such as inductors, and therefore show limited changes with scaling technology.

2.3 Basic Design of mm-Wave Circuits

In the previous sections, an overview of the actives and passives was given. Now, let us take a brief look at how an active transistor can be used to make a single amplifier and how multiple of these active amplifiers need to be connected.

2.3.1 A Capacitive Neutralized mm-Wave Amplifier

The capacitive neutralized pseudo-differential amplifier is the most commonly used amplifier in CMOS mm-wave designs, especially above 100 GHz. Figure 2.20 shows the pseudo-differential amplifier with neutralization capacitors, C_n . The 2 extra capacitors neutralize the gate-drain capacitor by creating a negative capacitance using the available differential signal.

The effects of C_n on MSG and MAG over frequency are shown for 3 capacitor values in Fig. 2.21a. The neutralization of the gate-drain capacitance leads to an improved differential mode stability, without the need of introducing losses. The intrinsic differential stability of this amplifier can be analyzed using the Rollet-stability factor [17]. The stability factor needs to be larger than 1, but the value itself does not hold any indication of stability. Instability always needs both gain and a form of feedback; the lower the gain, the more feedback is required and vice versa. Because of the reduced gain at a higher frequency, the amplifier is already unconditionally stable from a certain frequency, f_{stab} , and up. f_{stab} will

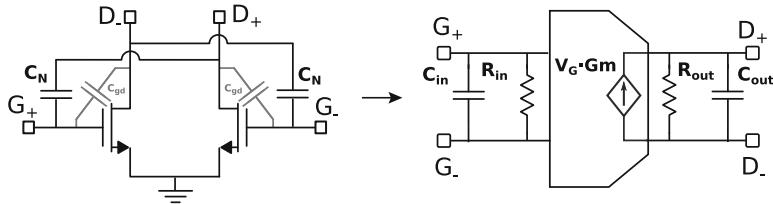


Fig. 2.20 The schematic of a capacitive neutralized pseudo-differential common-source amplifier and its simplified circuit model at the operating frequency

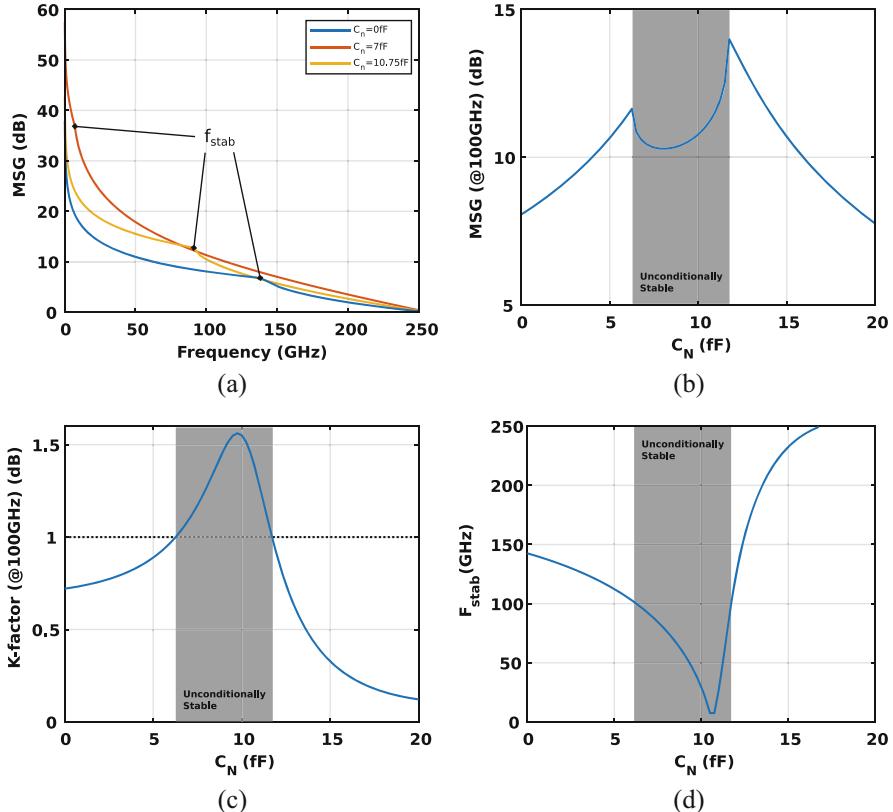


Fig. 2.21 (a) The MSG over frequency. (b) The MSG and (c) stability factor at 100 GHz together with (d) f_{stab} for increasing neutralization capacitance

be a more interesting value to look at when choosing the right neutralization capacitor. Besides the improved stability, the neutralization capacitor also increases the intrinsic gain of the amplifier. Figure 2.21 summarizes the gain and stability factor at 100 GHz and the minimum stable frequency for a capacitive neutralized amplifier, using two transistors with 4 unit transistors each. The amplifier shows

unconditional stable behavior for a neutralization capacitor from 6.25 fF to 11.75 fF. Although the maximum gain is achieved for a neutralization capacitor of 11.75 fF, it is right on the edge of unconditional stable behavior. A small perturbation on the neutralization capacitor or transistor, either by process variation or modeling error, can push the amplifier in conditional stable behavior and could lead to undesired stability problems. It also shows that as the operating frequency increases, the amplifier can tolerate more variation to its neutralization capacitor for unconditional stable behavior. A good trade-off between safety margin and gain is by taking the neutralization capacitor for the lowest minimum stable frequency.

The capacitive neutralized amplifier can be simplified around the operating frequency by a simplified circuit model as shown in Fig. 2.20. The input and output impedances are used to design the matching circuits.

2.3.2 Design of mm-Wave Interconnects

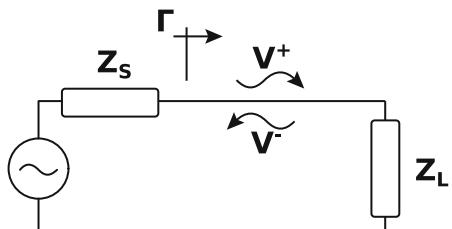
The increasing operating frequency results in a decreasing wavelength. For mm-wave circuits, the wavelength becomes so small, and the wave characteristics cannot be ignored anymore when designing on-chip interconnects. When a signal source, with impedance Z_s , is connected to a load with load impedance Z_L as shown in Fig. 2.22, the source will transmit a wave V^+ . An impedance mismatch between load and source, at the interface, will cause part of the incident wave to be reflected as V^- . The reflection coefficient, Γ , is given as the ratio between the reflected wave and the incident wave and can be translated to the impedance imbalance as follows: [17]

$$\Gamma = \frac{V^-}{V^+} \quad (2.34)$$

$$\Gamma = \frac{Z_L - Z_S}{Z_L + Z_S}. \quad (2.35)$$

A zero reflection coefficient means that no wave is reflected and a perfect power transfer is happening between source and load. The optimal source and load impedance for power transfer is called the conjugate match and is given as follows:

Fig. 2.22 A simple source and load impedance circuit with the forward and reflected waves



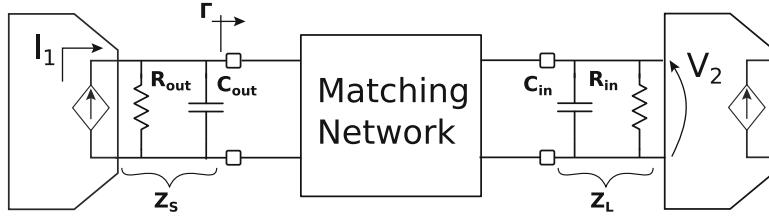


Fig. 2.23 The schematic overview of matching between two amplifier stages

$$Z_L = Z_S^* \quad (2.36)$$

$$R_L = R_S, X_L = -X_S. \quad (2.37)$$

The above matching assumes the load impedance can be chosen freely. Unfortunately, this is rarely the case. Let us apply the matching theory to the connection between 2 differential amplifier stages as shown in Fig. 2.23. The source impedance Z_S is now the parallel output resistor and capacitor with the transconductance as a current source. The load impedance is a parallel impedance of both the input capacitor and resistor, of the next stage. Between both stages, passive components act as a matching network that ideally would not provide any extra losses. The matching network transforms the load impedance into the ideal conjugate impedance for the source, to maximize power transfer and minimize reflections. To maximize the gain of the next stage, the voltage, V_2 , should be maximized at its gate. This translates to a maximization of the Z_{21} transfer function. If both impedances are the same, $R_{out} = R_{in}$, the matching network can consist of a single inductor, to tune out the total capacitance, leading to a conjugate match and a $Z_{21} = R_{in}/2$.

The input and output impedances are not the same for a lot of circuits, for example, input–output interfaces and the interface between a driver and a PA. A simple inductor can tune out the capacitors, but the unequal resistance, $R_{out} \neq R_{in}$, will lead to reflections and low $Z_{21} = R_{out}/R_{in}$. A matching network needs to be inserted, to transform the load impedance into the required conjugate impedance. The conjugate match will at the same time also maximize the transimpedance gain to $\max(|Z_{21}|) = \sqrt{(R_{out} \cdot R_{in})}/2$.

So far nothing has been said about the frequency response of the matching network. Matching networks have been studied extensively, and this has led to the Bode–Fano expression [18, 19]:

$$\int_0^{+\infty} \ln \left(\frac{1}{|\Gamma(\omega)|} \right) d\omega \leq \frac{\pi}{RC}. \quad (2.38)$$

The Bode–Fano expression relates the bandwidth of the well-matched frequency band to the RC product. This has several implications: (1) that a high-Q load is more difficult to match over a wide bandwidth than a low-Q load at the same frequency,

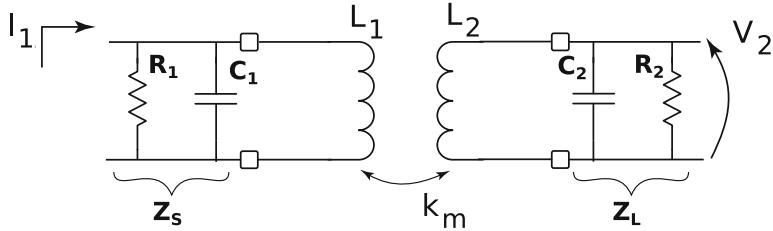


Fig. 2.24 The schematic of a transformer matching network: a magnetically coupled resonator

(2) the smaller the in-band ripple, the smaller the bandwidth will be, and (3) the better the match, the smaller $|\Gamma|$ at any frequency, the lower the resulting bandwidth.

The Bode–Fano expression provides an upper bound for the achievable bandwidth but says nothing about its implementation. One of the most widely used passives for matching in CMOS technology is the transformer.

2.3.2.1 Transformer Matching: A Coupled RLC Resonator

Two RLC networks can be coupled together in various ways; one of the most popular techniques uses the magnetic coupling between two inductors, better known as a transformer. Besides the ability to provide a compact 4th order matching network with a small footprint, the transformer provides a dc isolation between source and load, and a voltage can be applied to the center tap for differential circuits. Figure 2.24 shows the basic schematic of the magnetically coupled resonators. From this, the transimpedance Z_{21} of the network can be found as follows [20–25]:

$$Z_{21} = \frac{R_1 R_2 k_m \sqrt{L_1 L_2} s}{A \cdot s^4 + B \cdot s^3 + C \cdot s^2 - D \cdot s - R_1 R_2} \quad (2.39)$$

with

$$A = C_1 C_2 L_1 L_2 R_1 R_2 (k^2 - 1)$$

$$B = L_1 L_2 (C_1 R_1 + C_2 R_2) (k^2 - 1)$$

$$C = ((k^2 - 1) L_1 L_2 - R_1 R_2 (C_1 L_1 + C_2 L_2))$$

$$D = (L_1 R_2 + L_2 R_1).$$

From expression (2.39), assuming a high-quality factor or large R_1 and R_2 , the two angular resonance frequencies can be calculated as follows:

$$\omega_{L,H}^2 = \frac{1 + X \pm \sqrt{1 + X^2 + X(4k_m^2 - 2)}}{2L_2C_2(1 - k_m^2)} \quad (2.40)$$

with

$$X = \frac{L_2C_2}{L_1C_1}. \quad (2.41)$$

For the simple case where the resonator is balanced: $R_1 = R_2$, $C_1 = C_2 = C$, and $L_1 = L_2 = L$. The two frequencies can be simplified as follows:

$$\omega_L = \frac{1}{\sqrt{LC(1 + |k_m|)}} \quad (2.42)$$

$$\omega_H = \frac{1}{\sqrt{LC(1 - |k_m|)}}. \quad (2.43)$$

The amount of coupling will determine the spacing between the two resonances and, thus, also the bandwidth of the matching network. To achieve a certain bandwidth, the low and high resonance frequencies are set by the mutual coupling and can be expressed as follows:

$$k_m = \frac{\omega_H^2 - \omega_L^2}{\omega_H^2 + \omega_L^2}. \quad (2.44)$$

Figure 2.25a shows the effect of changing the coupling factor when all other parameters are kept the same. An increasing coupling will result in increased bandwidth. However, as the bandwidth increases so does the in-band ripple, this is in line with the Bode–Fano limit in Eq. (2.38). To reduce the ripple for a certain bandwidth, the quality factor needs to be reduced, as shown in Fig. 2.25b. The Q-factor of the input and output RC is reduced by decreasing the capacitor, resulting in the same peak $|Z_{21}|$ value. In practice, the quality factor of the source or load impedance can only be reduced by adding resistance. This added resistance will reduce the peak $|Z_{21}|$ and result in a reduced gain.

The simplification of a balanced resonator allows for easy analysis and to look at the effects of coupling and quality factor. The assumption, however, does not show how the required impedance transformation can be achieved. To achieve an n-fold impedance transformation such that $R_1 = nR_2$, the transformer needs to get an equal inductance ratio: $L_1 = nL_2$. To keep the resonant frequencies fixed, the capacitance ratio needs to change in an opposite way to the inductance $C_1 = C_2/n$. For CMOS transistors and amplifiers, the RC product stays constant independent of the transistor size. Therefore, no extra capacitance needs to be added to perform an impedance match between amplifier stages. At the interface, input or output pads, if the pads do not provide the necessary capacitance, extra capacitance will need to be added to be able to use a transformer impedance transformation.

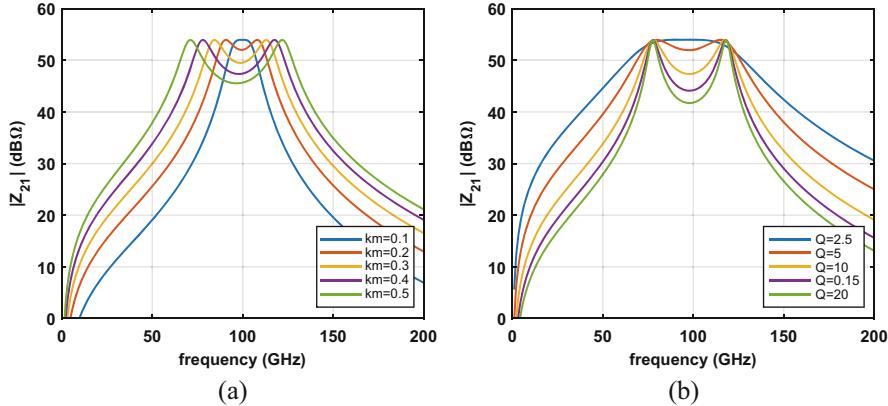


Fig. 2.25 The effect of (a) coupling factor k_m for a fixed $Q=10$ and (b) source/load quality for a fixed $k_m=0.4$ on the matching Z_{21} transfer function

So far the analysis assumed an ideal transformer with no losses. An implemented transformer will however come with its intrinsic losses expressed as series resistance or the coils Q-factor. The transfer losses from one coil to the other now depend not only on the Q-factor, but also on the coupling [26, 27]. The insertion loss (IL) of a transformer can be expressed as follows:

$$IL = 20 \log_{10} \left(\frac{k_m Q}{-1 + \sqrt{1 + (k_m Q)^2}} \right) \quad (2.45)$$

with

$$Q = \sqrt{Q_p Q_s}. \quad (2.46)$$

Figure 2.26 shows the IL for transformers against their coupling factor at different quality factors. A low insertion loss can be achieved with low coupling, as long as high-quality inductors are used, or with low-quality inductors as long as a high coupling is achieved.

The quality factor of transformers at mm-wave frequencies at 100 GHz and above is around 20. This implies that using a coupling factor from 0.2 to 0.5, to perform wideband matching, results in an increased IL of 1.6 dB to 0.54 dB. The increased insertion loss is bad, especially with the already low gain at 100 GHz and above but can be accepted for interstage matching in return for enhanced bandwidth. However, at the output of PAs (or input of LNAs), the increased IL is detrimental to the overall performance and a high coupling transformer is desired. The high coupling will put the two resonance frequencies so far apart that only the lower resonance is used to perform the matching network as shown in Fig. 2.27.

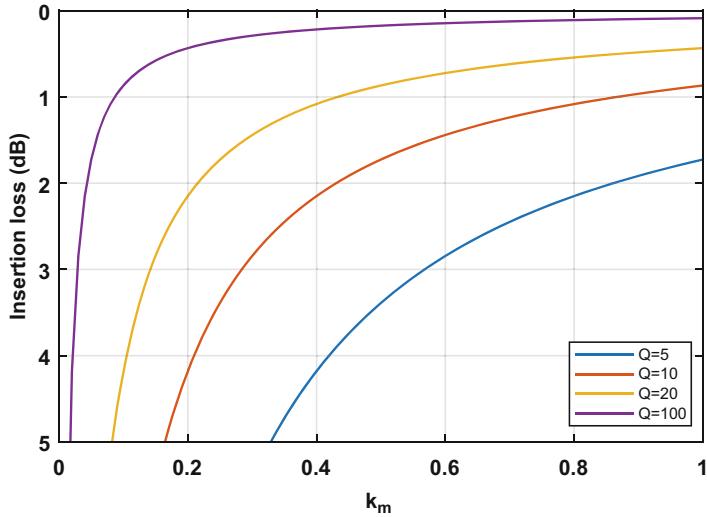


Fig. 2.26 The IL of a magnetically coupled transformer against its coupling factor for different quality factors

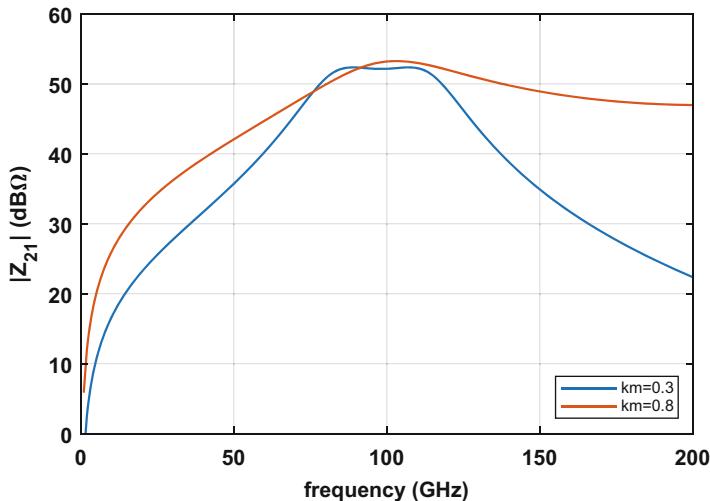


Fig. 2.27 The difference in $|Z_{21}|$ between a low-k, 2 resonance matching, and a high-k, 1 resonance matching

2.4 Conclusion

This chapter has focused on the basic active and passive components of mm-wave CMOS circuits. The effect of technology scaling on these components is discussed and its implications for mm-wave designs. Furthermore, the effects on the design

when transitioning from a planar to a FinFET technology for mm-wave actives are shown.

The active transistors are combined into an unconditional stable amplifier. A method to combine and match multiple of these stages together using transformers is shown. An in-depth analysis of the trade-offs in using transformers for interconnect matching is illustrated.

References

1. W.M.C. Sansen, *Analog Design Essentials*. Springer, Berlin (2006)
2. G.E. Moore, Cramming more components onto integrated circuits. *Electronics* **38**(8), 114 (1965)
3. Wikipedia, transistor count-Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Transistor_count. Accessed: 2020-06-12
4. H.P. Wong, K. Akarvardar, D. Antoniadis, J. Bokor, C. Hu, T. King-Liu, S. Mitra, J.D. Plummer, S. Salahuddin, A density metric for semiconductor technology [point of view]. *Proc. IEEE* **108**(4), 478–482 (2020a)
5. H.P. Wong, K. Akarvardar, D. Antoniadis, J. Bokor, C. Hu, T. King-Liu, S. Mitra, J.D. Plummer, S. Salahuddin, A density metric for semiconductor technology [point of view]. *Proc. IEEE* **108**(4), 478–482 (2020b)
6. P. Ye, T. Ernst, M.V. Khare, The last silicon transistor: Nanosheet devices could be the final evolutionary step for Moore's law. *IEEE Spectr.* **56**(8), 30–35 (2019)
7. Intel, Intel's revolutionary 22 nm transistor technology (2011). http://download.intel.com/newsroom/kits/22nm/pdfs/22nm-Details_Presentation.pdf. Copyright © Intel Corporation, Accessed: 2020-06-15
8. K.W.J. Chew, A. Agshikar, M. Wiatr, J.S. Wong, W.H. Chow, Z. Liu, T.H. Lee, J. Shi, S.F. Lim, K. Sundaram, L.H.K. Chan, C.H.M. Cheng, N. Sassiati, Y.K. Yoo, A. Balijepalli, A. Kumta, C.D. Nguyen, R. Illgen, A. Mathew, C. Schippel, A. Romanescu, J. Watts, D. Harame, RF performance of 28 nm PolySiON and HKMG CMOS devices, in *2015 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)* (2015), pp. 43–46
9. R.A. Wachnik, S. Lee, L.H. Pan, N. Lu, H. Li, R. Bingert, M. Randall, S. Springer, C. Putnam, Gate stack resistance and limits to CMOS logic performance, in *Proceedings of the IEEE 2013 Custom Integrated Circuits Conference* (2013), pp. 1–4
10. N. Lu, T.B. Hook, J.B. Johnson, C. Wermer, C. Putnam, R.A. Wachnik, Efficient and accurate schematic transistor model of FinFET parasitic elements. *IEEE Electron Device Lett.* **34**(9), 1100–1102 (2013)
11. G. Hueber, A.M. Niknejad (eds.), *Millimeter-Wave Circuits for 5G and Radar*. The Cambridge RF and Microwave Engineering Series. Cambridge University Press, Cambridge (2019)
12. I. Kwon, M. Je, K. Lee, H. Shin, A simple and analytical parameter-extraction method of a microwave MOSFET. *IEEE Trans. Microwave Theory Tech.* **50**(6), 1503–1509 (2002)
13. B. Philippe, P. Reynaert, A 126 GHz, 22.5% tuning, 191 dBc/HZ FOMt 3rd harmonic extracted class-F oscillator for D-band applications in 16 nm FinFET, in *2020 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)* (2020), pp. 263–266
14. J. Craninckx, M.S.J. Steyaert, A 1.8-GHz low-phase-noise CMOS VCO using optimized hollow spiral inductors. *IEEE J. Solid-State Circuits* **32**(5), 736–744 (1997)
15. R.P. Dixian Z., *CMOS 60-GHz and E-band Power Amplifiers and Transmitters* (Springer, Berlin, 2015)
16. A. Bevilacqua, Fundamentals of integrated transformers: From principles to applications. *IEEE Solid-State Circuits Mag.* **12**(4), 86–100 (2020)
17. D. Pozar, *Microwave Engineering Fourth Edition* (2005)

18. H. Bode, *Network Analysis and Feedback Amplifier Design* (D. Van Nostrand Company, New York, 1945)
19. R. Fano, *Theoretical Limitations on the Broadband Matching of Arbitrary Impedances*, vol. 249 (1950)
20. M. Dishal, Design of dissipative band-pass filters producing desired exact amplitude-frequency characteristics. Proc. IRE , 37(9), 1050–1069 (1949)
21. C.H. Li, Y.L. Liu, C.N. Kuo, A 0.6-v 0.33-mw 5.5-GHz receiver front-end using resonator coupling technique. IEEE Trans. Microwave Theory Tech. **59**, 1629–1638 (2011)
22. C.H. Li, C.N. Kuo, M.C. Kuo, A 1.2-V 5.2-mW 20-30-GHz wideband receiver front-end in 0.18- μm CMOS. IEEE Trans. Microwave Theory Tech. **60**, 3502–3512 (2012)
23. M. Babaie, R.B. Staszewski, A class-F CMOS oscillator. IEEE J. Solid-State Circuits, **48**(12), 3120–3133 (2013)
24. A. Mazzanti, A. Bevilacqua, On the phase noise performance of transformer-based CMOS differential-pair harmonic oscillators. IEEE Trans. Circuits Syst. I Regul. Pap. **62**(9), 2334–2341 (2015)
25. M. Vigilante, P. Reynaert, On the design of wideband transformer-based fourth order matching networks for *E*-band receivers in 28-nm CMOS. IEEE J. Solid-State Circuits **52**(8), 2071–2082 (2017)
26. T. Ohira, The kQ product as viewed by an analog circuit engineer. IEEE Circuits Syst. Mag. **17**(1), 27–32 (2017)
27. R.P. Vigilante M., *5G and E-Band Communication Circuits in Deep-Scaled CMOS* (Springer, Berlin, 2018)

Chapter 3

Frequency Generation



3.1 VCO Basics

A widely used view of a basic oscillator is the combination of a lossy resonator with a negative resistor [1]. If we have a LC-resonator as shown in Fig. 3.1a, a current impulse, $I_0\delta(t)$ given at time t, will result in a voltage $I_0\delta(t)/C$. Over time the capacitor and inductor in the resonator will periodically exchange energy and sustain an oscillation at frequency f_0 as follows:

$$f_0 = \frac{1}{2\pi\sqrt{LC}}. \quad (3.1)$$

Unfortunately, in practice, a lossless resonator does not exist, the passives used in the resonators have losses, and using the oscillating signal would extract energy from the resonator. These losses would result in a diminishing oscillation amplitude over time and end the oscillation. A solution to this problem is to use active components to constantly inject energy into the resonator to sustain the oscillation. Figure 3.1b represents the most simplified model of an active LC oscillator. All the losses are represented by the resistance R_T and are compensated by an equal but opposite active negative resistance. As described in Sect. 2.2, the losses of the passive components can be expressed as a quality factor. The total quality of the tank is as follows:

$$\frac{1}{Q_T} = \frac{1}{Q_C} + \frac{1}{Q_L}. \quad (3.2)$$

The negative resistance is created by placing an active amplifier in a positive feedback loop.

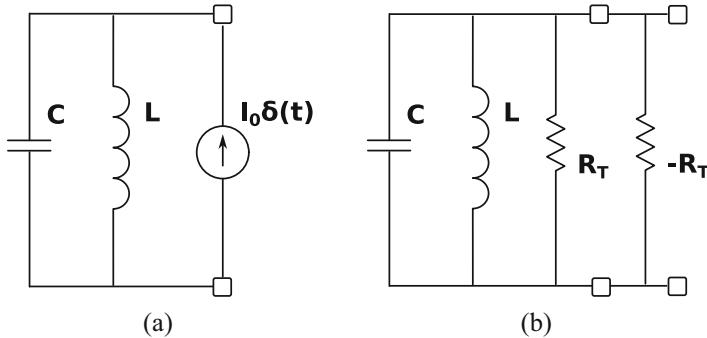


Fig. 3.1 (a) A lossless resonator and (b) its real-world implementation

3.1.1 Brief on Phase Noise

The ideal oscillator provides a single output tone at frequency f_0 . However, noise will affect the oscillator and disturb the output signal. The changes in output amplitude are rejected by the oscillator. On the other hand, changes in phase will accumulate. The changes in the output phase are of great importance to transmitter and receiver circuits. Phase noise, therefore, has been studied for years. Leeson was the first to describe the phase noise at an offset frequency from the carrier mathematically, called the Leeson's equation, and verified it experimentally [2]. This was further extended by many others [3, 4]. To model the effect of noise on the oscillator, an equivalent model was proposed. For an oscillator in steady state, the active $-R_T$ device can be replaced with a sinusoidal current source, and all noise sources are combined in a single injection source. The voltage across the tank when an instantaneous current is injected at two different times, t_1 and t_2 , in the tank is shown in Fig. 3.2. The moment of injection determines the effect of change on the output voltage. When the charge is injected on the peak at t_1 , the amplitude increases but the phase stays constant. The amplitude change is quickly rejected and no phase noise is created. In contrast, when a charge is injected at the crossing, t_2 , no change in amplitude is seen but a phase shift is created. This time variant sensitivity to injected charge is called the impulse sensitivity function (ISF) [3]. The total phase noise is thus the combination of the total injected noise current and the sensitivity to this injected noise. The study of different waveforms and oscillator topologies to improve the phase noise is out of the scope of this work [5–7], but the most fundamental LC oscillator is interesting as a reference, especially as this is the most widely used mm-wave oscillator. All this work has resulted in a mathematical derivation for the phase noise of an LC oscillator expressed as follows [4, 8–10]:

$$PN(\Delta f) = 10 \log_{10} \left(\frac{K_B T}{2N Q^2 P_{DC} \eta_{RF}} \left(1 + \frac{\gamma}{\alpha} \right) \left(\frac{f_0}{\Delta f} \right)^2 \right), \quad (3.3)$$

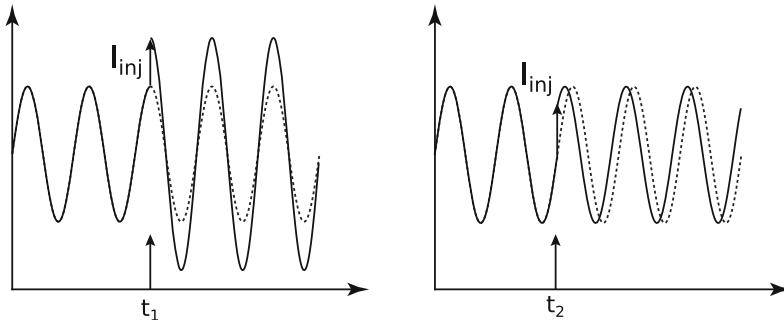


Fig. 3.2 The effect of an injected noise current at two different times t_1 and t_2 , resulting in pure magnitude and phase jump, respectively

where K_B and T are the Boltzmann constant and the temperature, respectively. Q is the quality factor of the resonator tank, and N signifies the number of coupled resonators. γ is the noise factor of the active device, and α is the noiseless gain between the tank and the active stage, for which P_{DC} is the DC-power consumption and η_{RF} is the DC-to-RF efficiency [8]. The combination of $P_{DC}\eta_{RF}$ expresses the oscillation amplitude across the resonator [4]. The oscillation frequency is given as f_0 , where the phase noise is measured at an offset frequency, Δf , from the carrier.

From this general result on the phase noise of an LC oscillator, trade-offs in the design can be discussed. (1) To achieve a low phase noise, a high-quality resonator is crucial. However, anything attached to the resonator tank will degrade the quality. Output buffers or frequency tuning devices will therefore increase the phase noise of the oscillator. Therefore, the larger the desired tuning range (TR) the worse the phase noise of an oscillator will be. (2) For a fixed resonator, a lower phase noise can only be achieved by an increased voltage swing that is mostly limited by the reliability of the active devices. At lower RF frequencies, the use of thick-oxide devices can help increase the maximal swing across the resonator and reduce the phase noise [11]. However, at mm-wave frequencies, these devices do not provide the necessary gain and cannot be used. Furthermore, as technology scales, the maximal reliable voltage swing on a transistor reduces leading to increased phase noise. However, as the maximum amplitude on the resonator is reached, an improved DC-to-RF efficiency will result in the same phase noise for a lower DC-power consumption. (3) By combining N resonators, an N improvement in phase noise can be achieved at the penalty of extra area and power consumption [12, 13]. (4) An improved passive voltage gain α will result in lower phase noise. (5) The higher the frequency the larger the phase noise will be even if all component performance stays the same. mm-wave frequency generating circuits will thus require extra effort with increasing frequency to reach the same phase noise goal.

3.1.2 Frequency Tuning

To use an oscillator in a practical design, tuning of the oscillation frequency is required. On the one hand, a frequency shift due to errors in the fabrication process needs to be recovered. On the other hand, tuning allows the oscillator to be used over a larger frequency range. From Eq. (3.1), the frequency can be adjusted by acting on either the capacitance or the inductance of the resonator.

3.1.2.1 Varactor Tuning

A varactor is a widely used variable capacitor, used in voltage-controlled oscillators to provide voltage control. A varactor is implemented as an accumulation-mode MOS device, where drain and source are connected, by controlling the gate voltage, and the oxide capacitance is modulated [1, 14]. The quality factor and variable capacitance of two NMOS varactors connected at the drain and source over frequency are shown in Fig. 3.3. The downside of using varactors for wide frequency tuning is apparent immediately. At mm-wave frequencies, varactors are plagued with very low quality factor, severely degrading the overall quality of the resonator and, thus, increasing the phase noise of the oscillator. Furthermore, a very limited ratio C_{max}/C_{min} results in a small tuning range. It is for these reasons that at mm-wave frequencies alternative frequency tuning techniques are used.

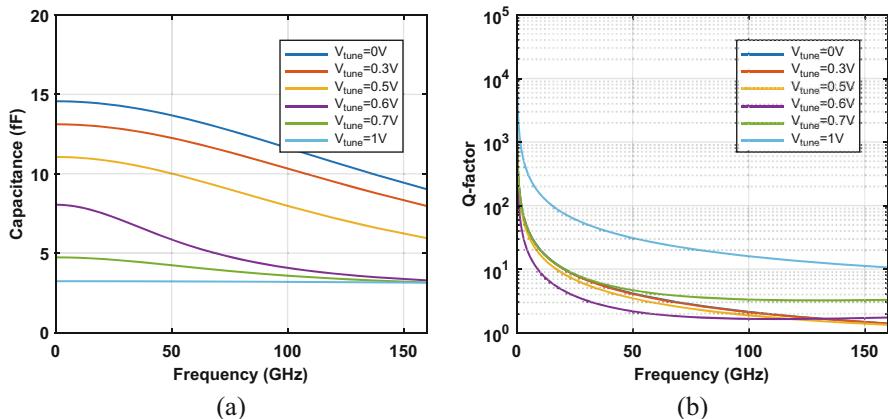


Fig. 3.3 The quality factor and capacitance of an NMOS varactor over its tuning range

3.1.2.2 Switched Capacitors

An alternative variable capacitor is a switched capacitor. A changing capacitance is created, by switching in and out a fixed capacitor as shown in Fig. 3.4. This results in discrete capacitance step [15]. Multiple switched capacitors need to be combined to create a usable tuning range. In Chap. 2, it is already established that the quality of capacitor and performance of the transistor degrade with frequency and scaling of the technology. The design of the switch shows a trade-off between tuning range and quality factor. For an optimal quality factor in the on-state, a large transistor is required, to provide a low on-resistance and a high-Q capacitor. The wide transistor comes with a large parasitic capacitance that increases the C_{min} and thus decreases the C_{max}/C_{min} ratio and the tuning range [10, 16]. Figure 3.5 shows the performance of two switched capacitors with different C_{fixed} and the same switch. A small capacitor with a large switch shows good Q but a bad tuning range compared to the larger capacitors. The performance of the switched capacitor is better than the varactor but still not great at mm-wave frequencies.

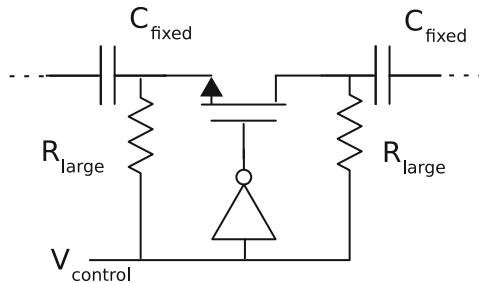


Fig. 3.4 Schematic of a switched capacitor

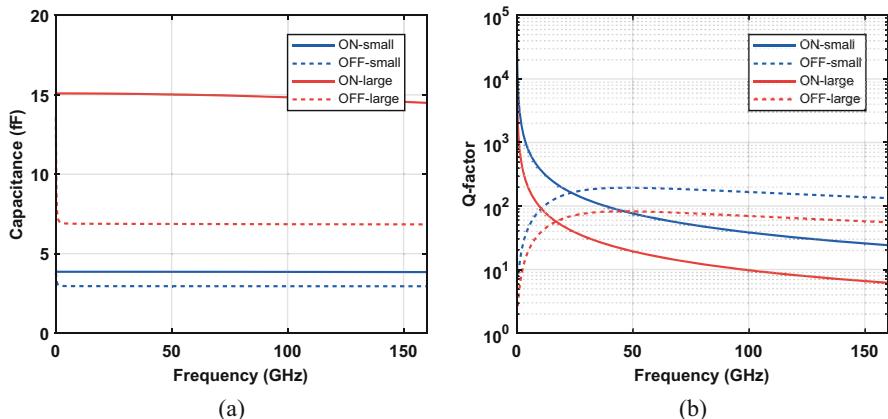


Fig. 3.5 The quality factor and capacitance of a switched capacitor in its on- and off-state for multiple sizes

3.1.2.3 Switched Inductors

In Chap. 2, inductors do not show the degradation of quality as the frequency increases. It therefore might be a promising candidate for frequency tuning at mm-wave frequencies [17]. Another more practical implementation of a switched inductor is the switched coupled inductor [18]. In this implementation, the switch is DC isolated from the core oscillator providing extra design freedom. Although the switch is still the bottleneck on the performance at mm-wave, a large switch can be used to limit the on-resistance. Although, the parasitic capacitance of the switch is disconnected from the frequency tuning. The off-capacitance will in this case reduce the quality of the resonator in the off-state by allowing some leakage current to flow in the inductor.

3.2 LO Architectures and Trends at mm-Wave

Using deeply scaled CMOS technologies at mm-wave frequencies for high-performing frequency generating circuits poses a lot of challenges. The increased frequency will inherently reduce the phase noise performance, on top of the reduced performance of both active and passive components. To tackle these issues and improve the overall system performance, several frequency generating circuit architectures have been proposed [19, 20]. These architectures can be categorized into three groups as shown in Fig. 3.6: (1) a phase-locked loop (PLL) with a voltage-

Fundamental oscillator	Oscillator	Divider / Multiplier
	<ul style="list-style-type: none"> ⊕ output frequency ⊖ Small tuning range ⊖ Bad switches ⊖ Low Q capacitors ⊖ Poor transistors perform. 	<ul style="list-style-type: none"> ⊕ No Multiplier needed ⊖ Bad divider ⊖ inductive tuned ⊖ injection locked ⊖ large area
	<ul style="list-style-type: none"> ⊕ fraction of output frequency ⊕ large tuning range ⊕ good switches ⊕ high Q capacitors ⊕ Good transistors perform. 	<ul style="list-style-type: none"> ⊖ Bad multiplier ⊖ inductive tuned ⊖ injection locked ⊖ low tuning range ⊖ large area ⊕ No extra divider needed
	<ul style="list-style-type: none"> ⊕ fraction of output frequency ⊕ large tuning range ⊕ good switches ⊕ high Q capacitors ⊕ Good transistors perform. 	<ul style="list-style-type: none"> ⊕ No Multiplier needed ⊕ No extra divider needed ⊖ Extra load on oscillator 2 buffers

Fig. 3.6 Overview of the 3 main LO generation architectures

controlled fundamental oscillator [6, 12, 18, 21–32], (2) a PLL with a frequency multiplier or injection-locked multiplier (ILM) [33–38], and (3) a PLL with a harmonic oscillator [13, 20, 39–46].

The first architecture combines a fundamental mm-wave oscillator and a mm-wave divider to downconvert the high-frequency signal for the PLL. This topology requires high-performance mm-wave transistors and a high-quality resonator to create a high-frequency LO. This is difficult to achieve above 100 GHz especially in a deeply scaled CMOS process. The performance of the transistors used as an active device or a switch for tuning degrades rapidly with increasing frequency. Furthermore, the quality factor of the integrated capacitors is reduced, which results in a poor resonator.

The second architecture combines a low-frequency LO with a frequency multiplying structure to generate the high-frequency signal. This topology benefits from the increased transistor performance and higher quality passives at lower frequencies and results in high-performance LO. However, this architecture poses stringent requirements on the mm-wave multiplier such as area, tuning range, and power consumption on top of its poor conversion gain.

The third architecture uses an oscillator that generates a higher harmonic besides the fundamental. This topology allows the oscillator to operate at a lower frequency and therefore benefits from the improved performance of transistors and passives. On the other hand, by extracting a higher harmonic, no mm-wave multiplier is necessary saving on area and power. However, this topology places a large load on the oscillator as both the divider and the output buffer are connected, reducing the tank's quality. Furthermore, depending on the harmonic oscillator's type, extra output filters and amplifiers are necessary.

Figure 3.7 reports the phase noise of the state-of-the-art oscillator designs over frequency at a 10 MHz offset. At frequencies up to 100 GHz, the phase noise of the oscillators follows the expected 20 dB/dec increase with frequency although most oscillators start to show worse PN at increased frequency. At frequencies above 100 GHz, fundamental oscillators quickly show deteriorating phase noise performance, and only harmonic or multiplier-based designs show published measurements. There do exist fundamental oscillators up to 300 GHz [21], but no phase noise is reported.

Although the phase noise performance is crucial, it does not say anything about power consumption or tuning range. For further comparison of these architectures, we consider two widely used figure of merits (FoM):

$$\begin{aligned} FoM &= |PN| + 20\log_{10}\left(\frac{f_0}{\Delta f}\right) - 10\log_{10}(P_{DC}(mW)) \\ FoMt &= FoM + 20\log_{10}\left(\frac{TR(\%)}{10}\right). \end{aligned}$$

The first FoM normalizes the PN against the power and oscillation frequency according to the formula of phase noise in Eq. (3.3). The second, FoMt, tries to

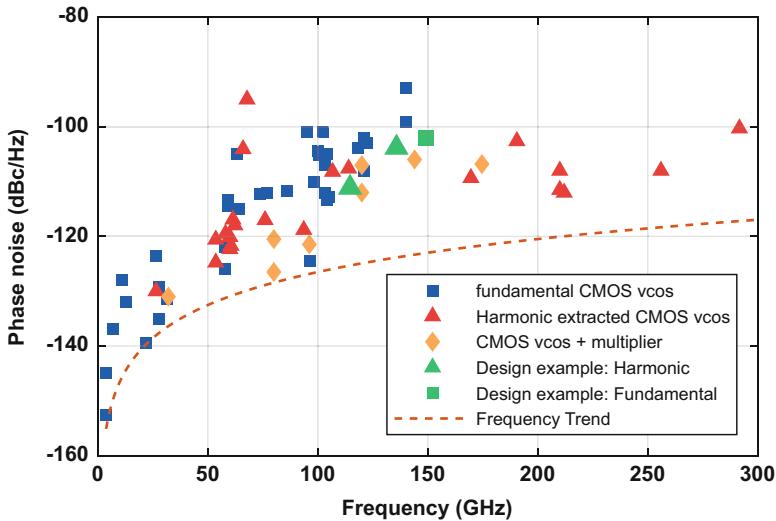


Fig. 3.7 The reported phase noise of the state-of-the-art CMOS oscillators normalized to 10 MHz offset against the oscillation frequency. Next to the expected phase noise trend extrapolated from the low phase noise designs

balance the PN penalty due to a reduced tank Q because of tuning elements. Figure 3.8 shows the current literature across the three architectures over the two figures of merits. The maximum FoM in the literature is achieved around 192 dBc/Hz up to 60 GHz. Above 60 GHz, the FoM overall frequency generators drop as indicated by the two lines. The fundamental oscillators have the steepest drop as transistors have no operating gain anymore. The difference between fundamental LOs and Harmonic or tripler-based LOs is seen more clearly in FoMt when the small tuning range of the fundamental oscillators above 100 GHz is taken into account.

3.3 Challenges of a Deeply Scaled Technology at mm-Wave

In Chap. 2, the effects of increased frequency and technology scaling on the performance of the passive and active components are already discussed. In this section, a deeper dive into the challenges of active design and its penalties on oscillator design is explored. The main purpose of the active components is the negation of the parasitic losses of the resonator R_T . The most popular topology to achieve this is a cross-coupled transistor pair. Figure 3.9 shows the circuit, next to its simplified small-signal model (without C_{GD} for simplicity). At very low frequencies, the admittance seen into this circuit is $-g_m/2$ or resistance of $R_T = -2/g_m$ is created. The transistors also come with parasitic capacitance

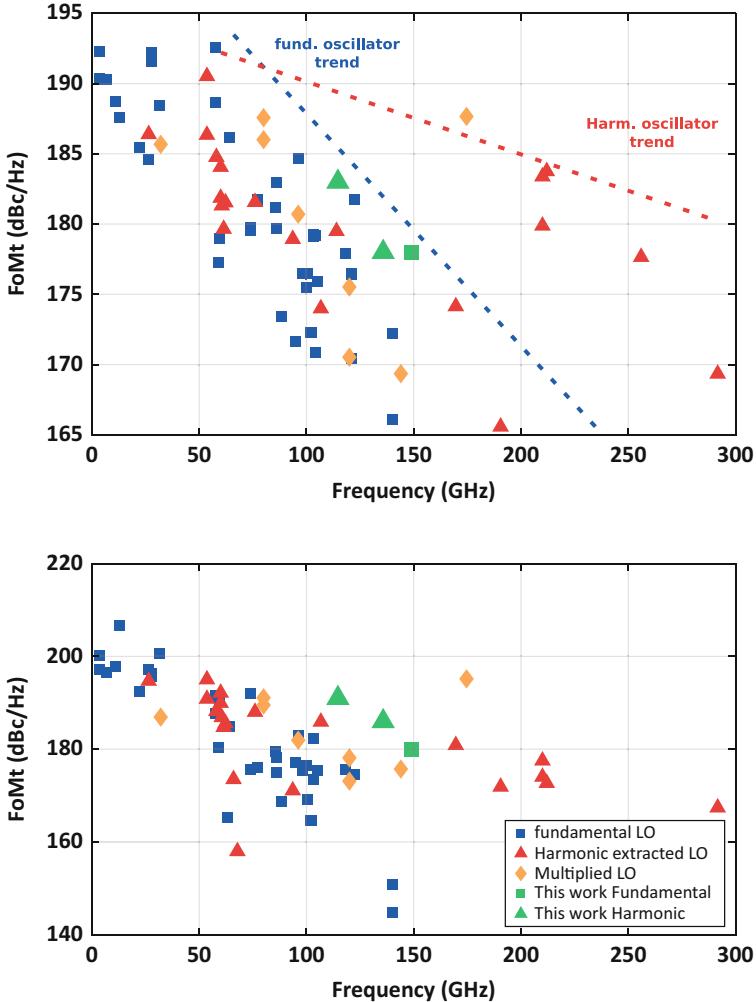


Fig. 3.8 The FoM and FoMt for different LO architectures from the literature and the 2 design examples in this chapter [6, 12, 13, 18, 20, 22–47]

resulting in a total capacitance seen as $C_{par} = (C_{GS} + C_{DS})/2$. This parasitic capacitance is added to the resonator as a fixed capacitance and leads to a reduced tuning range.

At mm-wave frequencies, the gate resistance plays an important role in the performance of the transistor as seen in the f_{max} . The total input admittance, Y_X , can be calculated from the small-signal model in Fig. 3.9 and is given as follows [21]:

$$Y_x = \frac{C_{GSS} - g_m}{2(R_G C_{GSS} + 1)} + \frac{1}{2R_{DS}} + \frac{C_{DSS}}{2} \quad (3.4)$$

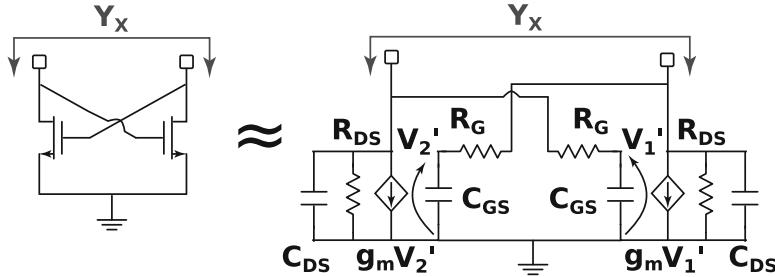


Fig. 3.9 A typical cross-coupled transistor pair acting as a negative resistor in an LC oscillator, next to its simplified small-signal representation

Decomposed as:

$$Re(Y_x) = \frac{-g_m}{2(1 + R_G^2 C_{GS}^2 \omega^2)} + \frac{R_G C_{GS}^2 \omega^2}{2(1 + R_G^2 C_{GS}^2 \omega^2)} + \frac{1}{2R_{DS}} \quad (3.5)$$

$$Im(Y_x) = \frac{(1 + g_m R_G) C_{GS} \omega}{2(1 + R_G^2 C_{GS}^2 \omega^2)} + \frac{C_{DS} \omega}{2}. \quad (3.6)$$

At low frequencies, both the R_G and the $C_{GS}\omega$ product are small and thus reduce to $-g_m/2$ and $C_{par} = (C_{GS} + C_{DS})/2$. As the frequency increases, the effective negative transconductance only reduces. To compensate for these losses and still achieve oscillation, the transistors need to produce more transconductance and thus their width needs to be increased. This increased size will result in increased power consumption and increased parasitic capacitance both reducing the oscillator's performance. The FoM will diminish as more power is consumed for the same performance, and the tuning range is reduced as more fixed capacitance is added to the resonator.

As the technology scales down, the parasitic gate resistance increases and results in a larger penalty at mm-wave frequencies. These trade-offs are visualized in Fig. 3.10, for a 40 nm and 16 nm FinFET technology. A cross-coupled pair is sized, by increasing the number of fingers, to achieve an effective transconductance of -3 mS at the operating frequency for a supply of 0.6 V and 0.8 V . The parasitic capacitance affecting the tuning range is shown in Fig. 3.10a, and the power consumption affecting efficiency and FoM is shown in Fig. 3.10b.

At low frequencies, both technologies show very similar performance; a lower supply results in a more efficient transconductance but results in a slightly larger parasitic capacitance. It will also result in a lower voltage swing improving reliability but degrading the phase noise [1]. As the frequency increases, both technologies need to increase the width of the transistor to keep the transconductance constant resulting in increased parasitic capacitance and increased power consumption. From 100 GHz and up, the 16 nm technology shows reduced performance compared to 40 nm. It shows increased gate resistance, and the gate resistance also increases

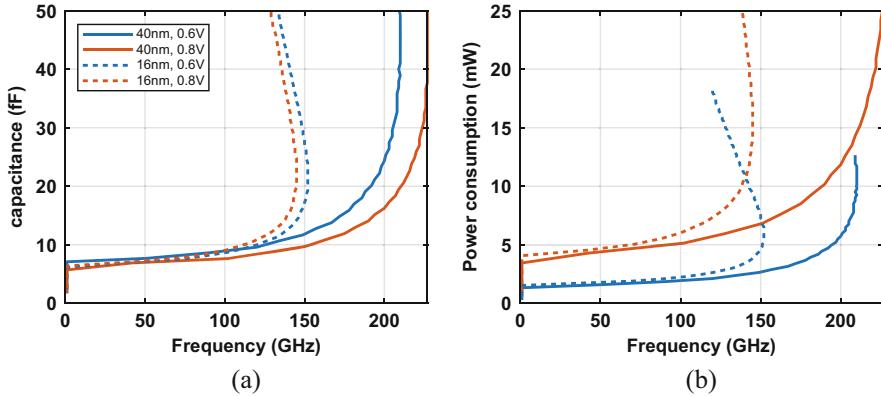


Fig. 3.10 Parasitic capacitance (a) and power consumption (b) of a cross-coupled transistor pair sized at an effective transconductance of -3 mS for a bulk 40 nm and 16 nm FinFET technology

more rapidly with increasing transistor size. For a 16 nm technology, a limit is reached at 150 GHz; at this point, the increase in transconductance is completely offset by the increased gate resistance, and no effective transconductance improvement can be gained by adding fingers. It must be noted that this crossing point also depends on the required transconductance and can be further improved upon as is discussed in the design example.

3.4 A 4th Order Transformer-Based Resonator

An LC-resonator only has one resonance limiting the design space of the oscillator. In recent years, higher-order resonators are used to expand the oscillators design space to improve performance. The two resonant frequencies of the transformer-based resonator have been used to extend TR, reduce phase noise, or enhance harmonic generation [6, 20, 25, 48–50].

The magnetically coupled RLC resonator is already discussed in Sect. 2.3.2.1 to perform interstage matching by looking at Z_{21} . In this section, the use of coupled resonator for oscillators is discussed. For an oscillator, the input impedance Z_{in} is of interest. Figure 3.11a shows the coupled resonator, and the main expression of the resonance frequencies is repeated below [6, 48, 49, 51]:

$$\omega_{L,H}^2 = \frac{1 + X \pm \sqrt{1 + X^2 + X(4k_m^2 - 2)}}{2L_2C_2(1 - k_m^2)} \quad (3.7)$$

$$X = \frac{L_2C_2}{L_1C_1}. \quad (3.8)$$

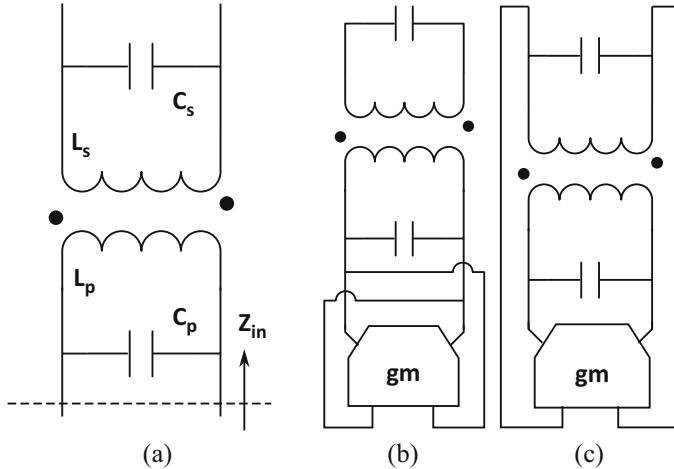


Fig. 3.11 Transformer coupled resonator (a) and its two oscillator implementations (b) and (c)

The frequency ratio between the two can be expressed as follows:

$$\frac{\omega_H}{\omega_L} = \sqrt{\frac{1 + X + \sqrt{1 + X^2 + X(4k_m^2 - 2)}}{1 + X - \sqrt{1 + X^2 + X(4k_m^2 - 2)}}}. \quad (3.9)$$

The frequency ratio between the two resonances is defined by the coupling factor k_m and the ratio X . Figure 3.12 evaluates the expression of frequency ratio ω_H/ω_L for different X and k_m values.

For the interstage matching in Sect. 2.3.2.1, an X value of 1 was always assumed leading to a single solution with two equal resonance peaks. For $X \neq 1$, the two resonance peaks are not equal in magnitude, and each X value for a fixed k_m also has a second solution at the same two resonances at its inverse X^{-1} . Figure 3.13 illustrates these two solutions in an example with $X=3$ and $k_m = 0.6$. Furthermore, the inductor and capacitor values are set as follows: $L_1 = 100 \text{ pH}$, $L_2 = 200 \text{ pH}$, $C_1 = 15 \text{ fF}$, and $C_2 = 22.5 \text{ fF}$. This results in $X=3$ and by switching L_1 with L_2 and C_1 with C_2 results in $X=0.33$. This results in a first resonance (f_1) at 70 GHz and a second resonance (f_2) at 175 GHz. Thus for an X of 3 or 0.33 and a $k_m=0.6$, a frequency ratio (f_2/f_1) of 2.5 is achieved. These two solutions effectively happen for a single resonator depending on which side of the resonator one is looking at. This is leveraged in [48] to achieve a dual-mode oscillator working on either of the two resonances. The imbalance can also be leveraged to equalize the frequency response of interstage matching [52].

A fourth order resonator can be used to create 2 types of oscillators, a one-port or a two-port oscillator as shown in Fig. 3.11b, c, respectively [48]. The one-port oscillator employs a cross-coupled pair to generate a negative transconductance to

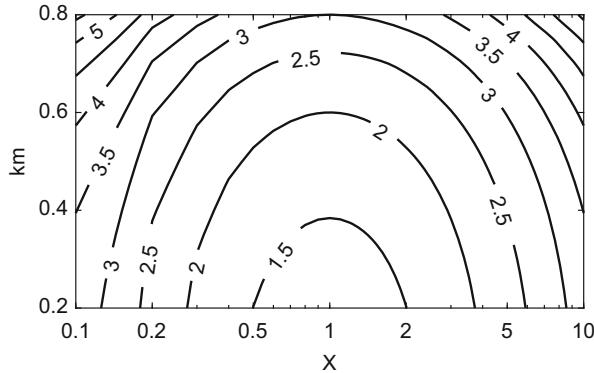


Fig. 3.12 The frequency ratio dependence on X and k_m

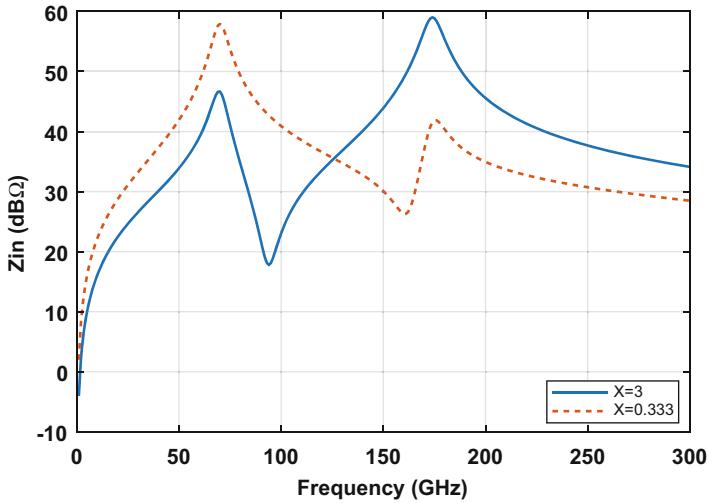


Fig. 3.13 Z_{in} for a resonator with $k_m=0.6$ and $X=1/3$ or $X=3$

negate the losses of the resonator and sustain oscillation. The start-up condition is therefore defined as follows at resonance.

$$-g_m + \operatorname{Re}(Z_{in}^{-1}) < 0 \quad (3.10)$$

$$\operatorname{Im}(Z_{in}) = 0. \quad (3.11)$$

The exact oscillation condition depends on the magnitude of both $-g_m$ and $\operatorname{Re}(Z_{in}^{-1})$ at ω_1 and ω_2 . Therefore, the oscillator can oscillate at either ω_1 or ω_2 if $\operatorname{Re}(Z_{in,\omega_1}) < g_m$ and $\operatorname{Re}(Z_{in,\omega_2} < g_m)$ or undesirably simultaneous oscillation [53].

The two-port oscillator employs feedback through the transformer, and the oscillation is set by the feedback network as the resonator Z_{21} transfer function

acts as a filter. Furthermore, by selecting a certain inductor ratio n , a passive voltage gain can be achieved improving the oscillator's performance [6].

3.5 Design Example: A Fundamental Oscillator

The previous section has shown that simply scaling a low-frequency oscillator will not result in a well-functioning fundamental mm-wave oscillator. In this section, the design of a D-band fundamental cross-coupled LC oscillator is tackled in a 16 nm FinFET technology. The design focuses on both improving the effective transconductance of the cross-coupled pair and improving the tuning range, without degrading the resonator's performance by using a 4th order transformer resonator.

3.5.1 The Cross-Coupled Pair

The first thing that can be done to improve the transconductance in a 16 nm FinFET technology is splitting up the transistor into smaller unit transistors, as discussed in Chap. 2. In this case, a cross-coupled transistor pair with transistors sized at 6 fins and 30 fingers generates a transconductance of -2 mS with a parasitic capacitance of 26 fF at 140 GHz . If the transistors are split up into two transistors, sized at 6 fins with 15 fingers and connected in the top metal layers, the performance improves and provides a transconductance of -4 mS with a parasitic capacitance of 26 fF .

The performance of the cross-coupled transistors can be further boosted by adding a source capacitor, C_c , and choke inductor, L_c , [18, 54] as shown in Fig. 3.14. The simplified small-signal model in Fig. 3.14 assumes L_c large enough to be considered an open, and further are the C_{GD} , R_{DS} , and C_{DS} omitted for simplicity. The input admittance, Y_x , can be expressed as follows:

$$Y_x = \frac{(C_{GS}s - g_m)C_{cs}}{2C_c C_g R_G s^2 + (2C_c + C_g)s + g_m} \quad (3.12)$$

with

$$Re(Y_x) = \frac{-g_m C_c \omega^2 (C_{GS} + C_c) + R_G C_c^2 C_{GS}^2 \omega^4}{2R_G^2 C_c^2 C_{GS}^2 \omega^4 + \left(2C_c^2 + 2(1 - g_m R_G)C_{GS}C_c + \frac{C_{GS}^2}{2}\right)\omega^2 + \frac{g_m^2}{2}} \quad (3.13)$$

$$Im(Y_x) = \frac{\left((g_m R_G + 1)C_c + \frac{C_{GS}}{2}\right)C_{GS}\omega^2 - \frac{g_m^2}{2}C_c\omega}{2R_G^2 C_c^2 C_{GS}^2 \omega^4 + \left(2C_c^2 + 2(1 - g_m R_G)C_{GS}C_c + \frac{C_{GS}^2}{2}\right)\omega^2 + \frac{g_m^2}{2}}. \quad (3.14)$$

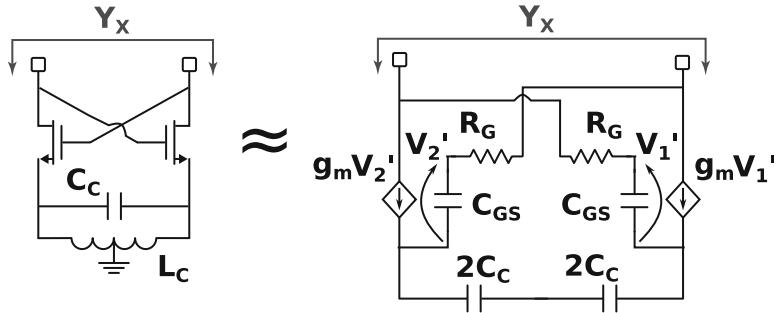


Fig. 3.14 A cross-coupled transistor pair with source capacitor, next to its simplified small-signal representation

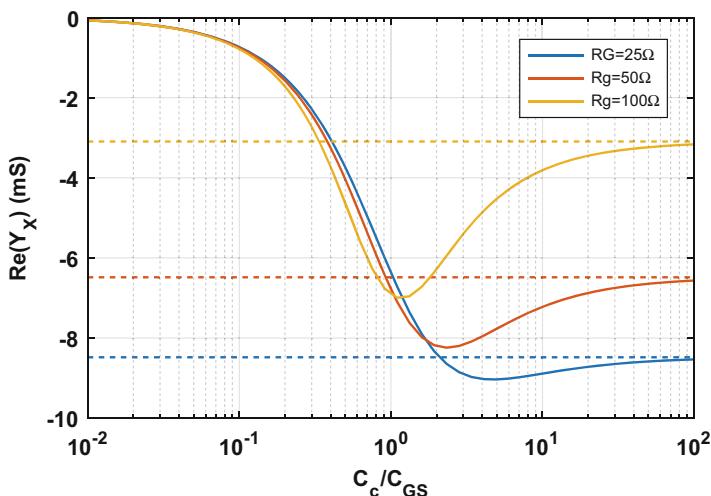


Fig. 3.15 The effect of the ratio C_c / C_{GS} on Eq. (3.13) for different gate resistances at 140 GHz with the dashed line indicating a grounded source

Figure 3.15 shows the effect C_c on the effective transconductance for an R_G of 25 Ω, 50 Ω, and 100 Ω at 140 GHz (with $C_{GS}=10 \text{ fF}$ and $g_m=20 \text{ mS}$). The more the cross-coupled pair is impacted by the gate resistance (or C_{GS}), the more benefit is gained from source degeneration. It can however not be completely offset with this technique. The technique becomes more useful as frequency increases and technology scales down and R_G increases. In the analytical model, the inductor is assumed large to act as choke; in the design, this is a 250 pΗ double winding inductor. The inductance is large enough to act like a choke at operating frequency but will change the performance at lower frequencies. In the design, a source capacitance of 15.5 fF is taken to place the optimal transconductance improvement at 140 GHz.

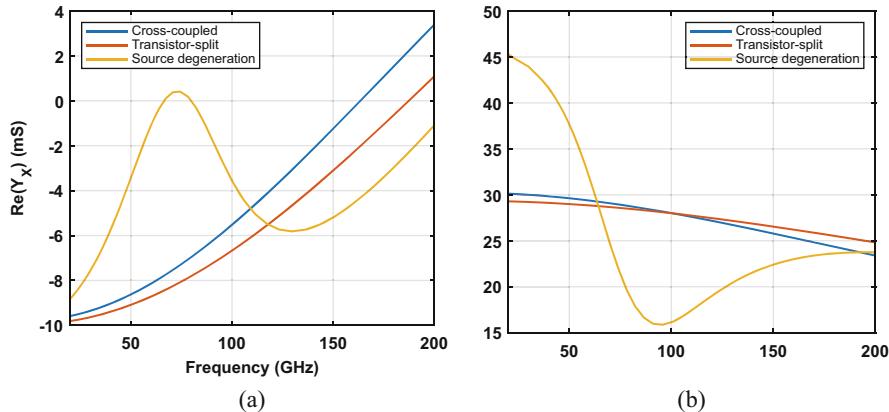


Fig. 3.16 Transconductance (a) and parasitic capacitance (b) of a base cross-coupled transistor pair. Showing the first improvement from splitting the transistors and further improvement by adding source degeneration

Figure 3.16a, b show the simulated effective transconductance and capacitance of the two techniques, compared to the original cross-coupled pair. The transconductance is almost 3 times better at 140 GHz using these two techniques. Splitting the transistors only works on the gate resistance and not on the gate capacitance, and no benefit can be achieved with this. However, the source degeneration also lowers the capacitance of the cross-coupled pair improving the tuning range of the oscillators or a better Q for the same tuning range.

3.5.2 The Resonator Design

A transformer acts both as the resonator for the oscillator and as the matching network to the output buffer. The transformer has winding inductances of $L_1 = 52 \text{ pH}$ and $L_2 = 103 \text{ pH}$ with a coupling factor of 0.32. A low coupling factor lowers the load of the buffer on the oscillator tank at a reduced power transfer. The schematic of the oscillator is shown in Fig. 3.17. The capacitance of the resonator is provided on one side by cross-coupled transistors as $C_1 = 26 \text{ fF}$ and on the other side by the input capacitance of the buffer, $C_2 = 5 \text{ fF}$. The buffer is a capacitive neutralized amplifier of one unit transistor, to minimize the load on the resonator. This puts the lower resonance at 131 GHz and the upper resonance at 246 GHz with $X=0.38$. Oscillation is ensured at the lower resonance due to the unavailable $-gm$ at the upper resonance. The resonance frequency of the resonator can be tuned by a variable capacitance and in this design implemented as a varactor, on either side of the resonator.

If the varactor is placed at the side of the cross-coupled transistors, the lowest frequency resonance is excited and tuned by the varactor as shown in Fig. 3.18 by

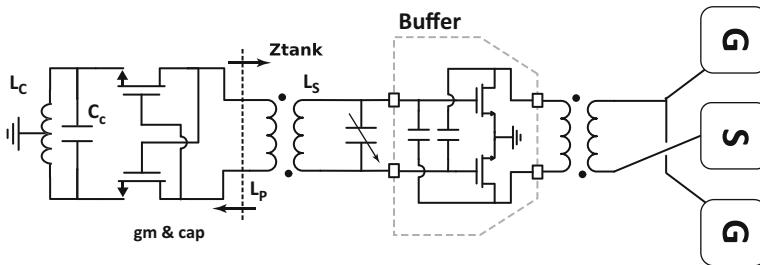


Fig. 3.17 Schematic of the fundamental oscillator

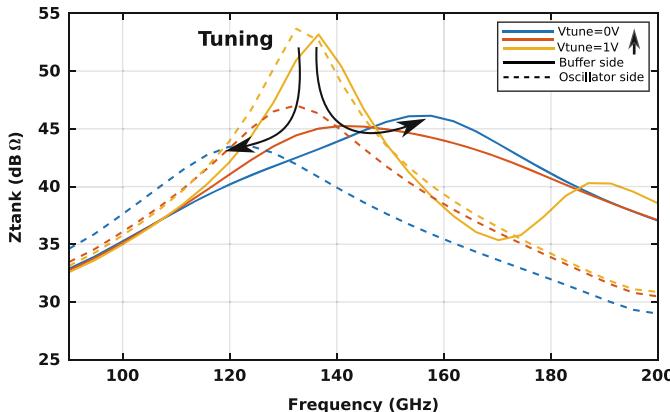


Fig. 3.18 The tank impedance over the tuning range for the varactor on the cross-coupled side (dashed line) and buffer side (solid line)

the dashed lines. As the capacitance is increased, the frequency lowers. Due to the bad performance of a varactor at mm-wave frequency, at increased capacitance, the quality factor of the resonator is reduced. This results in reduced performance and in this case for the highest capacitance even no start-up is possible.

If the varactor is placed on the buffer side, another behavior is possible. In this case for a low capacitance, the oscillator behaves the same as in the previous case. However, as the capacitance increases, the upper resonance decreases. At the same time, the magnitude of the upper resonance frequency increases as the X value increases. This results in a dynamic shift from the lower to the upper resonance over the tuning range as shown in Fig. 3.18. The worst performance of the oscillator is now not the lowest Q point of the varactor but the point where both resonances show equal impedance and provide a low-Q resonator. The lowest point of the resonance switching oscillator is better than the single resonance case and provides a wider tuning range. This results in a fundamental oscillator with improved tuning range.

3.5.3 Measurement Results

The design is fabricated in a 16 nm FinFET CMOS process and measures $165 \times 230 \mu\text{m}$. Figure 3.19 shows the die micrograph. The oscillator core consumes 5.18 mW from a 0.6 V supply, and the output buffer consumes 3.8 mW from a 0.8 V supply. The output of the die is probed and downconverted with a subharmonic mixer to the R&S FSW spectrum analyzer for frequency and phase noise measurements. The measured output frequency over the varactor's tuning range is shown in Fig. 3.20a. The oscillator operates from 139.2 GHz to 158.7 GHz, demonstrating a 13.1% tuning range. The phase noise at 140 GHz over offset frequency and the phase noise at 10 MHz offset over the tuning range are shown in Fig. 3.21a, b respectively.

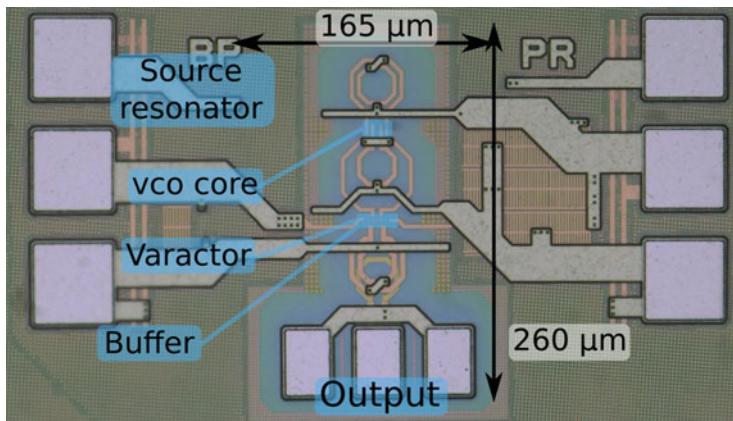


Fig. 3.19 Die micrographs of fundamental oscillator in 16 nm FinFET

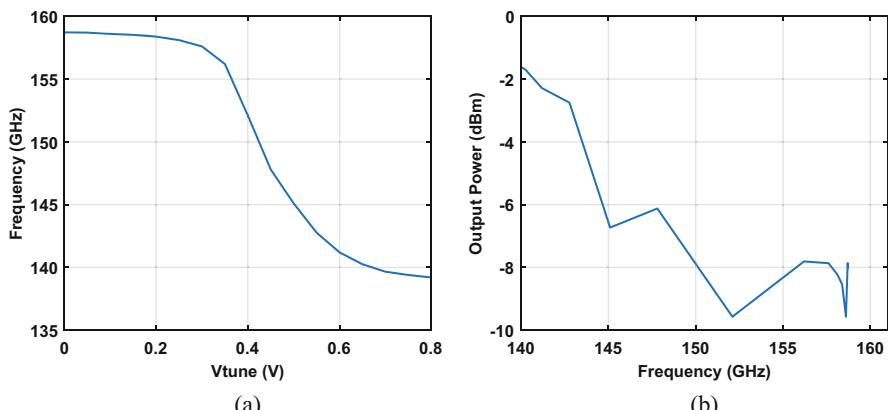


Fig. 3.20 Measured frequency for the tuning voltage (a) and the output power over the frequency range (b)

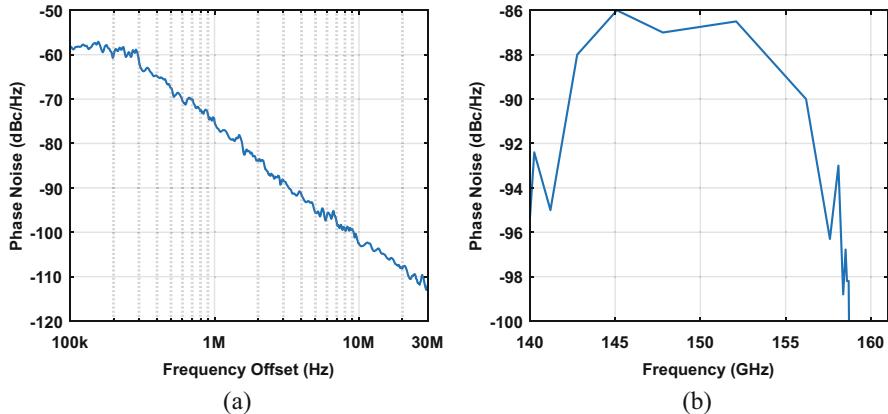


Fig. 3.21 Measured phase noise at 140 GHz over offset frequency (a) and over frequency range at 10 MHz offset (b)

The output power is measured with the PM5 power meter through probing. The measured output power is shown in Fig. 3.20b.

For a fundamental oscillator above 100 GHz, this design shows state-of-the-art tuning range and good phase noise although not constant over frequency.

3.6 Design Example: A Harmonic Oscillator

The conclusion from Sect. 3.2 was to lower the oscillation frequency to leverage the better performance of transistors and capacitors at lower frequencies. This requires extra circuits to increase the frequency back to the required frequency. In this design example, the frequency increase is achieved inside the oscillator itself, by promoting and extracting a higher harmonic. This has resulted in an oscillator operating from 36.2 to 48 GHz with an extracted third harmonic from 108.7 to 144 GHz.

3.6.1 Basic Principle

A harmonic oscillator oscillates at a fundamental frequency and outputs a higher harmonic to provide an n-factor frequency multiplication depending on the selected harmonic. The chosen harmonic number greatly impacts the design. A higher output harmonic allows the oscillator to operate at a lower frequency, improving its performance. However, a higher harmonic will have reduced output power and requires extra high-frequency buffers to increase the output power. The higher the frequency, the lower the gain of the buffers, therefore reducing the benefit of a higher

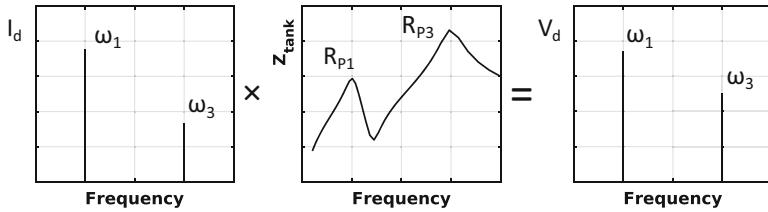


Fig. 3.22 Demonstration of the working principle of a harmonic oscillator

harmonic output. For differential oscillators, an even harmonic will help with the rejection of unwanted uneven harmonics reducing the need for filtering circuits. For the D-band frequency range, a harmonic number of three seems to be the sweet spot. One harmonic lower at the second harmonic would put the fundamental frequency between 55 and 85 GHz, and this is still high for the 16 nm FinFET process to create a good performing oscillator. One harmonic higher at the fourth harmonic would improve both the oscillator and the harmonic rejection but would require excessive buffering greatly increasing power and area. The basic concept of a third harmonic oscillator is shown in Fig. 3.22 and described in [6, 20, 38, 40]. A gm stage provides the necessary gain to start the oscillation. At oscillation, the transistor's non-linear behavior will generate 3rd harmonic currents at the output. These 3rd harmonic currents are turned into the desired 3rd harmonic voltages by the resonator. The resonator provides a resonance at the fundamental frequency (ω_1) and a second resonance at the third harmonic frequency (ω_3). The oscillator is implemented as a two-port oscillator with a transformer-based resonator. A coupling factor (k_m) around 0.6 is desired [20]; lowering k_m results in a reduced RP1 requiring extra gm to ensure oscillation and therefore results in larger power-hungry transistors and a reduced tuning range. Further increasing k_m is practically difficult and reduces RP3 resulting in a smaller 3rd harmonic voltage. If the mutual coupling is fixed, the ratio $X = \frac{L_s \cdot C_s}{L_p \cdot C_p}$ defines the frequency ratio between the resonances and can be set at 3. The ratio between the two windings of the transformer $\frac{L_s}{L_p}$ is chosen around 2 by employing a double winding. This double winding provides a passive voltage boost from the drain to the gate. A higher voltage at the gate will improve the phase noise [6] and the amount of harmonic currents. With the design choices of the transformer fixed, this leaves only the capacitor ratio $\frac{C_s}{C_p}$ to keep the resonance frequencies at a ratio of 3. The capacitance C_s will determine the lower resonance frequency and thus the frequency of the oscillator. The placement of tuning capacitors will therefore increase C_s for an increased tuning range. For the resonance ratio to stay fixed at 3 for each increase in C_s , an increase of C_p is required. However, the capacitance C_p impacts the resonance at the 3rd harmonic above 100 GHz. At these frequencies, the parasitics of the added capacitor will impact the impedance and thus the 3rd harmonic voltage.

3.6.2 Proposed Design

To maintain the correct resonance ratio, without adding C_p , two adaptations are made to the oscillator as shown in Fig. 3.23. The first adaptation removes the largest tuning capacitor to reduce C_s and substitutes the largest frequency jump with a switched leakage inductor to tune the resonator inductance [18]. The second adaptation reduces the size of the transistors M_G to reduce C_s . To maintain start-up and oscillation, two cross-coupled transistors M_C are added on the primary side. Furthermore, these transistors add capacitance to C_p , improving the balance between C_s and C_p . The trade-off between M_G and M_C is demonstrated by designing a 140 GHz harmonic oscillator with a fixed transformer and starting with M_G and M_C sized equally with 18 fingers. M_G is increased, while M_C is decreased to keep the total number of fingers constant at 36. Figure 3.23 shows the reduction in C_p and a slight increase of C_s , which provides extra tuning range. Figure 3.24 illustrates the effect on oscillators 3rd harmonic output voltage and the FoM. Exchanging M_G for M_C benefits the FoM of the oscillator. However, exchanging too much leads to a reduced 3rd harmonic voltage. Therefore in this design, a M_G with 24 fingers and a M_C with 12 fingers are chosen.

An overview of the designed oscillator core is shown in Fig. 3.25. The active core consists of two cross-coupled transistors of 12 fingers and 6 fins and two transistors, connected through the transformer, with 24 fingers and 6 fins. The transformer itself is a two-to-one transformer with a switchable leakage inductor on the outside, providing the largest frequency jump. To provide further fine frequency tuning a 4-bit, switchable binary scaled, the capacitor bank is provided on the secondary winding. Only one tunable capacitor is placed in the primary to minimize the load.

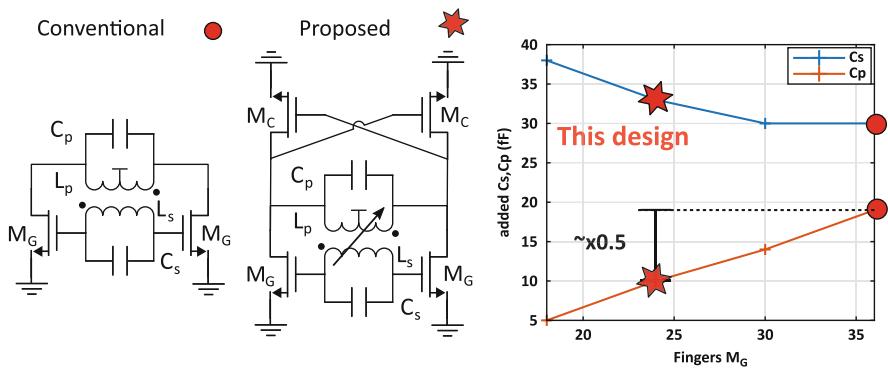


Fig. 3.23 Conventional class-F and proposed addition of M_C transistors and the effect on the required extra C_p

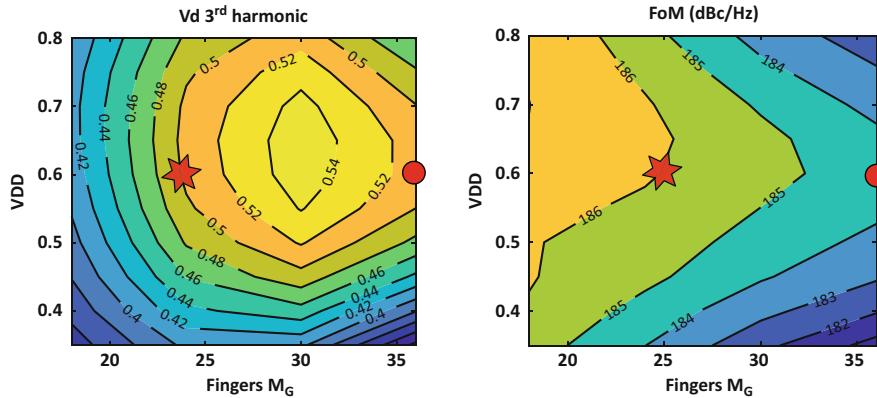


Fig. 3.24 The simulated difference in 3rd harmonic voltage and FoM for a different number of fingers in M_G , while keeping a total number of fingers constant $M_G + M_C = 36$. The conventional design is highlighted as circle and the proposed design with a star

3.6.3 Output Buffers

The oscillator now generates a large fundamental and 3rd harmonic tone. To use the generated signals, two buffers are needed to provide some power gain and filtering without loading the oscillator and degrading the performance; furthermore, the buffers need to have sufficient bandwidth to enable the whole tuning range.

3.6.3.1 Fundamental Buffer

The schematic of the fundamental buffer is shown in Fig. 3.26. The fundamental buffer consists of two neutralized common-source amplifier stages, the first stage has 2 unit transistors of 6 fins and 15 fingers, and the second stage is sized at two times 4 unit transistors. The input is capacitive coupled to the oscillator with a 142 fF capacitor. The 2 stages are matched together with a transformer of 420 to 405 pH and a mutual coupling of 0.26. The Q-factor of the output is reduced by placing a 2 k Ω resistor to improve the bandwidth. The output match is provided by an LC-transformer match. The simulated small-signal gain of the fundamental buffer is shown in Fig. 3.29a.

3.6.3.2 Third Harmonic Buffer

The third harmonic buffer is a 4-stage neutralized common-source amplifier as shown in Fig. 3.27. The first stage is moved inside the oscillator to be placed as close to the primary winding as possible. The effective impedance transformation

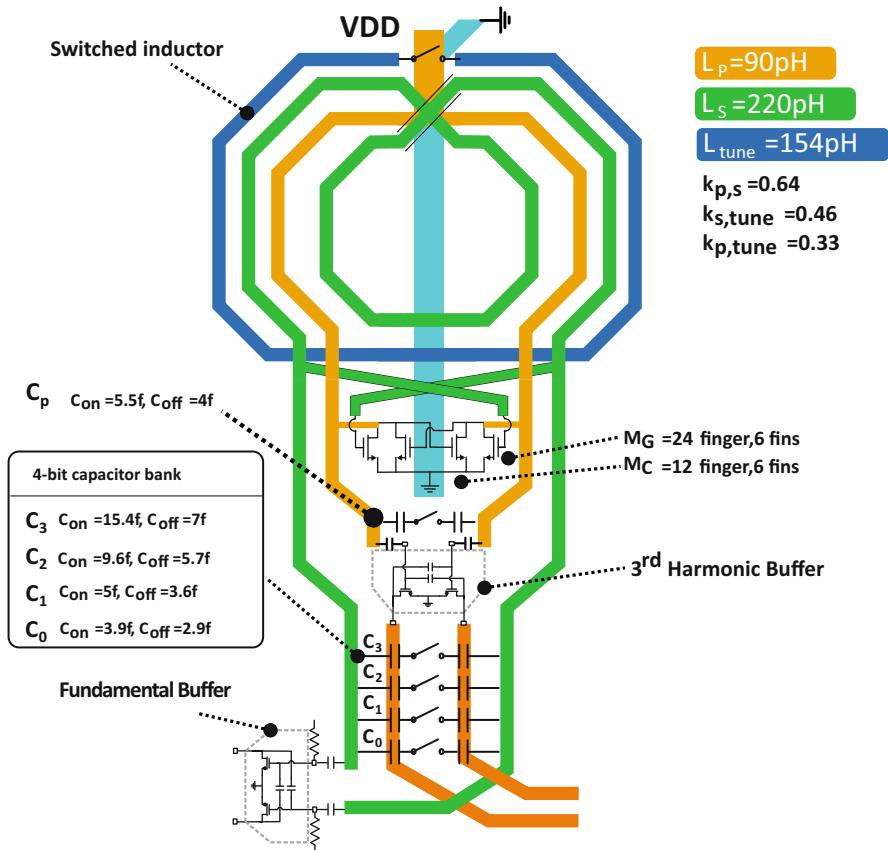


Fig. 3.25 Schematic and layout of the oscillator core

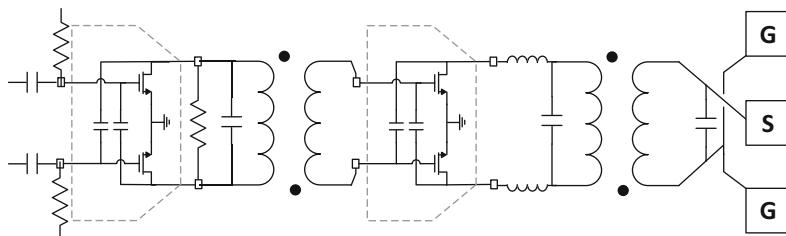


Fig. 3.26 Schematic of the fundamental buffer

caused by a transmission line connection between buffer and resonator is shown in Fig. 3.28. This shows that by moving the buffer as close as possible and placing the TL after the first buffer, the parasitic capacitance decreases, and the load on the resonator decreases. The stages are gradually sized up, from a two unit transistor of 6 fins and 15 fingers to 8 unit transistors, to be able to have a high input impedance

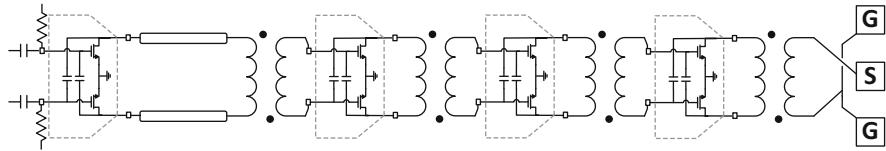


Fig. 3.27 Schematic of the harmonic buffer

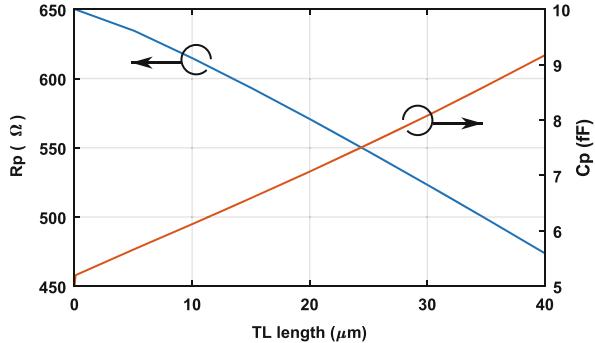


Fig. 3.28 The simulated load (parallel resistance and capacitance) of the buffer seen by the resonator depending on the distance between them

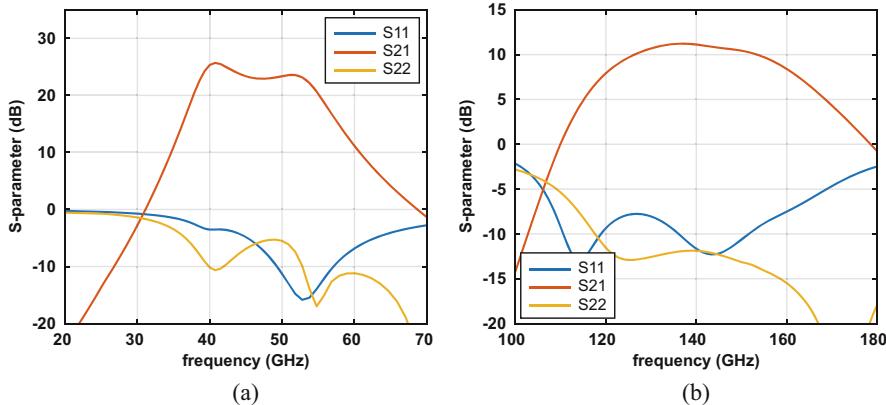


Fig. 3.29 The simulated s-parameters for the fundamental (a) and harmonic buffer (b)

for the oscillator and low enough output impedance to match to the $50\ \Omega$ output. To achieve the desired bandwidth, both staggering and moderately coupled transformer matching are employed [55]. The simulated small-signal gain of the third harmonic buffer is shown in Fig. 3.29b.

3.6.4 Measurement Results

The design is fabricated in a 16 nm FinFET process and measures $230 \times 800 \mu\text{m}$, with an oscillator core area of 0.013 mm^2 . The full die micrograph with a zoom of the core oscillator is shown in Fig. 3.30. The oscillator core consumes 7.2 mW from a 0.6 V supply and 0.6 V bias voltage. The fundamental and third harmonic output buffer stage each consume 3.5 mW from a 0.8 V supply. The fundamental and 3rd harmonic output amplifier consume 13.5 mW and 30.2 mW, respectively, from a 0.8 V supply. Figure 3.31 shows a picture and schematic overview of the measurement setup for characterization of the samples. The outputs of the oscillators are connected to an R&S FSW with PN-option or FSU for frequency and phase noise measurements. For the D-band outputs, a subharmonic mixer is used to downconvert the signal before connecting to the FSW. The output power below 110 GHz is measured with the R&S NRP2 power meter and above with the

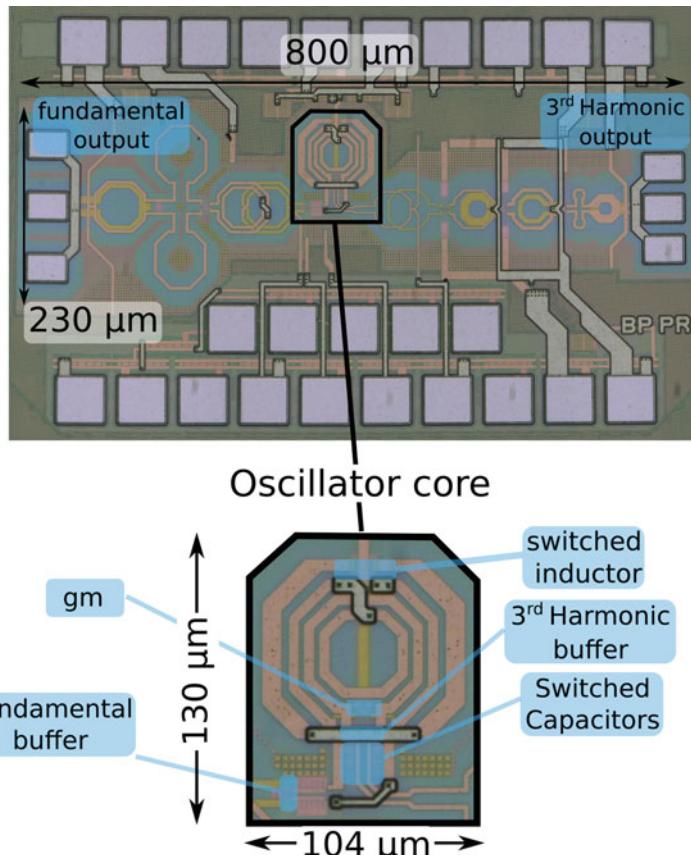


Fig. 3.30 Die micrograph of the harmonic oscillator with a zoom on the oscillator core

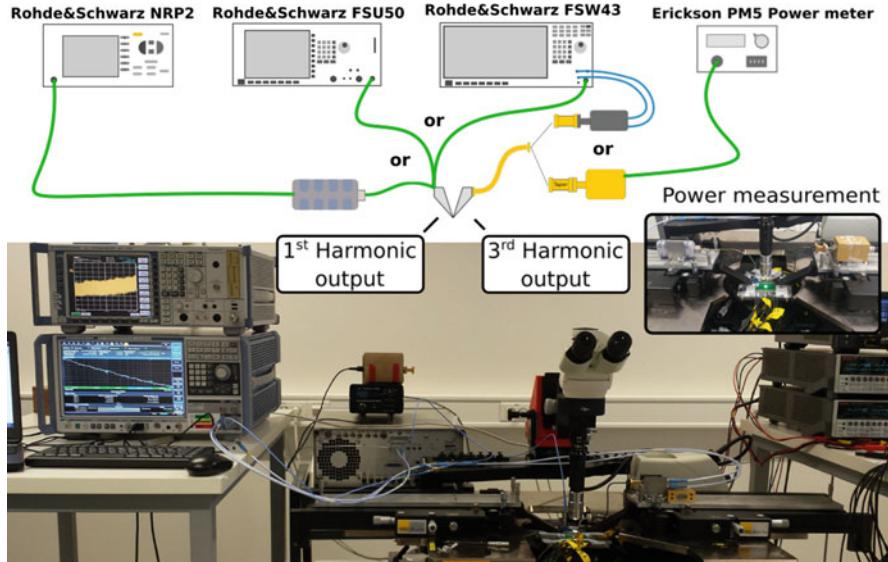


Fig. 3.31 Oscillator measurement setup

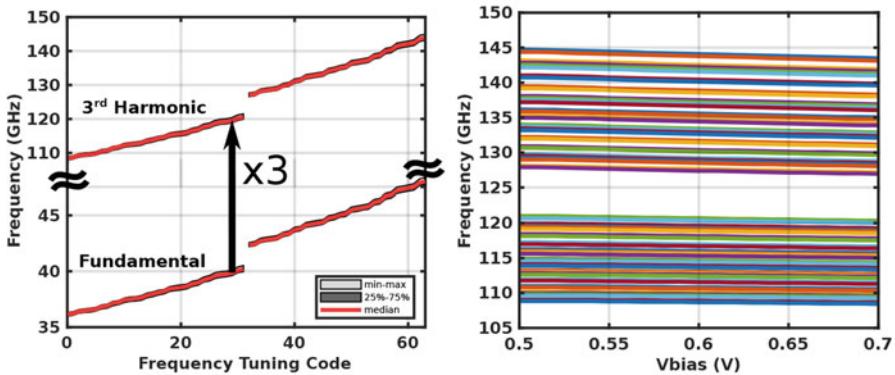


Fig. 3.32 Measured fundamental and 3rd harmonic frequency for the tuning codes

PM5 power meter. The measured frequency of the fundamental and third harmonic output is shown in Fig. 3.32 for the discrete tuning codes over 7 samples. The samples show little frequency difference; however, a discontinuity in the tuning range is measured compared to simulation. The frequency ranges from 108.7 to 144 GHz with a discontinuity from 120.6 to 127.5 GHz. This gap is due to a difference in modeling errors between the switched inductor and the switched capacitors. Continuous frequency tuning is achieved by adjusting the bias voltage of the oscillator core as shown in Fig. 3.32. The oscillator achieves a tuning range

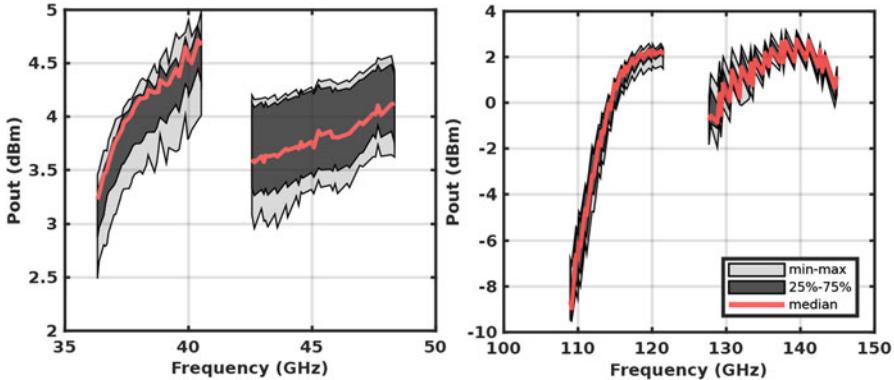


Fig. 3.33 Measured output power at the fundamental and 3rd harmonic output over the tuning range for 7 samples

of $11.9\text{ GHz}+16.5\text{ GHz}=28.4\text{ GHz}$ or 22.5%. The phase noise of the oscillator is measured and shown in Fig. 3.33. The phase noise is measured at both the fundamental tone of 36.3 GHz and the third harmonic of 109 GHz over the offset frequency. The measured phase noises of the fundamental tone are -97 dBc/Hz and -120.6 dBc/Hz at 1 MHz and 10 MHz offset, respectively. The measured phase noises of the third harmonic are -88 dBc/Hz and -111.1 dBc/Hz at 1 MHz and 10 MHz offset, respectively. The measurement shows the expected 9.5 dB difference between the fundamental and third harmonic with a noise corner of 2 MHz. The phase noise of the third harmonic at 10 MHz offset is shown over the tuning range in Fig. 3.33. A jump in phase noise performance is measured for the switched inductor as this degrades the quality of the resonator. The measured noise corner and FoM of the oscillator over the tuning range are shown in Fig. 3.34. The output power of the fundamental and third harmonic buffers is shown in Fig. 3.35.

3.7 Conclusion

Table 3.1 compares the two design examples to the current state-of-the-art oscillators. It shows the reduced performance in both tuning range and phase noise of fundamental oscillators compared to the harmonic or multiplier-based designs. The harmonic oscillator's performance holds up well against previous published designs, especially considering the higher operating frequency. Only [38] seems to outperform this work, but it does not employ output buffers and has therefore a low output power and reduced loading of the oscillator tank.

A fourth order transformer-based resonator is used in the design of a fundamental and harmonic oscillator at D-band frequencies in a 16 nm Finfet CMOS technology. The fundamental oscillator employs a source resonator to boost the performance at

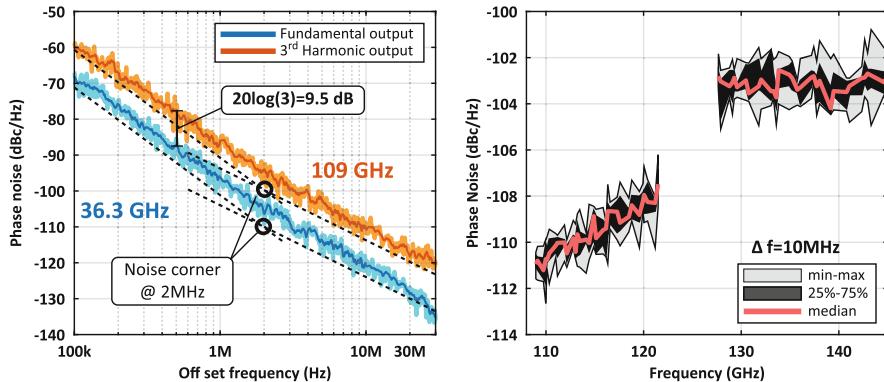


Fig. 3.34 Measured phase noise at 36.3 GHz and 109 GHz versus offset frequency and the measured phase noise of the 3rd harmonic output at 10 MHz offset over the frequency range for 7 samples

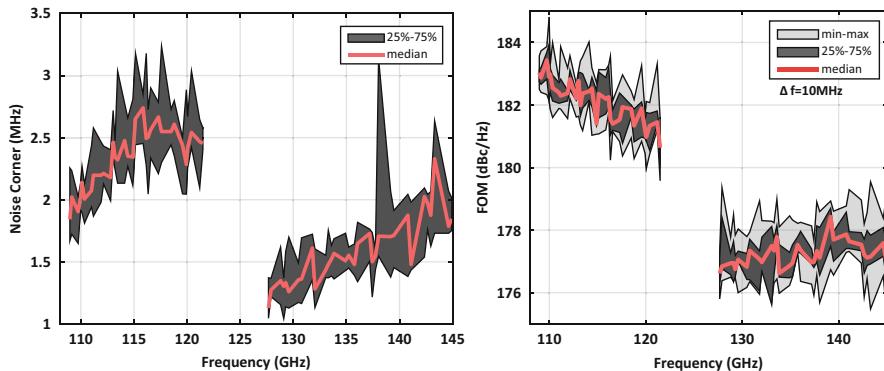


Fig. 3.35 Measured noise corner and FoM at the 3rd harmonic output over the tuning range for 7 samples

mm-wave frequencies and a varactor for resonance shifting frequency tuning. The fundamental oscillator has 13.1% tuning range with a peak FoMt of 178 dBc/Hz. The harmonic oscillator is adjusted for performance at D-band and outputs both the fundamental and the 3rd harmonic frequency. The harmonic oscillator achieves a 22.5% tuning range with a peak FoMt of 190.8 dBc/Hz. The harmonic oscillator outperforms the fundamental oscillator but has extra design challenges.

Table 3.1 Performance summary and comparison of previously published works

Reference	Technology	Oscillator type	Freq. (GHz)	Tuning range (%)	Power (mW)	Core area (mm^2)	Phase noise dBc/Hz @ Δf	FoM dBc/Hz @ Δf	FoMt dBc/Hz @ Δf	Phase noise dBc/Hz @ Δf	1 MHz	10 MHz	10 MHz
This work	16 nm FinFET	Fundamental	139.2-158.7	13.1	5.2	2	-75	-102	171	178	173.3	180.3	
RFIC'11 [22]	65 nm CMOS	Fundamental	122.5	4.4	2	0.22	-83	NA	181.8	NA	174.6	NA	
JSSC'17 [46]	130 nm SiGe	Fundamental	190.5	20.7	183	0.25	2	-102.6	2	165.6	2	171.9	
This work	16 nm FinFET	harmonic	108.7-120.6	22.5	7.1	0.013	-88	-111.1	180.2	183.3	187.8	190.8	
JSSC'17 [12]	55 nm BiCMOS	Multicore + quadrupler	127.5-144	80	15	50	0.598	-103.8	175.6	178	183.2	185.6	
JSSC'16 [20]	40 nm CMOS	Class-F 3rd HM extraction	60	25.4	13.5+10.5	0.067	-100.1	-122.3	181.9 ₊	184 ₊	190 ₊	192 ₊	
JSSC'16 [42]	130 nm CMOS	Self-mixing-VCO (class-C + Mixer)	52.8-62.5	16.8	4.1+3.5	0.06 _#	-100.6	-124.8	186.3	190.5	190.9	195	
RFIC'18 [38]	65 nm CMOS	Class-F 3rd HM + II Multiplier(3x)	153.7-195.3	23.7	2.5	0.08	-84.6 _*	-106.8 _*	NA	186.5	NA	194.1	
ESSCIR'C'19 [40]	16 nm FinFET	3rd HM extra.	68-84	21	10+10.2	0.13	-97	-117.2	181.6	181.8	188.1	188.3	
RFIC'13 [47]	45 nm CMOS	Mixing%	120	13.5	64	2	-87	-112	170.5	175.5	173.1	178	
TMFTT'18 [45]	90 nm CMOS	Triple push	210	5.1	28.6	2	-88	-111.5	179.9	183.4	174	177.5	
ISSCC'15 [41]	130 nm CMOS	Push-push	114	NA	8.4	2	2	-107.6	2	179.5	NA	NA	

γ Core area= Oscillator core and multipliers (no output buffers)

#estimated from die, +Includes first stage buffer, *Estimated from ILFM input, %Quadrature output, NA: not available

References

1. B. Razavi, *RF Microelectronics*. (Prentice Hall Press, Englewood, 2011)
2. D.B. Leeson, A simple model of feedback oscillator noise spectrum. Proc. IEEE **54**(2), 329–330 (1966)
3. A. Hajimiri, T.H. Lee, A general theory of phase noise in electrical oscillators. IEEE J. Solid-State Circuits **33**(2), 179–194 (1998)
4. A. Mazzanti, P. Andreani, Class-c harmonic CMOS VCOS, with a general result on phase noise. IEEE J. Solid-State Circuits **43**(12), 2716–2729 (2008)
5. S. Levantino, P. Maffezzoni, F. Pepe, A. Bonfanti, C. Samori, A.L. Lacaita, Efficient calculation of the impulse sensitivity function in oscillators. IEEE Trans. Circuits Syst. II Express Briefs **59**(10), 628–632 (2012)
6. M. Babaie, R.B. Staszewski, A class-F CMOS oscillator. IEEE J. Solid-State Circuits **48**(12), 3120–3133 (2013)
7. I. Chlis, D. Pepe, D. Zito, Phase noise comparative analysis of LC oscillators in 28-nm CMOS through the impulse sensitivity function, in *Proceedings of the 2013 9th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME)* (2013), pp. 85–88
8. L. Fanori, P. Andreani, Highly efficient class-C CMOS VCOs, including a comparison with class-b VCOs, IEEE J. Solid-State Circuits **48**(7), 1730–1740 (2013)
9. M. Garampazzi, S. Dal Toso, A. Liscidini, D. Manstretta, P. Mendez, L. Romană, R. Castello, An intuitive analysis of phase noise fundamental limits suitable for benchmarking LC oscillators. IEEE J. Solid-State Circuits **49**(3), 635–645 (2014)
10. R.P. Vigilante M., *5G and E-Band Communication Circuits in Deep-Scaled CMOS* (Springer, Berlin, 2018)
11. M. Babaie, R.B. Staszewski, A study of RF oscillator reliability in nanoscale CMOS, in *2013 European Conference on Circuit Theory and Design (ECCTD)* (2013), pp. 1–4
12. L. Iotti, A. Mazzanti, F. Svelto, Insights into Phase-Noise Scaling in Switch-Coupled Multi-Core LC VCOs for E-Band Adaptive Modulation Links. IEEE J. Solid-State Circuits **52**(7), 1703–1718 (2017)
13. J. Rimmelspacher, R. Weigel, A. Hagelauer, V. Issakov, A Quad-Core 60 GHz Push-Push 45 nm SOI CMOS VCO with -101 . 7 dBc Hz Phase Noise at 1 MHz offset, 19 Continuous FTR and -187 dBc Hz FoM T, in *Proceedings of the IEEE European Solid State Circuits Conference (ESSCIRC)* (2018), pp. 138–141
14. P. Andreani, S. Mattisson, On the use of MOS varactors in RF VCOs. IEEE J. Solid-State Circuits **35**(6), 905–910 (2000)
15. H. Sjoland, Improved switched tuning of differential CMOS VCOS. IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing **49**(5), 352–355 (2002)
16. S.D. Toso, A. Bevilacqua, A. Gerosa, A. Neviani, A thorough analysis of the tank quality factor in LC oscillators with switched capacitor banks, in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems* (2010), pp. 1903–1906
17. E. Mammei, E. Monaco, A. Mazzanti, F. Svelto, A 33.6-to-46.2GHz 32 nm CMOS VCO with 177.5dBc/Hz minimum noise FOM using inductor splitting for tuning extension, in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers* (2013), pp. 350–351
18. M. Vigilante, P. Reynaert, Analysis and Design of an E-Band Transformer-Coupled Low-Noise Quadrature VCO in 28-nm CMOS. IEEE Trans. Microwave Theory Tech. **64**(4), 1122–1132 (2016)
19. S. Kang, J.C. Chien, A.M. Niknejad, A W-Band Low-Noise PLL with a Fundamental VCO in SiGe for Millimeter-Wave Applications. IEEE Trans. Microwave Theory Tech. **62**(10), 2390–2404 (2014)
20. Z. Zong, M. Babaie, R.B. Staszewski, A 60 GHz Frequency Generator Based on a 20 GHz Oscillator and an Implicit Multiplier. IEEE J. Solid-State Circuits **51**(5), 1261–1273 (2016)
21. B. Razavi, A 300-GHz Fundamental Oscillator in 65-nm CMOS Technology. IEEE J. Solid-State Circuits **46**(4), 894–903 (2011)

22. W. Volkaerts, M. Steyaert, P. Reynaert, 118GHz fundamental VCO with 7.8 % tuning range in 65 nm CMOS, in *Digest of Papers—IEEE Radio Frequency Integrated Circuits Symposium* (2011), pp. 3–6
23. Y. Shu, H.J. Qian, A 169 . 6-GHz Low Phase Noise and Wideband Hybrid Mode-Switching Push $\ddot{\text{A}}$ Ş Push Oscillator. *IEEE Trans. Microwave Theory Tech.* **67**(7), 2769–2781 (2019)
24. R. Levinger, D. Ben-Haim, I. Gertman, S. Bershansky, R. Levi, J. Kadry, G. Horovitz, X-band NMOS and CMOS Cross-Coupled DCO's with a "Folded" Common-Mode Resonator Exhibiting 188.5 dBc/Hz FoM with 29.5% Tuning Range in 16-nm CMOS FinFET (2019), pp. 163–166
25. J. Yin, H.C. Luong, A 57.5-90.1-GHz Magnetically Tuned Multimode CMOS VCO. *IEEE J. Solid-State Circuits* **48**(8), 1851–1861 (2013).
26. H. Jia, B. Chi, L. Kuang, Z. Wang, A 47.6–71.0-GHz 65-nm CMOS VCO Based on Magnetically Coupled π -Type LC Network. *IEEE Trans. Microwave Theory Tech.* **63**(5), 1645–1657 (2015)
27. P. Chiang, O. Momeni, P. Heydari, A 200-GHz Inductively Tuned VCO With -7-dBm Output Power in 130-nm SiGe BiCMOS. *IEEE Trans. Microwave Theory Tech.* **61**, 3666–3673 (2013)
28. A. Mazzanti, P. Andreani, A push-pull class-C CMOS VCO. *IEEE J. Solid-State Circuits* **48**(3), 724–732 (2013)
29. C.H. Li, Y.L. Liu, C.N. Kuo, A 0.6-v 0.33-mw 5.5-GHz receiver front-end using resonator coupling technique. *IEEE Trans. Microwave Theory Tech.* **59**, 1629–1638 (2011)
30. M. Adnan, E. Afshari, *A 105GHz VCO with 9.5% Tuning Range and 2.8mW Peak Output Power Using Coupled Colpitts Oscillators in 65 nm Bulk CMOS* (2013), pp. 239–242
31. H. Guo, Y. Chen, P. Mak, R.P. Martins, 26.2 A 0.08 mm² 25.5-to-29.9 GHz Multi-Resonant-RLCM-Tank VCO Using a Single-Turn Multi-Tap Inductor and CM-Only Capacitors Achieving 191.6 dBc/Hz FoM and 130 kHz 1/f₃ Corner (2019), pp. 410–412
32. A. Bhat, N. Krishnapura, 26.3 A 25-to-38 GHz, 195dB FoMT LC QVCO in 65 nm LP CMOS Using a 4-Port Dual-Mode Resonator for 5G Radios (2019), pp. 412–414
33. Z. Chen, C. Wang, P. Heydari, W-band frequency synthesis using a Ka-band PLL and two different frequency triplers (2011), pp. 1–4
34. C. Wang, Z. Chen, P. Heydari, W-Band Silicon-Based Frequency Synthesizers Using Injection-Locked and Harmonic Triplers. *IEEE Trans. Microwave Theory Tech.* **60**, 1307–1320 (2012)
35. A. Tang, D. Murphy, G. Virbila, F. Hsiao, S.-w. Tam, H.-t. Yu, H.-h. Hsieh, C.P. Jou, Y. Kim, A. Wong, A. Wong, Y.-c. Wu, M.-c.F. Chang, D-Band Frequency Synthesis Using a U-band PLL and Frequency Tripler in 65 nm CMOS Technology, in *2012 IEEE/MTT-S International Microwave Symposium Digest* (2012), pp. 1–3
36. K. Tan, T. Chu, S.S.H. Hsu, A 76.2–89.1 GHz Phase-Locked Loop With 15.6% Tuning Range in 90 nm CMOS for W-Band Applications. *IEEE Microwave Wireless Compon. Lett.* **25**, 538–540 (2015)
37. H. Holisaz, S. Safavi-Naeini, A Low Noise D-Band VCO with a Wide Bandwidth and a Steady Output Power. *IEEE Microwave Wireless Compon. Lett.* **25**, 742–744 (2015)
38. X. Liu, A 0.3V 2.5mW 154 GHz–195 GHz CMOS IL LO Generation with -186.5dB FoM, in *Proceedings of the IEEE Radio Frequency Integral Circuits Symposium (RFIC)*, no. c (2018), pp. 196–199
39. Z. Zong, P. Chen, R.B. Staszewski, A Low-Noise fractional- N digital frequency synthesizer with implicit frequency tripling for mm-Wave applications. *IEEE J. Solid-State Circuits* **54**(3), 755–767 (2019)
40. F.-w. Kuo, Z. Zong, H.-n. R. Chen, L.-c. Cho, C.-p. Jou, M. Chen, R.B. Staszewski, A 77/79-GHz frequency generator in 16-nm CMOS for FMCW radar applications based on a 26-GHz oscillator with co-generated third harmonic, in *Proceedings of the IEEE European Solid State Circuits Conference (ESSCIRC)*, vol. 2 (2019), pp. 53–56
41. Y. Hu, T. Siriburanon, R.B. Staszewski, A 30-GHz Class-F23Oscillator in 28 nm CMOS using harmonic extraction and achieving 120 kHz 1/f₃Corner, in *Proceedings of the IEEE European Solid State Circuits Conference (ESSCIRC)* (2017), pp. 87–90

42. A.H. Masnadi Shirazi, A. Nikpaik, R. Molavi, S. Lightbody, H. Djahanshahi, M. Taghivand, S. Mirabbasi, S. Shekhar, On the Design of mm-Wave Self-Mixing-VCO Architecture for High Tuning-Range and Low Phase Noise. *IEEE J. Solid-State Circuits* **51** (2016), 1210–1222
43. P. Chiang, Z. Wang, O. Momeni, P. Heydari, A Silicon-Based 0.3 THz frequency synthesizer with wide locking range. *IEEE J. Solid-State Circuits* **49**, 2951–2963 (2014)
44. O. Momeni, E. Afshari, High power terahertz and millimeter-wave oscillator design: A systematic approach. *IEEE J. Solid-State Circuits* **46**(3), 583–597 (2011)
45. C.-I.S. Hsieh, J.Y.-c. Liu, A low phase noise 210-GHz triple-push ring oscillator in 90-nm CMOS. *IEEE Trans. Microwave Theory Tech.* **66**(4), 1983–1997 (2018)
46. R. Kananizadeh, O. Momeni, A 190-GHz VCO with 20.7% tuning range employing an active mode switching block in a 130 nm SiGe BiCMOS. *IEEE J. Solid-State Circuits* **52**(8), 2094–2104 (2017)
47. W. Volckaerts, M. Steyaert, P. Reynaert, A 120GHz quadrature frequency generator with 16.2GHz tuning range in 45 nm CMOS, in *Proceedings of the IEEE Radio Frequency Integral Circuits Symposium (RFIC)* (2013), pp. 207–210
48. A. Bevilacqua, F.P. Pavan, C. Sandner, A. Gerosa, A. Neviani, Transformer-based dual-mode voltage-controlled oscillators. *IEEE Trans. Circuits Syst. II Express Briefs* **54**(4), 293–297 (2007)
49. A. Mazzanti, A. Bevilacqua, On the phase noise performance of transformer-based CMOS differential-pair harmonic oscillators. *IEEE Trans. Circuits Syst. I Regul. Pap.* **62**(9), 2334–2341 (2015)
50. X. Liu, H.C. Luong, A 170-GHz 23.7% tuning-range CMOS injection-locked LO generator with third-harmonic enhancement. *IEEE Trans. Microwave Theory Tech.* **68**(7), 2668–2678 (2020)
51. S. Rong, H.C. Luong, Analysis and design of transformer-based dual-band VCO for software-defined radios. *IEEE Trans. Circuits Syst. I Regul. Pap.* **59**(3), 449–462 (2012)
52. M. Vigilante, P. Reynaert, On the design of wideband transformer-based fourth order matching networks for *E*-band receivers in 28-nm CMOS. *IEEE J. Solid-State Circuits* **52**(8), 2071–2082 (2017)
53. A. Goel, H. Hashemi, Frequency Switching in Dual-Resonance Oscillators. *IEEE J. Solid-State Circuits* **42**(3), 571–582 (2007)
54. X. Li, S. Shekhar, D.J. Allstot, Gm-boosted common-gate LNA and differential Colpitts VCO/QVCO in 0.18- μ m CMOS. *IEEE J. Solid-State Circuits* **40**, 2609–2619 (2005)
55. M. Vigilante, P. Reynaert, A Coupled-RTWO-Based subharmonic receiver front end for 5G *E*-Band backhaul links in 28-nm bulk CMOS. *IEEE J. Solid-State Circuits* **53**, 2927–2938 (2018)

Chapter 4

Power Amplification



4.1 Introduction

The power amplifier is the last active stage of the transmitter chain. The purpose of this amplifier, as the name indicates, is to amplify power and reach the required power level, to communicate with the receiver at enough SNR. If more power can be transmitted, the distance between the transmitter and the receiver can be increased. It is therefore interesting to look at what limits the maximum output power. Because of the large required output power levels, this component uses a major piece of the transmitter's power budget. Consequently, it is also critical to look at its efficiency. As the power amplifier is pushed toward higher output powers and higher efficiencies, it is critical that no other factors such as distortion would limit the signal's integrity and allow the receiver to do its job.

Figure 4.1 shows the peak power and peak PAE of the current CMOS PAs in the literature. A lot of research has been done to create high-power, high-efficient PAs at frequencies up to 60 GHz. However, as the frequency increases, both output power and efficiency decrease rapidly. Furthermore, the amount of research also diminishes.

4.2 Power Amplifier Basics

4.2.1 Gain and Power

In Chap. 2, conjugate matching is discussed. This describes the optimal load impedance for the maximal power transfer and thus the maximum power gain from the source (the amplifier) to the load (antenna). The output impedance of the amplifier used for conjugate matching is a small-signal model and does not assume

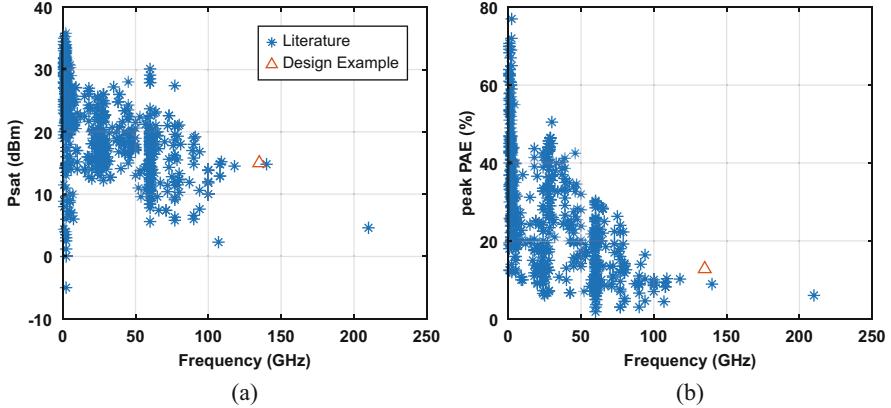


Fig. 4.1 An overview of published power amplifiers, (a) saturated output power and (b) peak PAE over frequency [1]

any limits on the maximum current and voltage of the amplifier. The maximum power deliverable to the load must take these large signal transistor parameters into account. The power is thus limited by the maximum current and the maximum voltage, and a transistor can deliver. Figure 4.2 shows the current through a transistor over the large signal gate and drain voltages. The load-line theory allows us to estimate the optimal real load impedance for maximal output power. The transistor is operated in the strong inversion and velocity saturation region. The minimal drain–source voltage is limited by the knee voltage, and the maximal drain–source voltage is limited by the reliability of the technology. The maximal voltage swing is thus restricted to $2 \cdot VDD - V_{knee}$. The maximal current, I_{max} , is set by the size and thus limited by other factors. This results in the following load impedance:

$$R_{max} = \frac{2 \cdot VDD - V_{knee}}{I_{max}}. \quad (4.1)$$

The difference between gain and power match is further shown in Fig. 4.3. The figure shows the output power as the input power is increased for both a conjugate match and a power match. It is clear that at small input powers, the small-signal model is valid and the maximum power is transferred into the load. This results in the maximum gain and a larger power compared to the power-matched amplifier. However, as the input power is increased, the large signal characteristic becomes more dominant, and the gain-matched transistor starts to saturate early compared to the power match. For larger input powers, the power match can deliver more output power.

One can also interpret this as once the amplifier starts to go into saturation, the output voltage swing is limited to approximately two times the supply voltage. The only method to get more power at full swing is by reducing the output impedance as is the case in the power-matched amplifier.

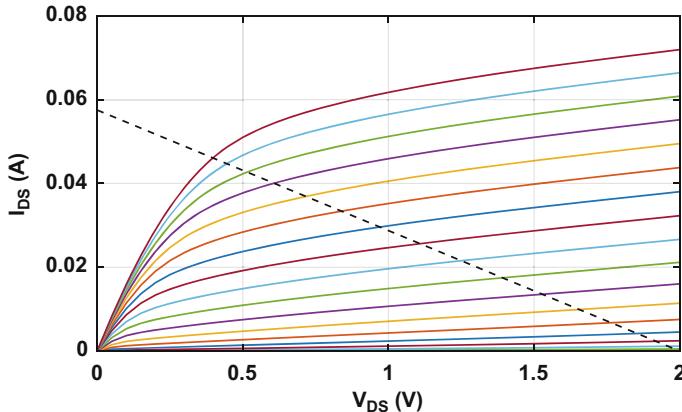


Fig. 4.2 Simulated load line of a 40 nm, 80 μm transistor, resulting in a R_{max} of 35 Ω

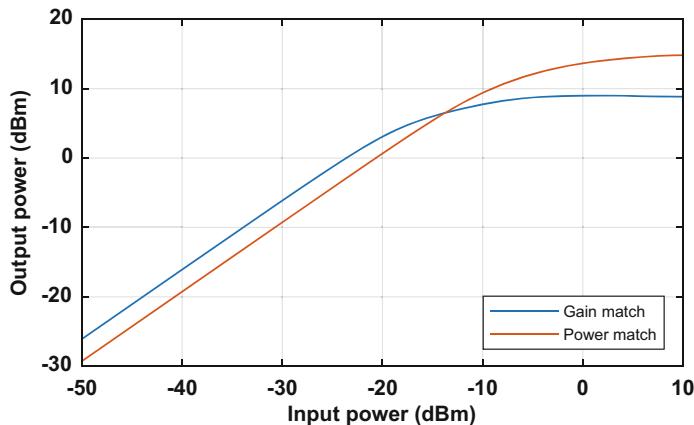


Fig. 4.3 The output power for an amplifier in gain and power match

4.2.2 Efficiency

The power amplifier is the most power-hungry building block of the transmitter. The efficiency of the power amplifier will immediately affect the overall efficiency of the transmitter and thus impacts system-level constraints such as battery life, thermal dissipation, and supply. Figure 4.4 shows a 2 stage amplifier with its DC-power and RF-power flows. A first efficiency metric used in power amplifiers is the drain efficiency (DE). The DE indicates the efficiency of the output stage and assumes a large enough power gain such that the input RF power can be ignored. The DE is expressed as follows:

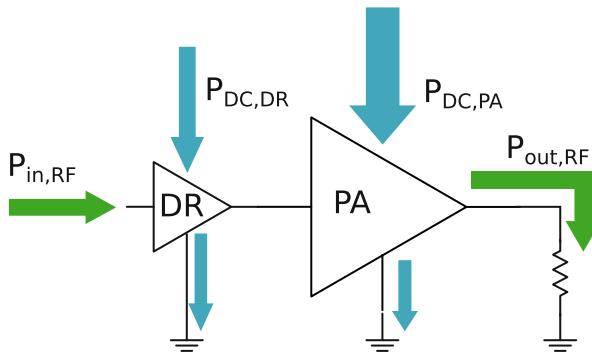


Fig. 4.4 The output power for an amplifier in gain and power match

$$DE = \frac{P_{out,RF}}{P_{DC,PA}}. \quad (4.2)$$

Unfortunately, at mm-wave frequencies especially above 100 GHz, the gain of the amplifier becomes limited. The DE is a bad indicator of the efficiency of the PA. A better metric is the power added efficiency (PAE). This metric also looks at how much RF power is needed to drive a PA and will thus also take into account the power consumption of driver stages, needed to increase the overall gain of the PA. The PAE for a PA with driver is given as follows:

$$PAE = \frac{P_{out,RF} - P_{in,RF}}{P_{DC,PA} + P_{DC,DR}}. \quad (4.3)$$

4.2.3 PA Linearity

A perfect power amplifier would take the input signal and amplify this to a perfect higher output power copy. This can be achieved as long as the power amplifier is used in the linear region away from the saturation point. However, this underutilizes the power amplifier, results in bad efficiency, and thus comes at the cost of large dc-power consumption.

The first major type of distortion is amplitude-to-amplitude distortion (AM–AM). This distortion is due to gain compression, as the power amplifier starts to saturate. A large signal measurement to indicate the AM–AM is the output referred 1 dB compression point (OP1dB). This is the output power at which the gain is compressed by 1 dB. The effect of gain compression on a constellation diagram for 16-QAM is shown in Fig. 4.5. The outer and largest amplitude constellation points move to the center.

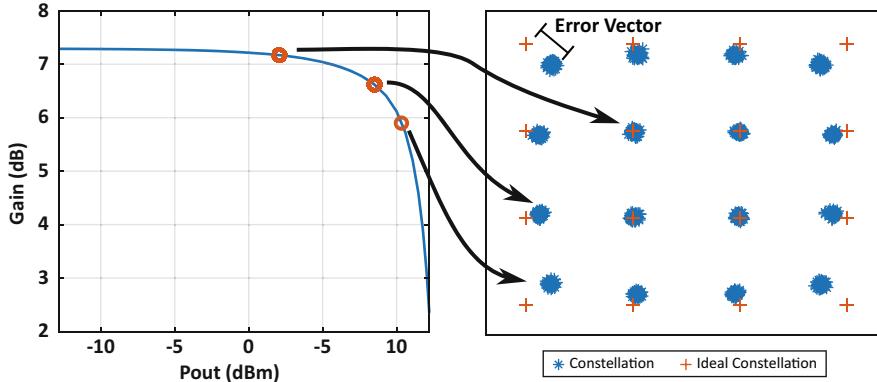


Fig. 4.5 The effect of AM–AM distortion on a 16-QAM constellation

A second major type of distortion is amplitude-to-phase distortion (AM–PM). For large signal outputs, the phase of the output changes causing the constellation to rotate depending on the symbol's magnitude as shown in Fig. 4.6. The origin of AM–PM distortion is less straightforward than the AM–AM distortion. The main cause of AM–PM distortion has been attributed to a non-linear input capacitance C_{in} and gain compression in combination with C_{GD} [2–5].

Both AM–AM and AM–PM distortions can happen at the same time as shown in Fig. 4.7. The total error of the transmitted symbol compared to its ideal constellation is captured in the error vector magnitude (EVM). The error vector is defined as the difference between the ideal and transmitted or measured symbols. For the RMS EVM, the average error over N symbols is compared to the RMS power of the ideal constellation of M symbols. Summarized in the following expression [6]:

$$\text{EVM}_{\text{RMS}} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N |S_{\text{ideal},i} - S_{\text{meas},i}|^2}}{\sqrt{\frac{1}{M} \sum_{i=1}^M |S_{\text{ideal},i}|^2}} \quad (4.4)$$

$$= \frac{V_{\text{error, RMS}}}{C_{\text{RMS}}} \quad (4.5)$$

4.2.4 PA Classes and Biasing

Power amplifiers have been categorized into many classes: A, AB, B, C, D, E, etc. The main focus of these classes is to improve the efficiency of the PA

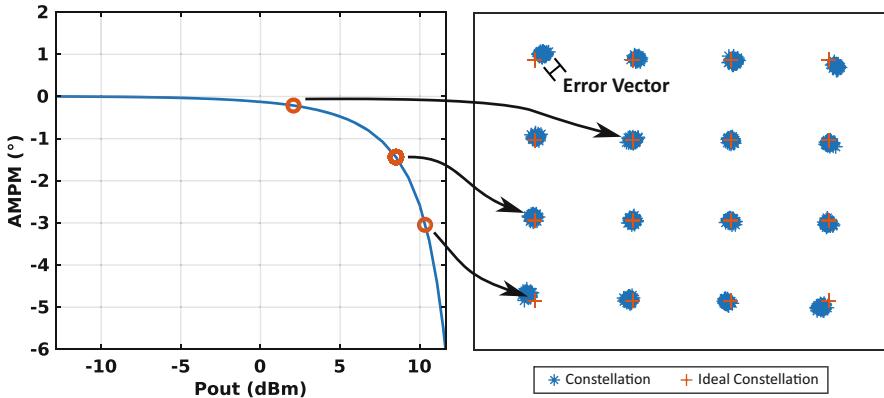


Fig. 4.6 The effect of AM–PM distortion on a 16-QAM constellation

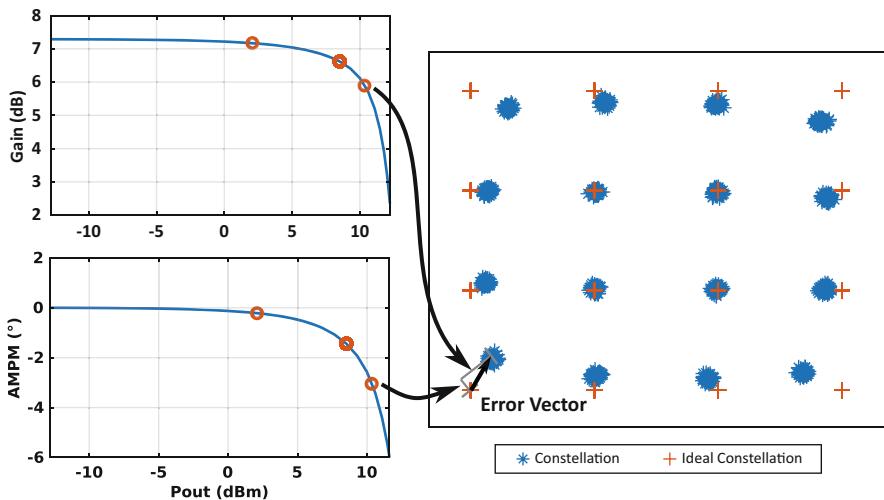


Fig. 4.7 The effect of AM–AM and AM–PM distortions on a 16-QAM constellation

by adjusting either the bias voltage, the harmonic content, or both [2, 7]. The manipulation of harmonics at the output of a PA is extremely difficult to achieve at mm-wave frequencies and especially above 100 GHz. First, these techniques require significantly higher harmonic content. This content is difficult to achieve due to the already low gain and large absolute frequency spacing to the second and third harmonics. Second, to sustain these higher harmonics, low loss higher-order matching networks need to be made, which are difficult in bulk CMOS technology.

For the reasons above, only the four biasing classes A, AB, B, and C are considered. These four are distinguished by their gate bias condition. The gate bias will determine how long the transistor will conduct current during a cycle, and this is represented by the conduction angle θ . The voltage and current waveforms for the 4 respective classes are shown in Fig. 4.8. Class A has a bias voltage well above the

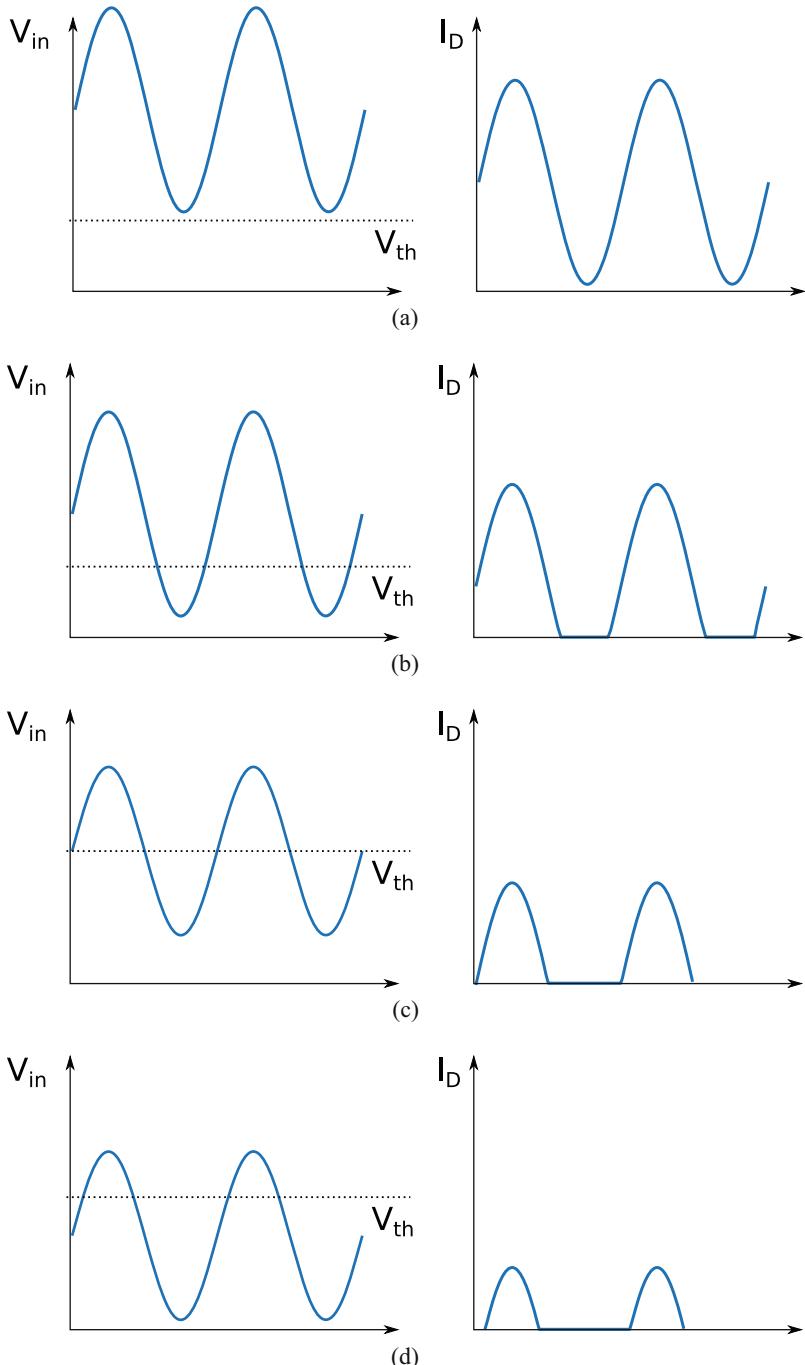


Fig. 4.8 Input voltage and drain current waveforms for (a) Class A, (b) Class AB, (c) Class B, and (d) Class C PA

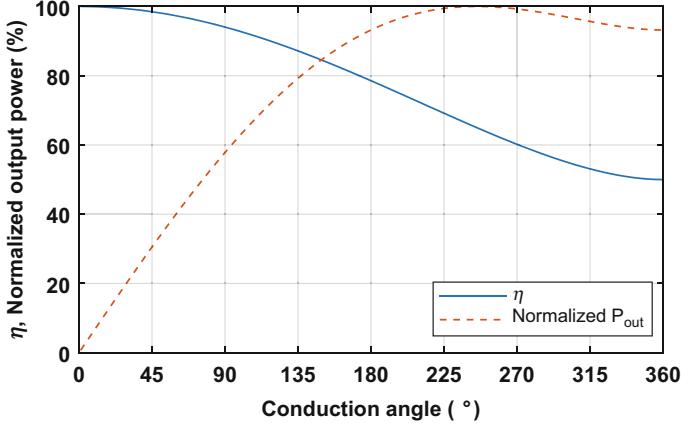


Fig. 4.9 The theoretical efficiency and normalized output power for different conduction angles

threshold voltage to ensure full conduction over the whole cycle. This results in a conduction angle of $\theta = 360^\circ$. Class B puts the bias voltage at approximately the threshold voltage, and this results in conduction during only half of the cycle. This results in a conduction angle of $\theta = 180^\circ$. The reduced conduction angle results in improved efficiency. Class AB is referred to as a bias point between Classes A and B or a conduction angle between $\theta = 180^\circ$ and $\theta = 360^\circ$. This allows a trade-off between linearity and efficiency. The theoretical efficiency and output power dependence to the conduction angle can be described as follows [2]:

$$\eta = \frac{1}{4} \cdot \frac{\theta - \sin \theta}{\sin(\theta/2) - (\theta/2) \cos(\theta/2)} \quad (4.6)$$

$$P_{out} \propto \frac{\theta - \sin \theta}{1 - \cos(\theta/2)}. \quad (4.7)$$

The theoretical efficiency and output power are shown in Fig. 4.9. For a class A PA, the maximal efficiency is 50%. When the conduction angle is reduced from class A toward class B, the efficiency improves from 50% up to 78.5% at Class B.

The theoretical look at bias classes up till now only discusses the efficiency and output power. However, the gate bias voltage also affects the small-signal gain and linearity of the power amplifier [8–10]. A common-source capacitive neutralized amplifier as shown in Fig. 4.19 is biased at different gate voltages between 0.3 V and 0.7 V. The amplifier is sized at 2 transistors, and each build-up of 12 unit transistors sized at 6 fins and 15 fingers. The large signal simulations for this power amplifier stage at 140 GHz are shown in Fig. 4.10. The maximum achievable gain for this amplifier is limited to 7.5 dB. For a reduced gate bias voltage, the small-signal gain

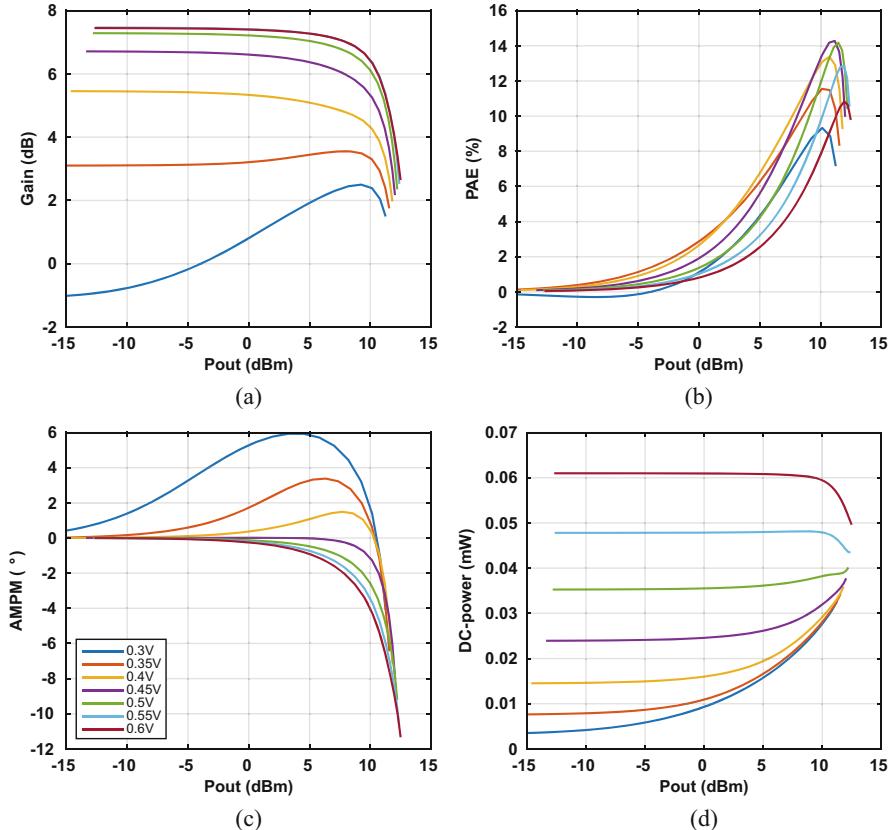


Fig. 4.10 The effect of gate bias on (a) gain, (b) PAE, (c) AM-PM, and (d) DC-power consumption over the output power at 140 GHz

reduces, once the bias voltage hits 0.35 V, the AM-AM linearity has improved, and a small gain expansion for large output power is observed. Further reducing the bias voltage toward class B results in larger gain expansion but a negative small-signal gain. The AM-AM linearity and expansion come with a similar characteristic for the output phase change. However, biasing deep toward class-A results in early AM-PM compression; although the gain curves stay fairly unchanged, the AM-PM compression further regresses resulting in more AM-PM distortion and worse linearity.

The trade-off between gain and linearity is well known and used at lower mm-wave frequencies to design high linearity power amplifiers [9, 10]. Unfortunately, above 100 GHz, there is little headroom in small-signal gain to exchange it for better linearity. This becomes more obvious if we look at a summary of the large signal parameters over the gate bias voltage in Fig. 4.11. As the bias is reduced, the gain starts to reduce by the time, the 1 dB output referred compression point has moved

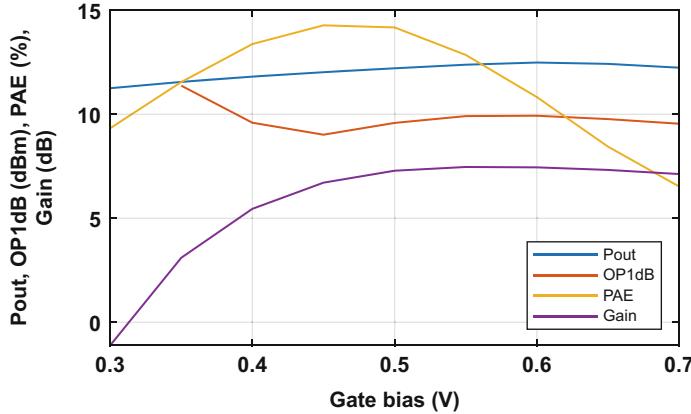


Fig. 4.11 The summarized effect of gate bias for a 140 GHz amplifier stage on the main PA specifications

closer to the saturated output power, and the gain has dropped below 5 dB. The low gain will negatively impact the area as more driver stages will be necessary and reduce the peak PAE. The peak PAE indeed starts to reduce for low bias voltage. The theoretical reduction of bias voltage to improve the efficiency thus does not hold at mm-wave.

4.2.5 Challenge at mm-Wave in a Deeply Scaled Technology

As discussed in Chap. 2, scaling has improved transistors up to a point where they can be used for mm-wave circuits beyond 100 GHz. Unfortunately, the performance has started to stagnate for mm-wave transistors as the loss in the interconnects starts to dominate. Furthermore, the decreased channel lengths come with a reduction in maximum power supply voltage and thus further limits the maximum linear output swing and power for a fixed current. Furthermore, it is shown that the gain degrades rapidly with frequency putting more strain on the circuits.

The combination of all these factors results in PAs with reduced output power, PAE, and linearity above 100 GHz.

4.2.5.1 The Effect of Drain Resistance

In Chap. 2, the main focus was put on the effect of the gate resistance on the transistor's performance. However, as the frequency increases, the parasitic drain resistance in combination with the drain–source capacitance results in a reduced output impedance. If the parasitic drain resistor, R_d , is added to the small-signal

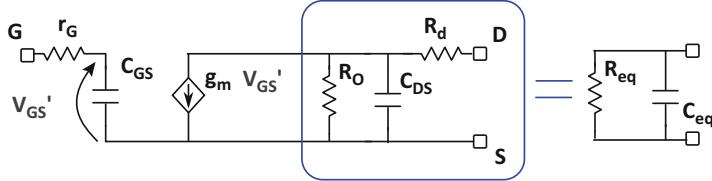


Fig. 4.12 The small-signal model with parasitic drain resistance and its equivalent reduction

model of Chap. 2 as shown in Fig. 4.12, the model can be converted back to its original form with an equivalent output resistor, R_{eq} , and capacitor, C_{eq} . These equivalent output resistor, R_{eq} , and capacitor, C_{eq} , can then be written as follows and simplified assuming $R_d \ll R_o$:

$$R_{eq} = \frac{\omega^2 C_o^2 R_o^2 R_d^2 + R_o^2 + 2R_o R_d + R_d^2}{\omega^2 C_o^2 R_o^2 R_d + R_o + R_d} \quad (4.8)$$

$$\approx \frac{\omega^2 C_o^2 R_o R_d^2 + R_o}{\omega^2 C_o^2 R_o R_d + 1} \quad (4.9)$$

$$C_{eq} = \frac{C_o R_o^2}{\omega^2 C_o^2 R_o^2 R_d^2 + R_o^2 + 2R_o R_d + R_d^2} \quad (4.10)$$

$$\approx \frac{C_o}{\omega^2 C_o^2 R_d^2 + 1}. \quad (4.11)$$

The effect of R_d on the equivalent output impedance over frequency is shown in Fig. 4.13, and the output impedance values are estimated from a parasitic extracted transistor model. The extracted values result in a $C_o = 41 \text{ fF}$, $R_o = 300 \Omega$, and $R_d = 3.5 \Omega$. The capacitance stays fairly constant over frequency, while the equivalent output resistance drops. The simulated model follows the trend of the parasitic extracted transistor model.

The effect of the magnitude of R_d on the reduction of output impedance is shown in Fig. 4.14. At low frequencies, below 50 GHz, the effect of R_d on the output impedance is minimal. However, as the frequency increases above 100 GHz, the output impedance is reduced significantly. This frequency dependence is of course dependent on the magnitude of the parasitic drain capacitance, and a larger drain capacitance will result in a larger penalty on the drain resistance.

The reduced output impedance will impact the performance of the PA on different fronts. The first and most straightforward is reduced gain as given in Eq. (2.8) when a conjugate match is applied. A second problem is a reduction in output power as proportionally more power is consumed in R_{eq} instead of the load. The reduction in useful power in the load will result in a reduction in PAE. Another way of looking at this is as a reduction in the quality factor of the output and thus an increased loss, when used in a tuned design as is the case for all mm-wave designs.

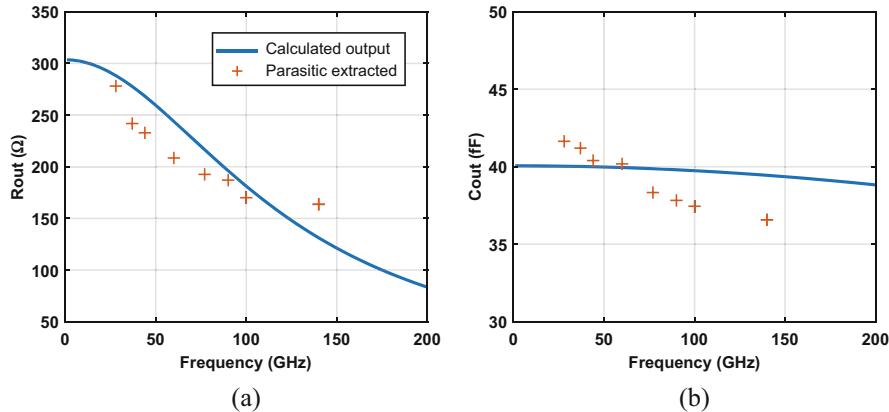


Fig. 4.13 Comparison of the simulated output model and parasitic extracted results for the output impedance (a) and capacitance (b)

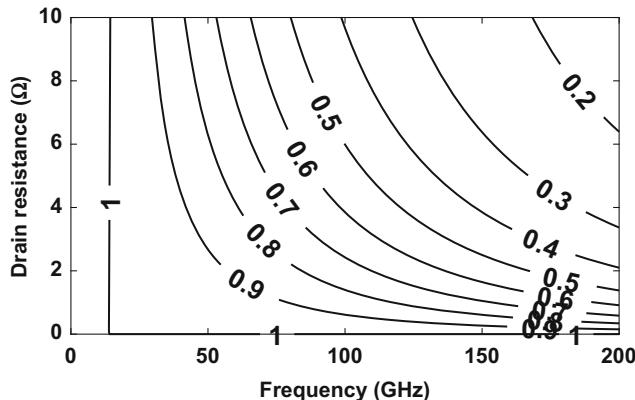


Fig. 4.14 The effect of parasitic drain resistance on the output impedance expressed as the ratio R_{eq}/R_o or the reduction in output resistance

This effect worsens with smaller technology nodes, as more metal layers and thinner metals result in larger interconnect resistances, and the parasitic output capacitance increases. Although a better channel control in FinFET results in large R_o , the increased interconnect resistance and parasitic capacitance reduce these gains at higher frequencies.

The effects on the output impedance at higher frequencies can be further demonstrated by looking at the difference between the gain or PAE contours and the power contours of the same amplifier as shown in Fig. 4.15. At 25 GHz, there is a big difference in optimal impedance for PAE and output power. The amplifier is thus open to load modulation or gain can be exchanged for extra output power. For increasing frequency to 60 GHz and 120 GHz, the difference between the optimal

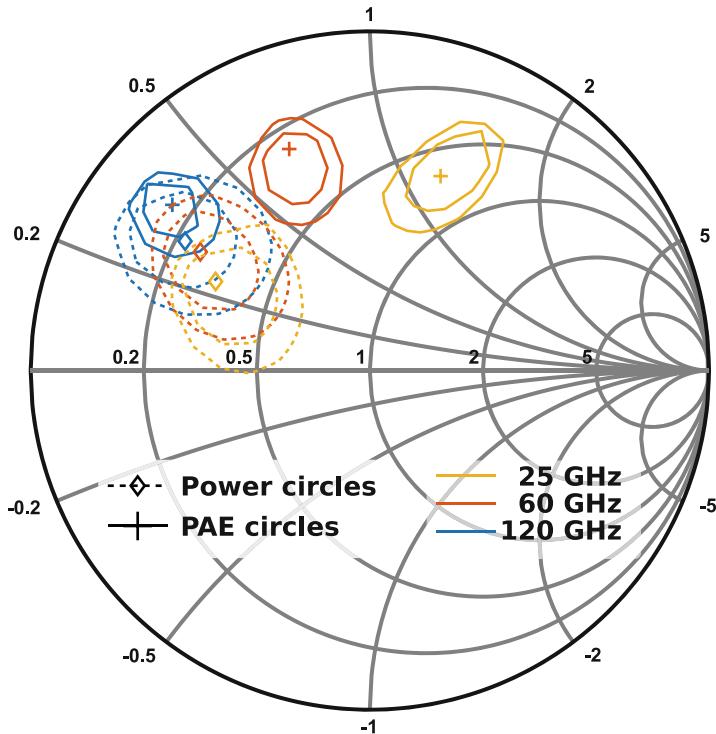


Fig. 4.15 The power and peak PAE contours for a common-source amplifier at 25 GHz, 60 GHz, and 120 GHz

impedance for PAE and the power diminishes to a point where there is barely any difference and both cannot be exchanged. The main reason for this is a reduction of the output impedance of the amplifier at increased frequencies.

4.2.5.2 Limits on Output Power

If we look at CMOS mm-wave PAs, there is a clear trend in the output power over the operating frequency as seen in Fig. 4.1.

The output power is defined by the maximum voltage swing and the maximum output current. The voltage swing is limited by the reliability of the technology. For decreasing gate lengths, the supply has been reduced. A method to increase the supply is by employing transistor stacking [11–13]. Multiple stacked transistors allow the supply to increase by distributing the supply load over multiple transistors.

The maximum current is set by the total width of the transistor. However, as the total width of the transistor is increased by a factor n , the input and output capacitances also increase with a factor n . In a tuned amplifier, the output

capacitance is resonated out by an inductance. The inductance needed to resonate out the input or output capacitance is given as

$$L = \frac{1}{4\pi^2 C f^2} = \frac{1}{4\pi^2 n C_{unit} f^2}. \quad (4.12)$$

The required inductance thus decreases linearly with increased transistor size. Unfortunately, for mm-wave circuits, as we try to increase the frequency, the required inductance reduces with the square of the frequency. For example, an equally sized amplifier at 140 GHz compared to 28 GHz will require a 25 times smaller inductance to tune out the same input or output capacitors. The main issue with small inductances is their inherent high loss when implemented. As a transformer needs to be sized down to reduce the inductance, at a certain point, the insertion loss of the transformer increases drastically as shown in Fig. 4.16. The

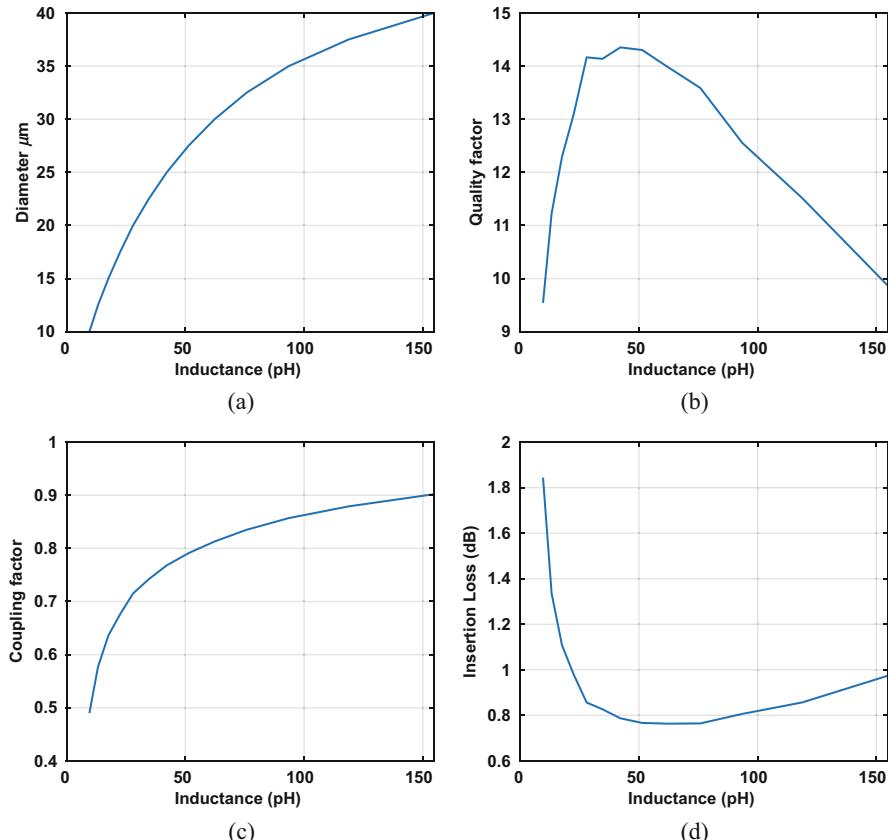


Fig. 4.16 (a) The diameter of a one-to-one transformer for a certain inductance at 140 GHz. (b) The quality factor, (c) mutual coupling, and (d) insertion loss for differently sized transformers

drastic increase in IL of a transformer puts a practical lower limit on the inductance. The lower limit in inductance therefore also places a maximum capacitance limit, and limits the maximum size and output power of a single amplifier stage.

4.3 Design Example: A Class-A Two-Way Power Combining D-Band PA in 16 nm FinFET

As a design example, a D-band power amplifier in a 16 nm FinFET technology is designed and measured after fabrication. The challenges of designing in 16 nm FinFETs for high gain and high power above 100 GHz are tackled and overcome. The maximum power limit of a single stage is met as described in the previous section. The power is further extended by employing a power combining scheme, and the powers of two output amplifiers are combined to further extend the output power by 3 dB. The best power combining topology is chosen and designed.

4.3.1 Design of the 16 nm FinFET Transistor

In Chap. 2, the basic design trade-offs of a transistor for mm-wave applications are described. It is clear that to maximize the performance of the transistors, we need to minimize the parasitic gate resistance. As large transistor widths are required to create large output powers, the layout parameters need to be optimized for low gate resistance. For this reason, power amplifier transistors sometimes have larger finger widths compared to LNAs. Unfortunately due to scaling of the BEOL, a 16 nm FinFET technology shows a dramatic decrease in transistor performance as the size increases. To demonstrate the penalty paid when designing with 16 nm FinFETs compared to bulk 40 nm transistors, let us take a regular capacitively neutralized differential common-source amplifier. The size of the transistors is increased by increasing the number of fingers. For each size, the small-signal gain is reported against its output referred 3 dB compression point in Fig. 4.17. For a 40 nm technology, finger widths of 1 μm and 1.4 μm are used. Increasing the size of the transistor has a very limited effect on the gain indicating that the interconnect gate resistance is not the limiting factor. Only when reaching about 10 dBm of output power, the gain starts to reduce. Furthermore, the benefit from a larger finger width becomes apparent as the transistor with 1.4 μm fingers now crosses over the transistor with 1.0 μm fingers.

For a 16 nm technology, the gain starts to reduce immediately as the number of fingers is increased. Attempting to change the trade-off between fins and fingers by increasing the number of fins per finger (finger width) from 6 to 8 does not improve the performance. The described performance penalty is mainly due to the resistive interconnects in the thin lower metals as large transistors will result in long

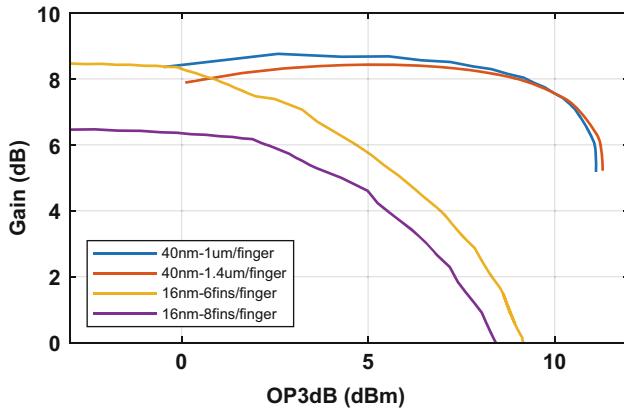


Fig. 4.17 The small-signal gain and output referred 3 dB compression point for a 40 nm and 16 nm amplifier

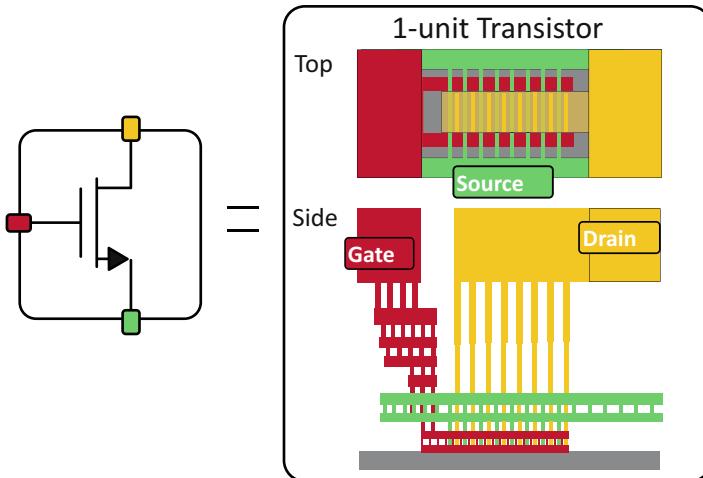


Fig. 4.18 Layout of a unit transistor of 6 fins and 15 fingers used in all stages

highly resistive interconnects. This penalty can be minimized by combining small transistors with a low number of fins and fingers in the upper low resistive metals. A unit transistor of 6 fins per finger and 15 fingers is chosen. The transistors layout, as shown in Fig. 4.18, goes up to the first thicker metal and is designed to allow easy tiling to increase the total transistor size. The added loss of the metalization reduces the gain from 7.5 dB to 6.9 dB.

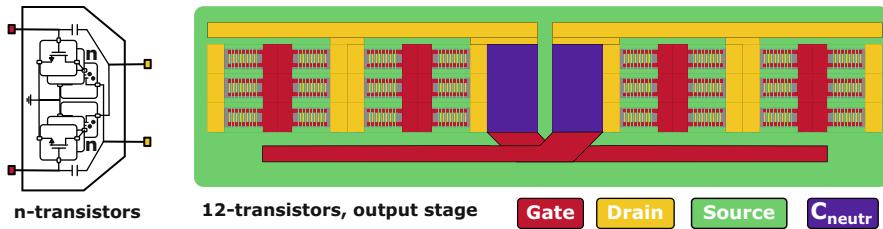


Fig. 4.19 Layout of the capacitively neutralized differential common-source output stage. Each branch is the build-up of 12 unit transistors

4.3.2 Build-up of the Power Amplifier

The output power of the power stage is limited by the design of the output transformer. At 140 GHz, this limit is set at combining 12 unit transistors into a single transistor as shown in Fig. 4.19. These transistors are combined in a common-source amplifier with two neutralization capacitors of 31 fF. The output impedance, Z_{out} , of this amplifier is $Z_{out} = 4 - 21j$ or has an equivalent output capacitance of 52 fF. Due to parasitic output resistance, the conjugate match and power match are almost identical. From Eq. (4.12), the required inductance should only be 25 pH, which in Fig. 4.16 is right at the edge where the IL starts to increase dramatically.

4.3.2.1 Power Combining to Extend the Power Limit

To increase the output power above the limit of a single amplifier stage, the power of multiple stages can be combined. Most CMOS power amplifiers rely on this technique due to the low supply voltage. At higher frequencies, this becomes even more crucial as the power of a single stage is further limited. Power combining can be implemented using several methods [12, 14–17].

A first distinction is between isolating and non-isolating power combiners. An isolating combiner ensures isolation between the combining amplifiers. Any change or defect of a power stage will not affect any other stage. If one power stage is turned off, the output impedance seen by the other amplifiers stays the same, and they can continue to deliver power as before. A Wilkinson power combiner or a hybrid coupler can be used as isolating power combiners [14]. Despite the beneficial isolation, all isolating combiners employ $\lambda/4$ transmission lines and a resistor. Therefore, they have a larger footprint and have a large insertion loss compared to non-isolation combiners.

On the other hand, non-isolating combiners provide no isolation between the amplifiers. Since all amplifiers influence each other, extra care must be taken during the design. One of the most straightforward power combining methods is zero-length transmission line or parallel power combining where all outputs are assumed to be in phase and all currents are added [18]. If we combine n number of amplifier

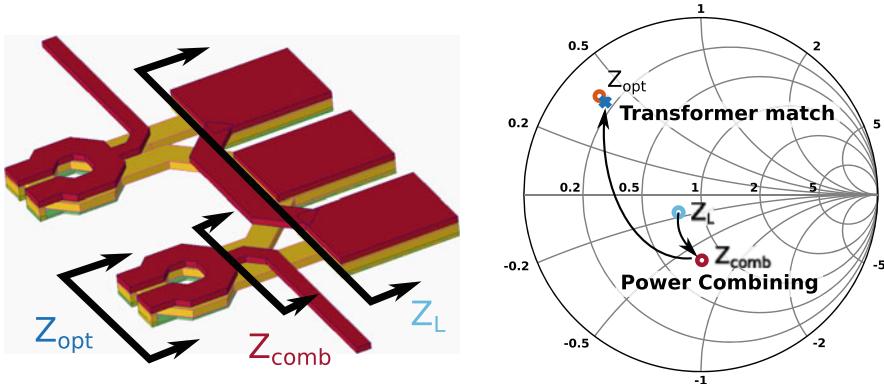


Fig. 4.20 Layout of the output matching network and its intermediate impedances shown on the Smith chart

stages, every amplifier sees the same output voltage, but the output current is multiplied with a factor n , the resulting output impedance seen by the amplifier is then also n times the output load. The opposite and dual case is also possible where voltages are added but the current is kept the same. In this case, the resulting output impedance is reduced with a factor n . Both cases are summarized below.

$$Z_{Comb, Series} = \frac{Z_L}{n} \quad (4.13)$$

$$Z_{Comb, parallel} = n \cdot Z_L. \quad (4.14)$$

Therefore, power combining can be employed as part of the impedance transformation network. The impedance transformation from a parallel power combiner helps create a low loss output matching network of the PA and increases the impedance seen by each individual PA. The output matching network of the 2-way power combining network and transformer matching is shown in Fig. 4.20. The 50Ω load including the capacitance of the probe pads transforms the load impedance, Z_L , into $Z_L = 38 - 8j$ or 40Ω in parallel with 6 fF at 140 GHz . The optimal load seen by the power amplifiers is $Z_{opt} = 8.5 + 21j$ or 68Ω in parallel with 25 pH . From this, it is clear that by employing a parallel power combiner a factor 2 impedance transformation is performed to bring the load impedance seen after the power combiner, Z_{Comb} , close to the desired Z_{opt} . Therefore, the output transformers have to perform almost no impedance transformation, allowing them to be designed as a highly coupled 1-to-1 transformer, resulting in a lower IL. To further reduce the insertion loss of the transformer, a highly coupled transformer is made by adding an extra winding in the secondary. This has resulted in a total insertion loss of only 1.14 dB for the whole output network.

4.3.2.2 Completing the Circuit with Drivers

To increase the gain of the power amplifier, the output stage is preceded by 2 sequential driver stages. The driver stages are scaled down to save power and improve efficiency. The driver stages are sized at 8 and 4 unit transistors with their layout shown in Fig. 4.21. Transformer matching is used in between stages to provide the impedance transformation and also the required bias and supply connections. To improve the stability, a resistor is placed at the center tap of each biasing inductor. The schematic of the full power amplifier is shown in Fig. 4.22 with transformer values shown in Table 4.1.

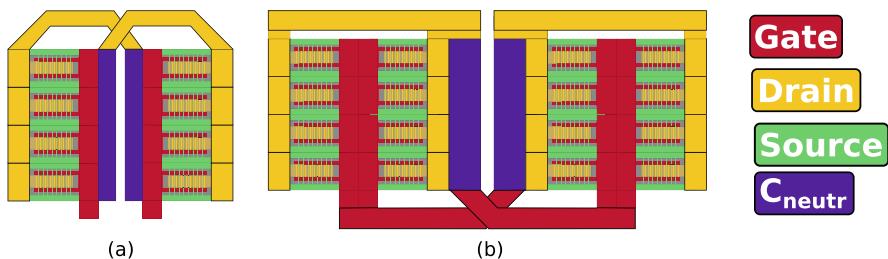


Fig. 4.21 The layout of the unit transistors and neutralization capacitor for the 4 unit input stage (a) and the 8 unit driver stage (b)

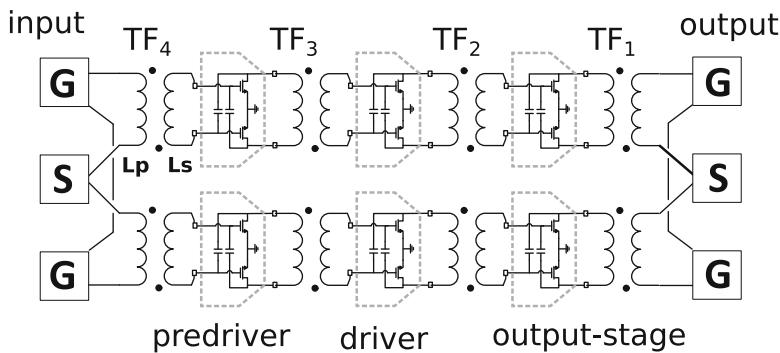


Fig. 4.22 The schematic of the full PA

Table 4.1 Matching
transformer parameters

	TF_1	TF_2	TF_3	TF_4
L_p (pH)	21	31	90	46
L_s (pH)	26	20	32	61
k_m	0.73	0.343	0.4	0.57

4.3.3 Measurement Results

The die micrograph of the fabricated D-band PA in a 16 nm FinFET CMOS technology is shown in Fig. 4.23. The active area of the PA measures 460 by 135 μm , and this covers the manual dummy filled area including the input and output RF probe pads.

All measurements are performed by probing the sample. The small-signal measurements are performed with a vector network analyzer (VNA) connected to ZVC F-band frequency converters or ZVC D-band frequency converters as shown in Fig. 4.24. The small-signal measurements performed for a single sample with both F-band and D-band measurement setups are shown in Fig. 4.25. At the overlap between both bands, the measurements show good correspondence providing extra confidence in the measurements and calibration. For stability, Rollet-stability factor and Mu1 values are shown in Fig. 4.26, and for both bands, unconditional stability is measured. The measurements are performed on multiple samples for increasing power supply. The measurement results for 12 samples at 0.8 V nominal supply and 1.0 V boosted supply are shown in Fig. 4.27. The PA shows a peak gain of 24 dB at

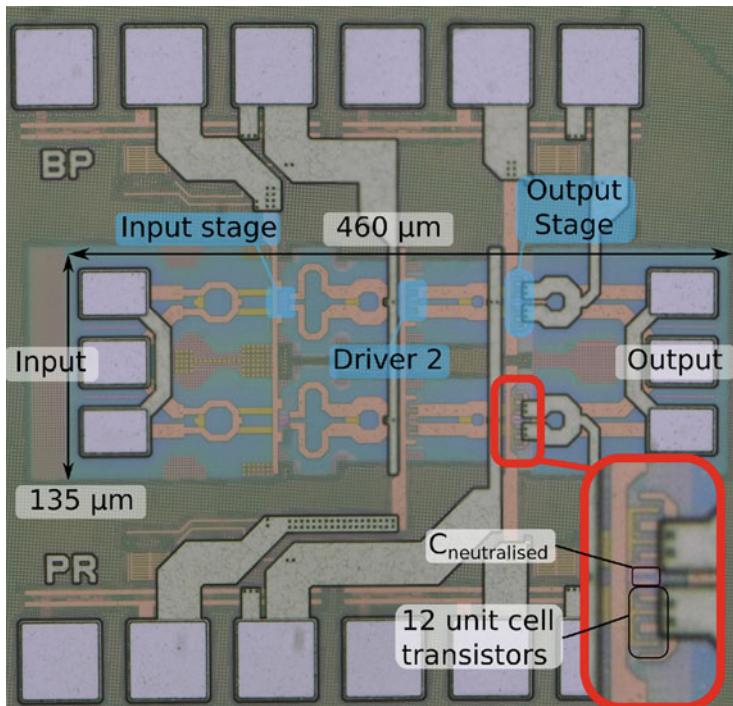


Fig. 4.23 Die micrograph of the produced power amplifier

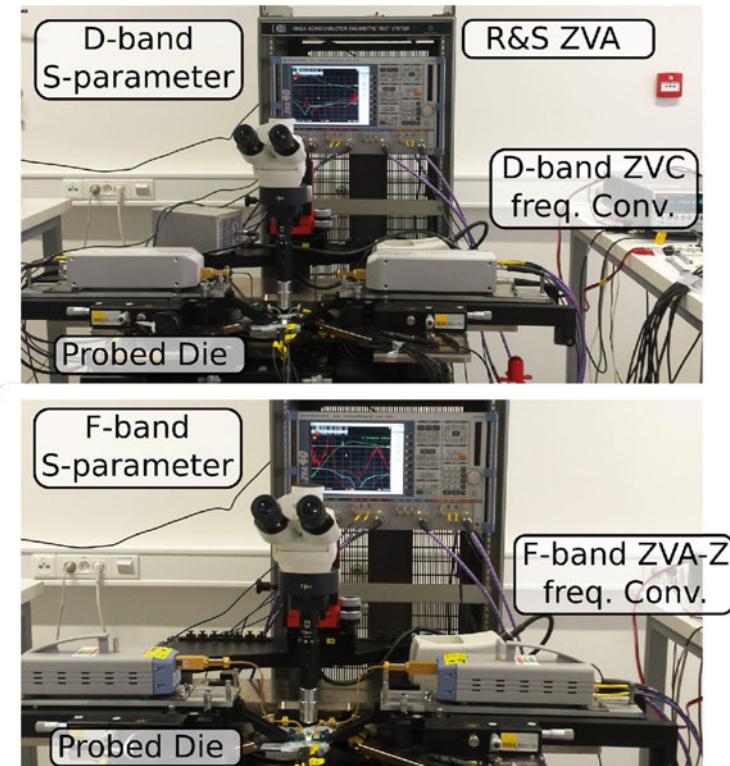


Fig. 4.24 Measurement setup for probed small-signal measurements

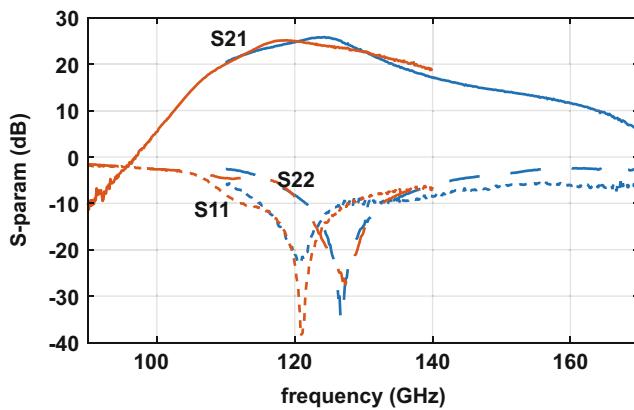


Fig. 4.25 Measured s-parameters using F- and D-band extender modules

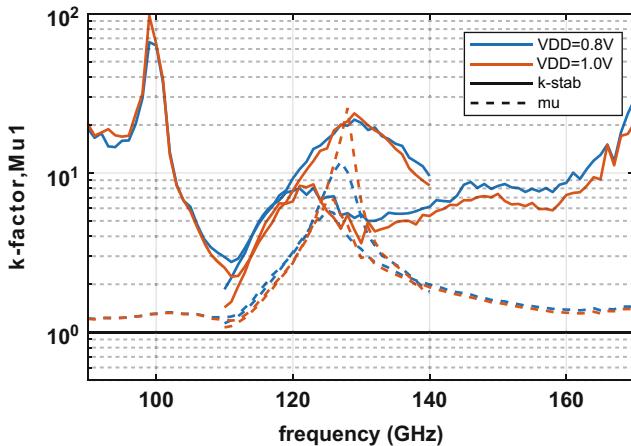


Fig. 4.26 Measured Rollet-stability factor and Mu₁ values for both F- and D-band

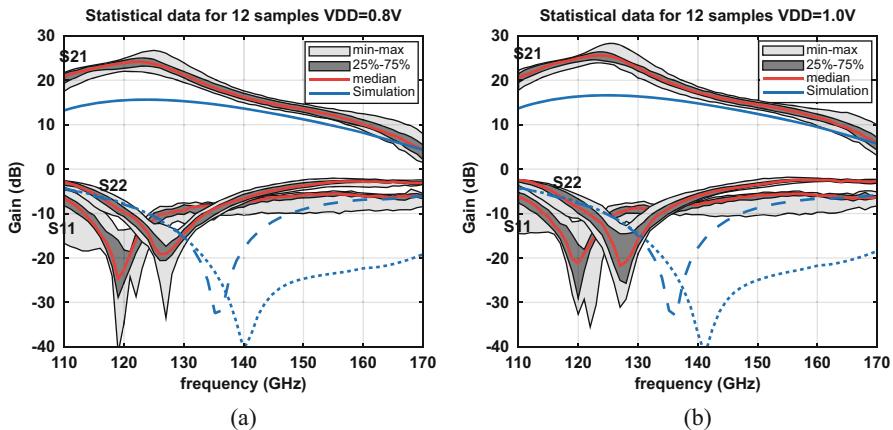


Fig. 4.27 Measured s-parameters for multiple samples' demonstration of the spread for a supply of 0.8 V(a) and 1.0 V(b)

115 GHz, with a 3-dB bandwidth of 22 GHz at 0.8 V. The gain increases to 24 dB for the boosted supply.

The large signal measurements are performed using a D-band ZVC frequency converter connected to a Keysight PSG to provide a variable RF-input power up to 5 dBm at the GSG input and a PM5 power meter to perform calibration and measure the output power. The measurement setup and calibration steps are shown in Fig. 4.28. The output power, large signal gain, PAE, and 1 dB compression point are shown against the provided input power in Fig. 4.29 for 115, 125, 135, and 145 GHz at a supply of 1.0 V. The saturated output power measures around 15 dBm with a peak PAE above 12% and a 1 dB compression point around 10 dBm output

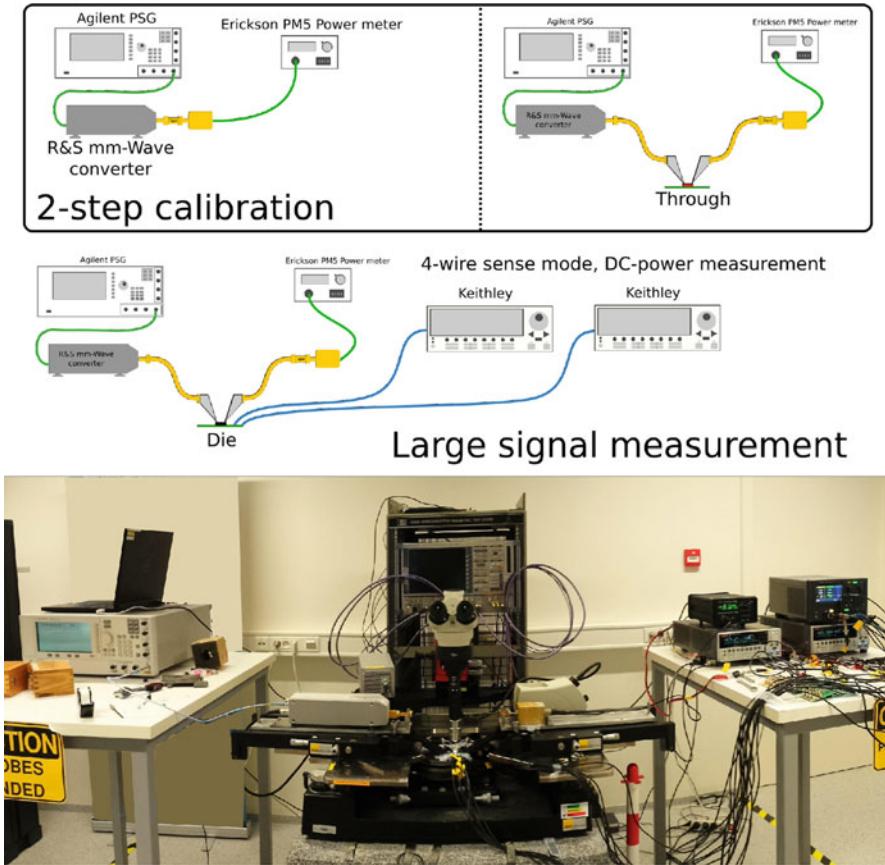


Fig. 4.28 Measurement setup for probed large-signal measurements

power. A summary of the measured saturated output power, large signal gain, PAE, and 1 dB compression point over frequency is shown in Fig. 4.30. The gain measured using the large signal measurement setup is compared to the gain in the small-signal measurement setup, and they show good correspondence except for the high part of the D-band. The DC-power consumption of the PA is shown in Fig. 4.30b, with a breakdown for the different stages. The fixed DC-power consumption of the output stage shows the class-A characteristic of the PA. Once more the measurements are performed for 13 samples, and the results for 0.8 V nominal supply and 1.0 V boosted supply are shown in Fig. 4.31. As expected, a boost in power supply benefits a PA dramatically.

The measured results are summarized and compared to the state-of-the-art D-band and FinFET power amplifiers in Table 4.2. Thanks to the proposed techniques, the PA shows a state-of-the-art mm-wave performance at D-band. The presented PA achieves similar or better performance compared to other finFET designs [19, 20] at

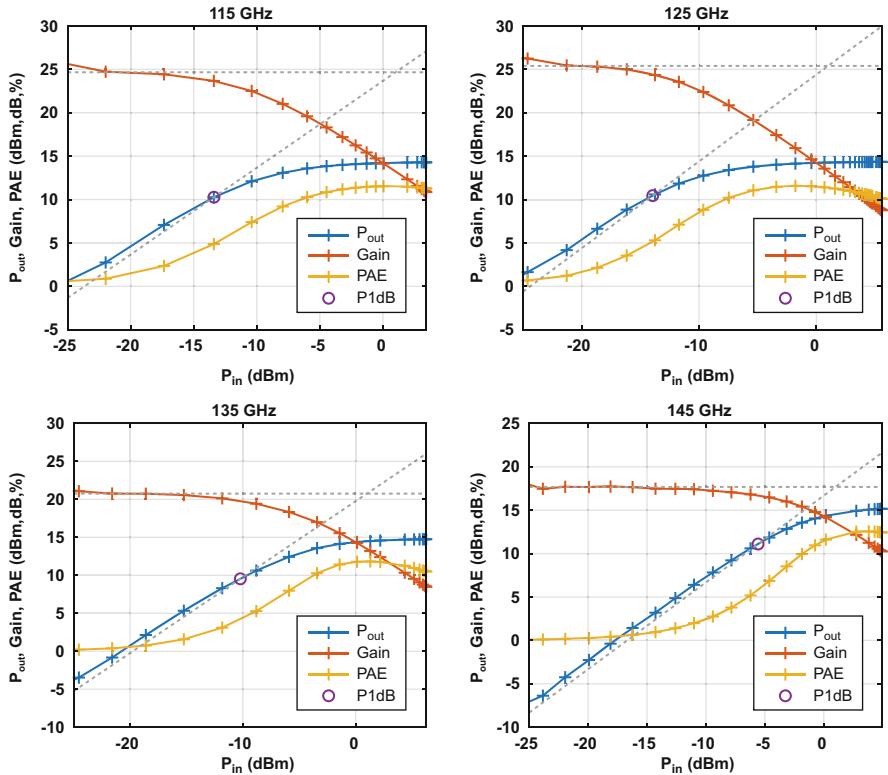


Fig. 4.29 Large signal measurements of P_{out} , Gain, PAE, and $P_{1\text{dB}}$ at 115, 125, 135, and 145 GHz

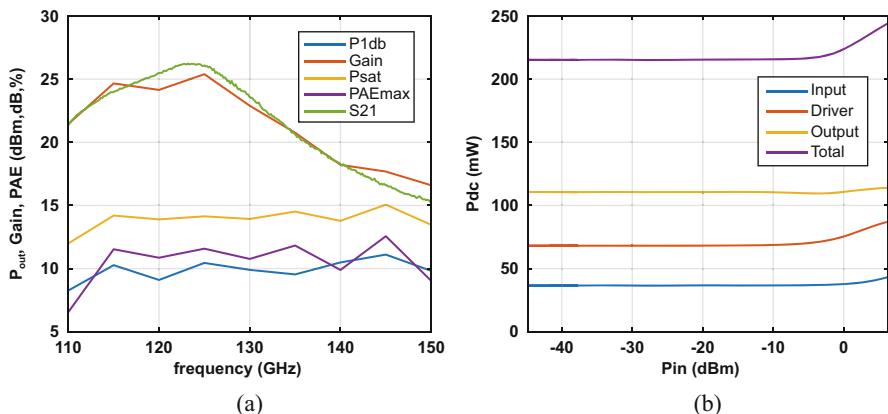


Fig. 4.30 Summary of large signal measurement of over frequency (a) and the DC-power consumption over input power (b)

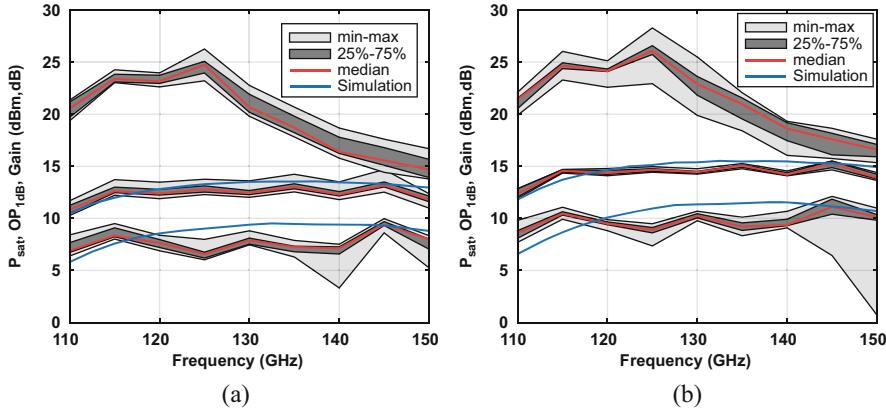


Fig. 4.31 Summary of large signal measurements for multiple samples' demonstration of the spread for a supply of 0.8 V (a) and 1.0 V (b)

Table 4.2 State-of-the-art power amplifier comparison table

Ref.	This work		RFIC'17 [19]	JSSC'19 [20]	RFIC'18 [21]	RFIC'15 [22]
Tech.	16 nm FinFET	14 nm FinFET	22 nm FinFET	40 nm CMOS	120 nm SiGe	
VDD (V)	0.8	1	1	1	1	3.5
Pdc (mW)	162	207	30	44.6	305	560
Freq. (GHz)	135	135	71	75	140	124
Gain (dB)	19	20.5	16.7	16.6	20.3	32
OP_{1dB} (dBm)	7.1	9.2	2	5.7	10.7	13.5
P_{sat} (dBm)	13.1	15	7.4	12.8	14.8	17.8
PAE _{max} (%)	11	12.8	8.9	26.3	8.9	4.3
Area (mm ²)	0.062		0.1	0.054	0.125	1.92

twice the frequency. The median performance for 13 samples shows a peak gain of 25.6 dB, a P_{sat} of 15 dBm, and a PAE_{max} of 11.7% for an area of only 0.062 mm². It is the first D-band PA in FinFET showing the capabilities of FinFET for future systems in this frequency range

4.4 Conclusion

This chapter has discussed the basics of mm-wave power amplifiers. The challenges of implementing high-frequency PAs in deeply scaled technologies, benefits, and issues were discussed. A design example of a power combining PA in 16 nm FinFET at 135 GHz was given. This has resulted in PA with 15 dBm saturated output power and a peak PAE of 12.8%, demonstrating that through design 16 nm FinFET could still be used for mm-wave PAs above 100 GHz.

References

1. N.S.M.J.L.E.G.D.M.E.L.Y.L.B.L.M.E.S.L.F.W.A.S.A.C.S.S.L.H.T.N.M.E.D.S.H. Wang, T. Huang, Power amplifiers performance survey 2000-present. https://gems.ece.gatech.edu/PA_survey.html. Accessed: 2021-03-12.
2. B. Razavi, *RF Microelectronics*. Prentice Hall Press, New York (2011)
3. S. Kulkarni, P. Reynaert, A 60-GHz power amplifier with AM-PM distortion cancellation in 40-nm CMOS. *IEEE Trans. Microwave Theory Tech.* **64**(7), 2284–2291 (2016)
4. S. Golara, S. Moloudi, A.A. Abidi, Processes of AM-PM distortion in large-signal single-FET amplifiers. *IEEE Trans. Circuits Syst. I Regul. Pap.* **64**(2), 245–260 (2017)
5. M. Vigilante, P. Reynaert, A wideband class-AB power amplifier with 29–57-GHz AM-PM compensation in 0.9-v 28-nm bulk CMOS. *IEEE J. Solid-State Circuits* **53**(5), 1288–1301 (2018)
6. M. Vigilante, E. McCune, P. Reynaert, To EVM or two EVMs?: An answer to the question. *IEEE Solid-State Circuits Mag.* **9**(3), 36–39 (2017)
7. S.C. Cripps, *RF Power Amplifiers for Wireless Communications*. Artech House, Inc., Norwood (2006)
8. G. Hueber, A.M. Niknejad (eds.), *Millimeter-Wave Circuits for 5G and Radar*. The Cambridge RF and Microwave Engineering Series (Cambridge University, Cambridge, 2019)
9. Y. Zhang, P. Reynaert, A high-efficiency linear power amplifier for 28 GHz mobile communications in 40 nm CMOS, in *2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pp. 33–36 (2017)
10. B. Park, S. Jin, D. Jeong, J. Kim, Y. Cho, K. Moon, B. Kim, Highly linear mm-wave CMOS power amplifier. *IEEE Trans. Microwave Theory Tech.* **64**(12), 4535–4544 (2016)
11. S. Pornpromlikit, J. Jeong, C.D. Presti, A. Scuderi, P.M. Asbeck, A watt-level stacked-FET linear power amplifier in silicon-on-insulator CMOS. *IEEE Trans. Microwave Theory Tech.* **58**(1), 57–64 (2010)
12. B. Hanafi, O. GÄijrbÄijz, H. Dabag, S. Pornpromlikit, G. Rebeiz, P. Asbeck, A CMOS 45 GHz power amplifier with output power >600 mw using spatial power combining, in *2014 IEEE MTT-S International Microwave Symposium (IMS2014)* (2014), pp. 1–3
13. J. Cools, P. Reynaert, A 40nm bulk CMOS line driver for broadband communication, in *ESSCIRC Conference 2016: 42nd European Solid-State Circuits Conference* (2016), pp. 273–276
14. D. Pozar, *Microwave Engineering Fourth Edition* (2005)
15. P. Reynaert, A.M. Niknejad, Power combining techniques for RF and mm-Wave CMOS power amplifiers, in *ESSCIRC 2007 - 33rd European Solid-State Circuits Conference* (2007), pp. 272–275
16. R. P. Dixian Z., *CMOS 60-GHz and E-band Power Amplifiers and Transmitters* (Springer, Berlin, 2015)
17. H.T. Nguyen, D. Jung, H. Wang, 4.9 a 60 GHz CMOS power amplifier with cascaded asymmetric distributed-active-transformer achieving watt-level peak output power with 20.8% PAE and supporting 2Gsymb/s 64-QAM modulation, in *2019 IEEE International Solid- State Circuits Conference—(ISSCC)* (2019), pp. 90–92
18. W. Tai, L.R. Carley, D.S. Ricketts, A 0.7w fully integrated 42 GHz power amplifier with 10% PAE in 0.13um SiGe BiCMOS, in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers* (2013), pp. 142–143
19. S. Callender, S. Pellerano, C. Hull, A 73 GHz pa for 5G phased arrays in 14 nm FinFET CMOS, in *2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)* (2017), pp. 402–405
20. S. Callender, S. Pellerano, C. Hull, An e-band power amplifier with 26.3% PAE and 24-GHz bandwidth in 22-nm FinFET CMOS. *IEEE J. Solid-State Circuits* **54**, 1266–1273 (2019)

21. D. Simic, P. Reynaert, A 14.8 dbm 20.3 db power amplifier for d-band applications in 40 nm CMOS, in *2018 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)* (2018), pp. 232–235
22. R.B. Yishay, D. Elad, A 17.8 dbm 110–130 GHz power amplifier and doubler chain in SiGe BiCMOS technology, in *2015 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)* (2015), pp. 391–394



5.1 Efficiency Enhancement Techniques

The power amplifier is the most power-hungry block in a transmitter. The efficiency of the PA has therefore received a lot of focus, which has resulted in many techniques, each with its pros and cons. Furthermore, the peak to average power ratio (PAPR) of the transmitted signal only increases with higher modulation schemes required for better bandwidth efficiency. To amplify this signal, the PA has to operate in back-off further reducing output power and efficiency. The PAPR can easily be calculated from the constellation diagram following [1]. The PAPR for an increasing number of QAM constellation points is shown in Fig. 5.1. The PAPR increases further by applying a square-root-raised-cosine filter, for a typical roll-off factor of $\alpha = 0.35$, and the PAPR increases with 4 dB [2]. In the following sections, we will very briefly go over the existing (back-off) efficiency enhancement techniques applied to mm-wave PAs: Doherty and outphasing PA. What is required to implement them and identify where mm-wave PAs above 100 GHz have trouble with.

5.1.1 Common Efficiency Enhancement Techniques

5.1.1.1 Doherty Power Amplifier

One of the most well-known back-off efficiency enhancement techniques is the Doherty PA as invented by William Doherty in the 1930s [3]. The Doherty PA employs two PAs: a main or carrier PA and an auxiliary or peaking PA as shown in Fig. 5.2. The main PA is the overall workhorse of the PA, and it is always on amplifying the input signal. The peaking amplifier as its name suggests only

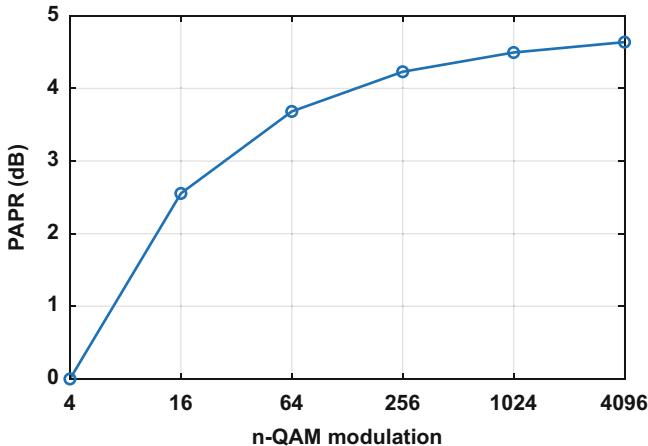
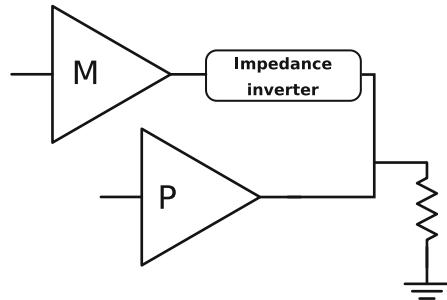


Fig. 5.1 The PAPR for different QAM modulation indices, without RRC filtering

Fig. 5.2 The basic schematic of a Doherty PA



operates at the peaks or when more power is required than the main PA can deliver. The help from an equally sized peaking PA, in the ideal case, can only add 3 dB of extra output power and shows limited performance enhancement. The real beauty of the Doherty PA comes when taking the extra $\frac{\lambda}{4}$ transmission line between the two PAs into account. The peaking PA will perform an active load-pull on the main PA further increasing the possible output power. The load-pull effect can be expressed as follows:

$$Z_{main} = \frac{Z_0^2}{R_L} \left(1 + \frac{I_{peak}}{I_{main}} \right)^{-1} \quad (5.1)$$

$$Z_{peak} = R_L \left(1 + \frac{I_{main}}{I_{peak}} \right). \quad (5.2)$$

Several CMOS mm-wave Doherty PAs have been designed up to 77 GHz [4–9]. Although these PAs show Doherty behavior, as the frequency increases, the gains of the Doherty technique reduce as the ideal requirements for a Doherty PA cannot

be met by the technology. In the most basic design, a Doherty PA requires the three following elements from the amplifier stage to work properly:

- The main PA needs to demonstrate load modulation. As the frequency increases, the difference between peak efficiency and peak power impedance diminishes as shown in Chap. 4.
- The second requirement assumes the peaking PA does not load the main PA at back-off, or $Z_{off} = \infty$.
- Third, the output power of the peaking PA needs to ramp up quickly. Most designs achieve this by biasing the peaking PA toward class-C. In Chap. 4, the effect of decreasing the bias of a PA at 140 GHz is shown, and the simulated class-C biased PAs shows low gain. This problem already appears at lower frequencies and is solved by adding a rectifier, to dynamically change the bias voltage [5], but adds extra complexity and power consumption.

5.1.1.2 Outphasing PA

The outphasing PA or transmitter has gained some attention for mm-wave and terahertz data transmission in the recent years [10–14]. The concept of outphasing relies on the decomposition of an modulated signal S_{out} into two constant envelope signals S_1 and S_2 . The amplitude modulation is now captured in the phase angle, 2φ , between both signals. When both signals are combined, the in-phase or out-of-phase signals will lead to constructive or destructive interference, reconstructing the amplitude modulation at the output. The simplified schematic and its phasor diagram are shown in Fig. 5.3. The outphasing signals can then be described as follows:

$$S_1 = \frac{A}{2} \cos(\omega t + \theta + \varphi) \quad (5.3)$$

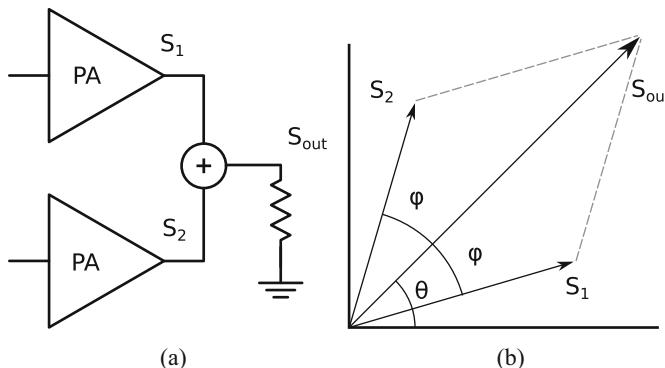


Fig. 5.3 A simplified schematic of: (a) an outphasing PA and (b) its phasor diagram

$$S_2 = \frac{A}{2} \cos(\omega t + \theta - \varphi) \quad (5.4)$$

$$S_{out} = S_1 + S_2 = A \cos(\varphi) \cos(\omega t + \theta). \quad (5.5)$$

The system allows the two power combining PAs to have a constant envelope and does not need linear amplification. All the modulation information is captured by the phase of the signal. This technique is sometimes also referred to as “Linear amplification using nonlinear components” or LINC. It is this technique that allows for that last stage to be a non-linear multiplier in a THz transmitter [13, 14].

Although highly efficient non-linear PAs could be used at back-off, the two PAs still consume a lot of power and RF power is wasted. The PBO efficiency will therefore not be as much as with the Doherty PA. However, by choosing the right non-isolating output combiner network, load modulation can be achieved to further improve the PBO efficiency as in a Doherty PA [15].

The efficiency enhancement techniques of an outphasing PA thus rely on the following two elements to work:

- The two PAs need to show a high efficiency by operating as a hard-switching non-linear amplifier. The high base efficiency comes from a high harmonic content together with a high-quality output impedance. For a 5 GHz PA, the 5th harmonic is still only at 25 GHz, where the transistors still demonstrate gain. For a 100 GHz PA, the second and third harmonic frequencies are too high for the transistor to still show gain. A hard-switching efficient PA is therefore not possible to create. Furthermore, higher frequency PAs have higher parasitic output losses further reducing the maximal achievable efficiency.
- The power-back-off (PBO) efficiency can be further enhanced by making use of load modulation on both PAs. With the increasing frequency, the impedance difference between peak efficiency and peak power diminishes resulting in limited possible load modulation.

5.1.1.3 Conclusion: Difficulties for >100 GHz Power Amplifiers

Most mm-wave PBO-efficiency improvement techniques rely on load modulation to move the PA, between the optimal load for efficiency at back-off and power at the peak output power. Unfortunately, as the frequency increases, the difference in impedance between peak efficiency and peak power reduces, and almost no improvement can be gained by load modulation.

Due to the reduced intrinsic gain at higher frequencies, it is difficult to trade gain for other improvements as a reduction in gain requires more power from the driver and will not result in the desired efficiency improvement.

The parasitic drain resistance reduces the output impedance when a PA is turned off. A turned-off PA will always provide a power-consuming impedance to other PAs. Non-isolated power combined PAs cannot be selectively turned off to reduce the output power as in [8, 16].

The combination of these properties leads to a simple PBO-efficiency improved amplifier such as the sequential power amplifier [17]. The amplifier does not use load modulation and uses an isolating power combiner. The PA will be discussed in detail in the next section.

5.1.2 A Sequential Power Amplifier

A sequential power amplifier (SPA), as described by Cripps [17], consists of a main PA and a peaking PA combined through a directional coupler as shown in Fig. 5.4. The main PA is always on and provides the power amplification over the whole input range, and the peaking PA only turns on for the large output powers and assists the main PA. The increased output power of the peaking PA extends the gain when the main PA goes into compression. The design of this PA is centered around the output coupler. The coupling factor will determine the characteristic of the PA at back-off. It will also provide isolation between the main and peaking PA. This isolation means no load modulation is happening but also the main PA is not loaded by the peaking PA in the off-state, keeping the efficiency high at power back-off [18].

Figure 5.5 shows the schematic of a 140 GHz sequential PA. Both input and output consist of a quadrature coupler, as both paths from input to output need to

Fig. 5.4 The basic schematic of a sequential PA [17]

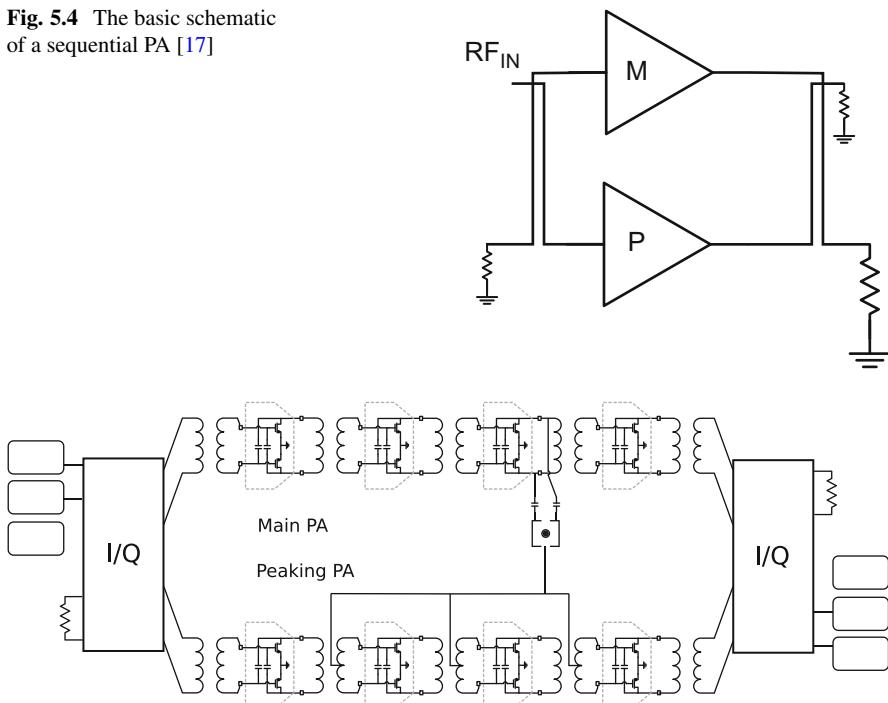


Fig. 5.5 The schematic of a 4-stage sequential PA with a rectifier enabled peaking PA

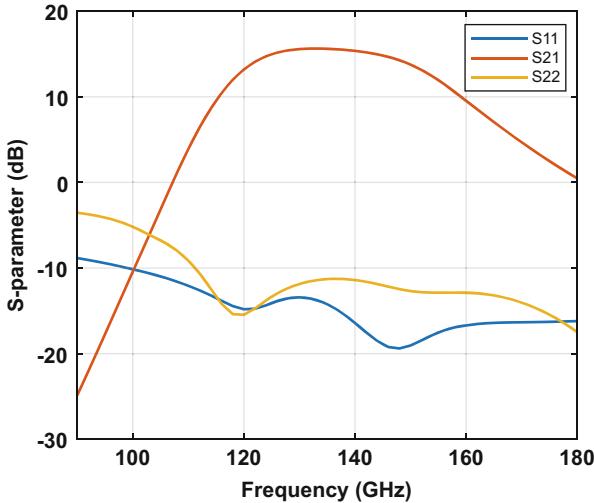


Fig. 5.6 The simulated S-parameter characteristics of the sequential power amplifier

see the same phase shift to combine constructively. The input therefore also benefits from excellent input matching, as with a balanced PA, the reflections get absorbed by the dummy load. The PA uses 4 stages to provide enough small-signal gain. For small input powers, the peaking PA is off, and input power is still split equally between both branches losing 3 dB. The peaking PA normally operates in class-C to get the required power ramp-up. However, in Chap. 4, it is shown that a PA at 140 GHz does not operate well enough when biased in Class-C operation. The PA shows reduced gain and saturated output power compared to class-AB PAs. A solution to this problem is to apply an adjustable bias voltage. This dynamic bias voltage created by a rectifier tracks the output power of the main PA driver [5, 19–21]. The dynamic bias voltage is applied to the pre-driver, driver, and output stage of the peaking PA.

The simulated S-parameters of the sequential PA are shown in Fig. 5.6. It shows a peak gain of 16 dB at 130 GHz. Both input and output matches are good over the whole bandwidth due to the coupler at input and output.

Figure 5.7 shows the large signal characteristic of the sequential PA. The simulated PAE shows a clear improvement compared to an ideal reference class-A PA demonstrating the possible back-off improvement. Furthermore, the gain compression is also improved, bringing OP1dB closer to the saturated output power. One of the problems seen from this simulation is the low-peak PAE; this could be attributed to extra DC-power consumption of the extra driver stage and rectifier. However, the main problem is the insertion loss of the output matching network.

The layout of the output matching network is shown in Fig. 5.9a, and a transformer matching is performed between the amplifier stage and the coupler. An extra LC matching is necessary between the output pads and the couplers

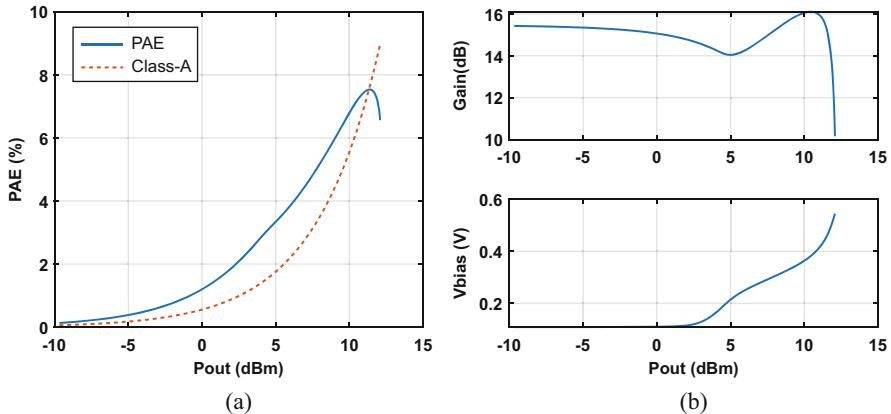


Fig. 5.7 The simulated PAE with equivalent class-A PA (a) and large signal gain and peaking bias voltage (b)

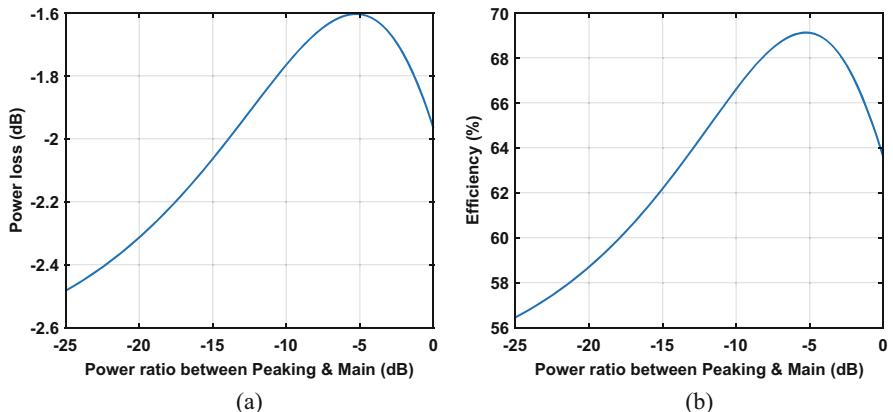


Fig. 5.8 The (a) insertion loss and (b) efficiency of the output coupler over the power ratio

output. The complete output matching network shows an S31 from Main PA to output of -2 dB and S32 from the Peaking PA to the output of -6.6 dB . The insertion loss and efficiency of the coupler is shown in Fig. 5.8. At peak power, when both PAs are on, an IL of 2 dB is shown with an output efficiency of 64% . The peak efficiency of 69% is reached for a power ratio of -5.5 dB . For comparison, the output matching network for parallel power combing as shown in the design example Chap. 4 and shown in Fig. 5.9b has an insertion loss of only 1.2 dB at peak power. The insertion loss of the output matching network has a detrimental effect on the overall performance of the PA. This increased loss will reduce the output power and reduces the PAE, in the most obvious examples, a 3 dB loss will half the output power and half the PAE, and a theoretical 100% efficient PA suddenly only achieves 50% . The effect of IL on PAE is shown in Fig. 5.10b.

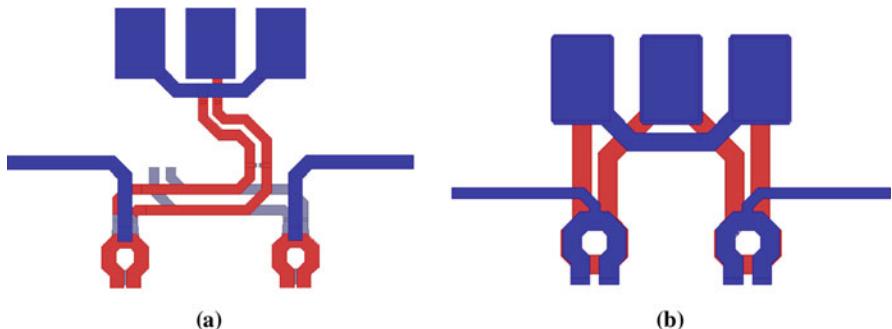


Fig. 5.9 Coupler-based output matching (a) and zero length TL power combining (b)

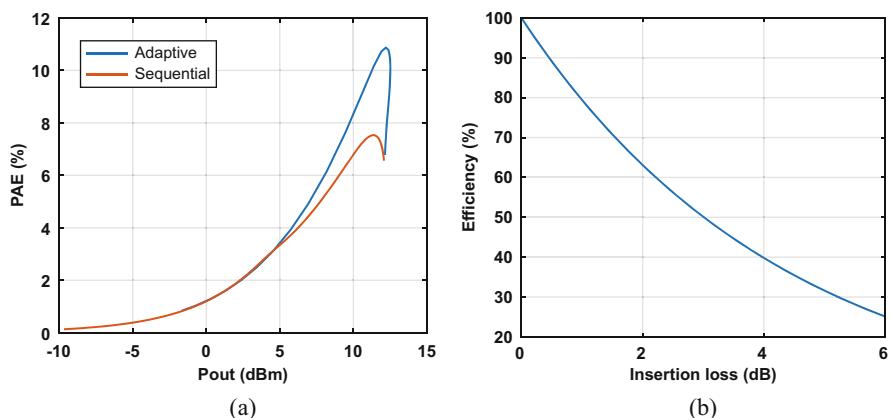


Fig. 5.10 The PAE of a sequential PA compared to simple PA (a) and the efficiency penalty to insertion loss (b)

The parallel power combiner shows a lower insertion loss, which improves the overall PA's performance. However, it does not provide isolation between the main and peaking PAs; therefore, the network will degrade the performance at back-off by load modulation and undesired loading. The PA is adapted to use a parallel power combiner by exchanging the input and output couplers for a parallel splitter and combiner. Figure 5.10(a) shows the simulated PAE of the sequential and this parallel power combining adaptation. It is immediately clear that the parallel combiner outperforms the sequential PA. Although the peak to back-off efficiency improvement is less compared to the sequential PA, the reduced IL improves the overall PAE such that the PAE even at back-off never goes below the sequential PA.

5.1.3 Conclusion

From the investigation into efficiency enhancement techniques such as the sequential PA, we can conclude that efficiency enhancements are possible, but one cannot lose sight of the added power consumption or increased output IL created. It is for this reason a simple power combining PA is chosen as the final design that trades off back-off efficiency enhancement for the overall efficiency.

5.2 Linearization Techniques by Dynamic Bias

Next to the efficiency of the PA, linearity is also important. The effect of amplitude and phase distortion on the constellation diagram was discussed in Chap. 4 and is captured in the EVM. In this section, some techniques to improve the linearity of the PA with the help of a peaking PA providing variable power are discussed.

5.2.1 AM-AM Compensation

Amplitude distortion shows itself on the constellation diagram by the outer constellation points moving inward and results in an increased EVM. The inward motion of the outer constellation points is automatically compensated to an extent by the receiver. The receiver will try to map the given constellation points to an ideal constellation with the lowest EVM. Instead of using the ideal constellation diagram of the distorted diagram, a scaled version will result in a reduced EVM. This effect is demonstrated in Fig. 5.11. This works because scaling the ideal constellation diagram dramatically improves the EVM of the outer constellation

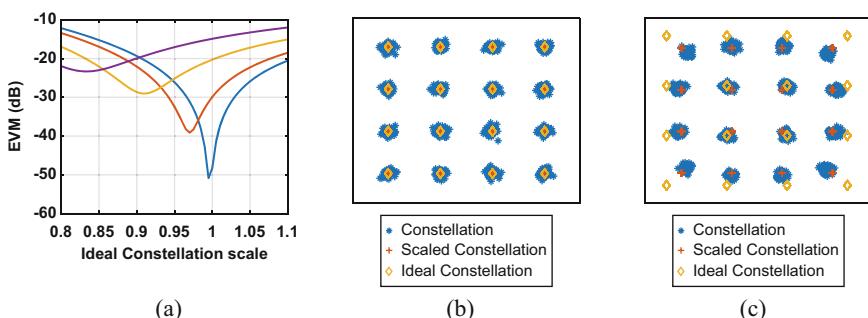


Fig. 5.11 The simulated (a) EVM against the scaling factor of the ideal constellation for different back-off powers and the constellation diagrams for (b) complete back-off and (c) a constellation in compression with a scaling factor of 0.83

points with only a minor degradation for the inner points. Although the receiver will already compensate some of the amplitude distortion, improving the linearity is still beneficial for every PA especially since the large targeted bandwidth of mm-wave PAs is costly for digital pre-distortion (DPD).

The combination of main and peaking PAs allows for some AM–AM compensation at the cost of a decreased small-signal gain. At low input powers, the main PA is the only amplifier providing gain; however, the input power is split between the main and peaking PAs reducing the gain by 3 dB for an equal split. When the main PA starts to compress, the peaking PA is turned on, power is added, and the gain is thus extended for 3 dB before also the peaking PA starts to compress.

5.2.2 ***AM–PM Compensation***

The origins of AM–PM distortion have been the highlight of several works [22–25], and a multitude of compensation techniques have been developed [23, 25–27].

Two techniques try to compensate for the non-linear variation of the input capacitance of the NMOS transistor. The first technique employs a complementary PMOS amplifier [23]. The NMOS and PMOS devices can be combined on top of each other or two separate amplifiers can be power combined. Although a lot of work has been done to improve the RF performance of the PMOS device, it is still not as good as an NMOS transistor. PMOS devices have been used for mm-wave application up to 76 GHz [28, 29], but they are still not good enough to use for PAs above 100 GHz. A second technique to compensate for the input capacitance is by employing a PMOS transistor or PMOS varactor to create an opposite capacitor. In contrast to the previous technique, the PMOS device does not do any active power generation or amplification, and it only acts as a non-linear capacitor [25]. The control voltage is used to ensure correct compensation after process, voltage, and temperature variation (PVT). Even at the presented frequency of 28 GHz the varactor shows low Q and results in a loss of gain. The performance of varactors above 100 GHz is bad, and again PMOS varactors are worse compared to their NMOS counterparts. It is for these reasons PMOS varactors are not the ideal technique for mm-wave PAs above 100 GHz.

Employing source degeneration on CS amplifier to improve linearity is another widely used technique. This technique improves both AM–AM and AM–PM by adding a source inductor [26, 30]. A well-known downside of using source degeneration is a gain reduction. This reduction is not a problem for the mm-wave PAs at 28 GHz since the transistors can still deliver a lot of gain. Unfortunately, the gain of the transistors operating at 100 GHz and above is already low, and adding source degeneration would reduce this even further.

Although a lot of research has been done on AM–PM compensation for mm-wave power amplifiers for 5G and E-band applications, these techniques are not as easily applied to mm-wave PAs above 100 GHz. For most published PAs this is not an issue as mm-wave PAs above 100 GHz, such as the example in Chap. 4,

are biased in class-A to maximize output power and gain. The class-A PA comes automatically with good AM-PM performance. However in this case, we try to improve the PA, and thus techniques to improve the AM-PM need to be considered. The peaking PA with its changing bias is prone to large AM-PM distortion. It can however also be leverage for AM-PM compensation, and two different approaches are discussed below.

5.2.2.1 Capacitive Compensation

The peaking PA is dynamically biased causing the input capacitance of the controlled amplifiers to increase as the bias is increased. This will result in an uncontrolled phase expansion of the peaking PA. To control the phase expansion and dissociate it from the gain control, a varactor as in [25] can be added. PMOS varactors show poor behavior at mm-wave frequencies and are not always supported by the PDK. To get a similar behavior as a PMOS varactor, an NMOS varactor can be used as the gate to source and drain voltage is reversed. This can be achieved by switching the gate to source and drain connection. The downside of this approach is the increased fixed capacitance of the well to substrate capacitance [22]. This added capacitance can be absorbed by the matching network as long as the Q-factor of the input does not increase [25]. The simulated input capacitance and quality factor of a neutralized CS amplifier with NMOS compensation is shown in Fig. 5.12. The gray area shows the point of operation by the dynamic bias circuit; we cannot just compensate the PA but also overcompensate. The overcompensation allows the varactor of one stage to compensate multiple stages at once. The gain penalty due to the losses of the varactor is now limited to only one stage. Furthermore, since this is

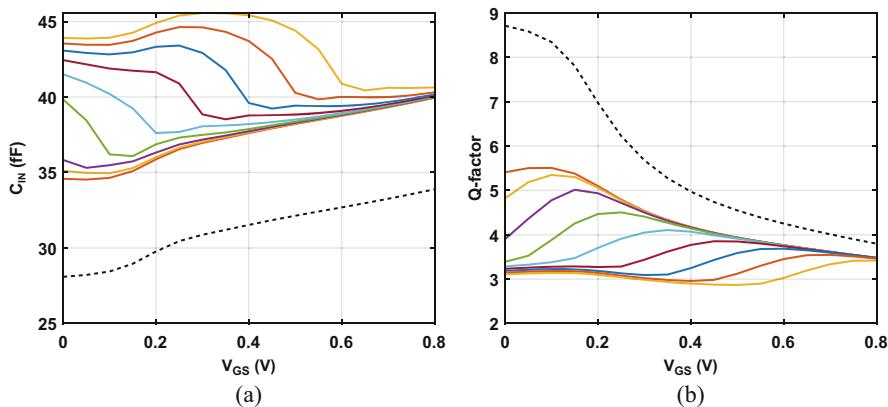


Fig. 5.12 The effect of adding a varactor to (a) the input capacitance and (b) the quality factor. The dashed blacked line indicates the reference without varactor

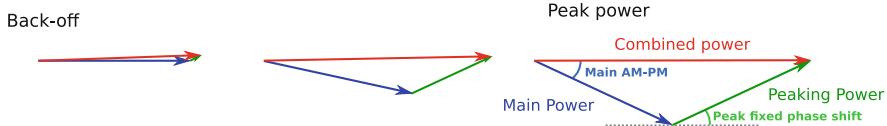


Fig. 5.13 Simplified phasor diagram visualizing the AM-PM compensation using a fixed phase offset with a variable peaking power branch

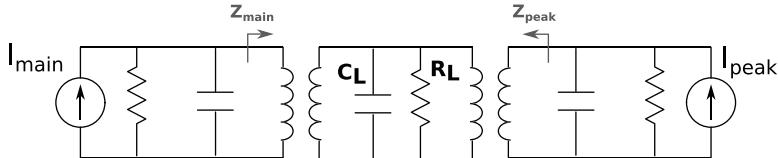


Fig. 5.14 Parallel output combining network

a pre-driver stage, it has limited effect on the overall PAE determined by the output and driver stage.

5.2.2.2 Fixed Phase Offset

The rapid increase in power from the peaking PA allows for another phase compensating technique employing only a fixed phase offset. Figure 5.13 provides a simplified phasor diagram of the PA's output power decomposed into the main and peaking PAs. For perfect power combining, both main and peaking PAs should be in phase; in contrast here, the PAs are designed with a small phase offset. At low output powers, at back-off, the main PA is the only one providing power. The output phase is only determined by the main PA. When the power increases, the main PA goes into compression, and it shows an undesired phase distortion. At the same time, the peaking PA starts to provide power, and, because of the phase shift, can now compensate for the change in phase from the main PA. The advantage of this technique is that no amplitude-varying components are necessary just a fixed phase shift between main and peaking is required. The disadvantage is that by not power combining in phase a small amount of power is lost.

If the power combining network is simplified as in Fig. 5.14, the output power and phase can be derived, with $I_{main} = I_0$ and $I_{peak} = \alpha I_0$, for which alpha represents how much the peaking PA is turned on. The phase of the main PA is ϕ_1 , and the phase difference between the two PAs is $\Delta\phi$. The transformers are assumed ideal one-to-one transformers that resonate out the capacitors. The total combined current's magnitude and phase through the load are than given as follows:

$$|I_{tot}| = I_0 \sqrt{\alpha^2 + 2\alpha \cos(\Delta\phi) + 1} \quad (5.6)$$

$$\angle I_{tot} = \phi_1 - \arctan \left(\frac{\alpha \sin(\Delta\phi)}{1 + \alpha \cos(\Delta\phi)} \right) \quad (5.7)$$

$$\approx \phi_1 - \arctan \left(\frac{\alpha \Delta\phi}{1 + \alpha - 0.5\alpha \Delta\phi^2} \right). \quad (5.8)$$

To get the output voltage and power, the output impedance is necessary. The output impedance of the peaking PA is dependent on the on-state and is taken as $R_{peak} = R_0/\alpha$. With this and the above equations, the output voltage and delivered power can be found as follows. Furthermore, if a conjugate match at peak power is assumed with $R_0 = 2R_L$, the output power can be further simplified.

$$|V_{out}| = \frac{R_L R_0}{R_0 + (1 + \alpha) R_L} I_0 \sqrt{\alpha^2 + 2\alpha \cos(\Delta\phi) + 1} \quad (5.9)$$

$$P_{out} = \frac{4I_0^2 R_L (\alpha^2 + 2\alpha \cos(\Delta\phi) + 1)}{(3 + \alpha)^2} \quad (5.10)$$

$$Z_{main} = R_L \sqrt{\alpha^2 + 2\alpha \cos(\Delta\phi) + 1} \quad (5.11)$$

$$Z_{peak} = R_L \frac{\sqrt{\alpha^2 + 2\alpha \cos(\Delta\phi) + 1}}{\alpha}. \quad (5.12)$$

Figure 5.15 shows the output power and phase, and impedance over the turn-on range of alpha. At back-off, some power is lost due to a mismatch between the main PA and the load. This however does not take into account load modulation, which will be limited due to the high frequency and a lower off-impedance of the peaking PA. Despite this, it is clear that a small fixed offset angle between the two will barely impact the output power but allows creating an AM–PM characteristic however we like.

All the techniques discussed will be used and implemented in the following design example.

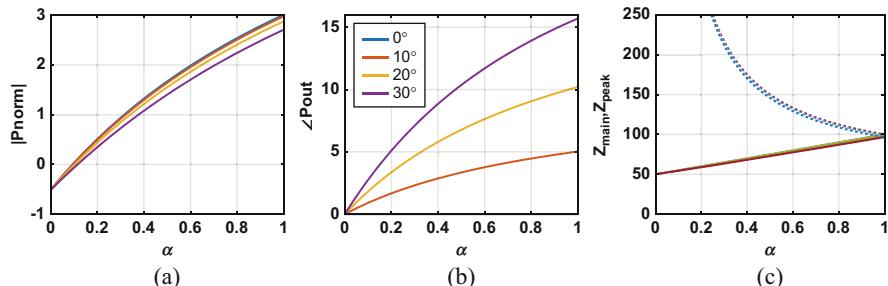


Fig. 5.15 The simulated normalized output power (a), the output phase (b), and output impedance (c) over the on-state alpha of the peaking PA

5.3 Design Example: A Direct-Conversion TX with Enhanced PA

The D-band transmitter is implemented in a 16 nm FinFET CMOS technology, with RF-enhanced transistors. The design of the transmitter will be discussed in detail and split into two sections, the PA and the modulator. The PA will employ main and peaking PAs to enhance its performance compared to the class-A PA example from Chap. 4. The modulator consists of an I/Q mixer with on-chip quadrature generation and a frequency tripler. The design is fabricated, and measurements are discussed below.

5.3.1 The Dynamic Bias Enhanced PA

The designed PA consists of main and peaking PAs, power combined using a parallel power combiner for minimal IL. Figure 5.16 shows the schematic of PA. The main and peaking PA branch consists of 4 CS-capacitive neutralized amplifier stages. The PA was designed with 4 stages instead of 3 in the previous chapter to compensate the 3 dB loss in gain due to the input splitter.

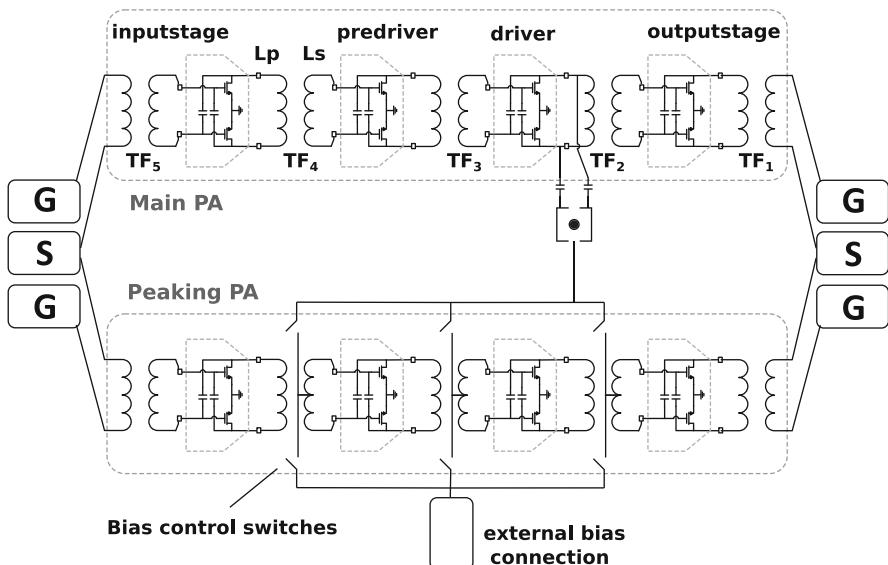


Fig. 5.16 The schematic of PA

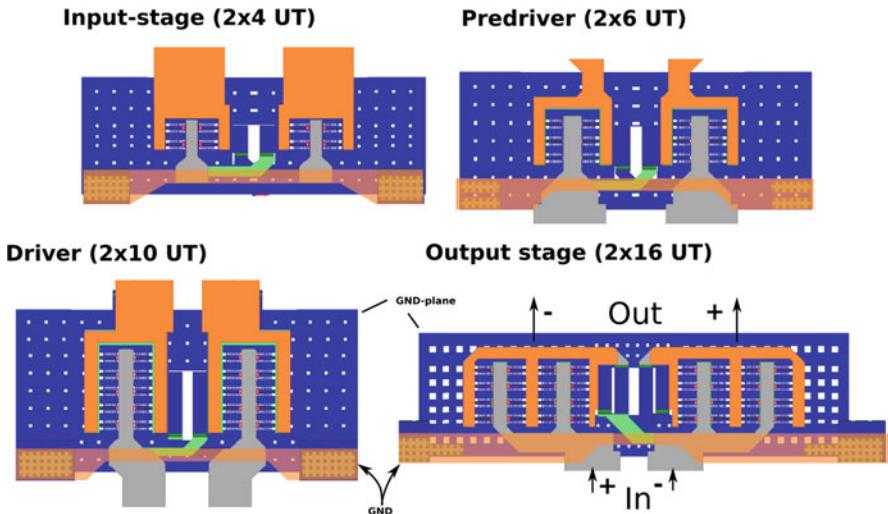


Fig. 5.17 The layout of the 4 amplifier stages: input, pre-driver, and output stage

5.3.1.1 Amplifier Stages and Interstage Matching

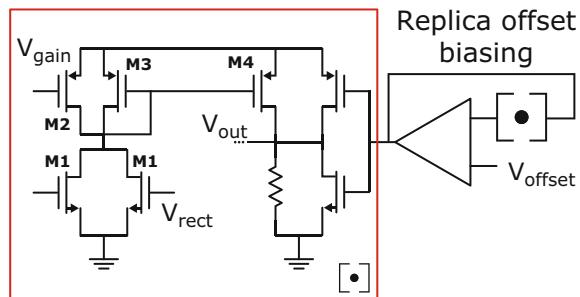
The design example in Chap. 4 describes how to build up the amplifier stages from smaller unit cell transistors. Although a new RF-enhanced 16 nm FinFET transistor is used in this design, the amplifier still benefits from small unit cell transistors. The unit transistor is sized at 15 fingers with 6 fins per finger, and the layout goes up to the first non-x-metal layer using the layout as in Chap. 4. All the amplifier stages are differential capacitive neutralized common-source amplifiers. The layout of all the stages is shown in Fig. 5.17, and the output stage uses 2x16 unit transistors (UTs) with two neutralization capacitors of 42.5 fF. The driver, pre-driver, and input stage are scaled down to 2 × 10 UTs, 2 × 6 UTs, and 2 × 4 UTs to improve the PAE. The ground plane around the unit transistors is made up of 3 x-metal layers and connects on both sides of the amplifier to the top metal layer to provide a low resistive ground.

The output and interstage matching is designed with transformers, and their respective values are shown in Table 5.1. The input split is done with a parallel power combiner as this provides a better impedance transformation compared to series splitting.

Table 5.1 Matching transformer parameters

	TF_1	TF_2	TF_3	TF_4	TF_5
L_p (pH)	21	27.6	45	51	61.4
L_s (pH)	26	16	21.5	28	45.8
k_m	0.73	0.48	0.33	0.5	0.57

Fig. 5.18 The schematic of the rectifier connected at the output of the driver stage



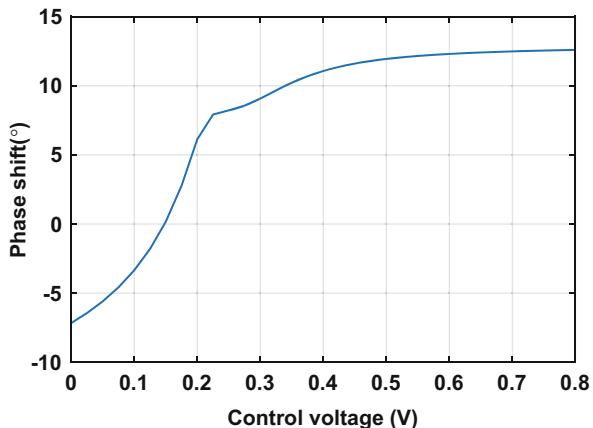
5.3.1.2 The Dynamic Bias Circuit

The peaking PA needs to be dynamically turned on with the increasing input power. To achieve this, the bias voltage of the peaking PA is controlled by a rectifier circuit that follows the input power. Figure 5.18 shows the schematic of the rectifier, and the input signal is coupled through two capacitors of 52.2 fF to the input gates of transistors M1. Biasing is applied through two 15.4 kΩ resistors. The input transistors are sized at 6 fins and 15 fingers. The conversion gain of the rectifiers is controlled by their gate voltage (V_{rect}), and the DC current is controlled by transistor M2 (V_{gain}) sized at 6 fins and 20 fingers. The current is then mirrored and amplified with the CS amplifier M4 sized at 2 transistors of 9 fins and 15 fingers and a 100 Ω resistor. This extra amplifier stage is necessary to improve the drive strength to drive the large gate capacitance of the peaking PA. The output of the rectifier is connected to the pre-driver, driver, and output stages with a transmission line. The bias voltage is connected to the amplifiers through the center tap of the coupled gate inductor. The peaking PA can also be controlled with an external control bias voltage. The two bias voltages are separated by two switches. Because the bias voltage does not need to go higher than 0.4 V, an NMOS only switch can be used. The NMOS is sized at 2 transistors of 14 fins and 32 fingers.

The external bias override allows the PA to operate in two extra operation modes: low power and full power. Low power mode sets the bias voltage of the peaking PA to zero. The PA will now only operate with the main PA, and this is the operation for the dynamic PA in back-off. The full power mode sets the peaking bias voltage to the same as the main PA. The PA now operates as a two-way power combining PA similar as in Chap. 4. There is no power-back-off efficiency enhancement and no linearization happening now.

The input of the dynamic bias circuit is connected at the output of the main PAs driver stage. The rectifier's conversion gain is dependent on the input power level.

Fig. 5.19 The simulated phase shift between the main and peaking PAs with the varactor's control voltage



5.3.1.3 AM–PM Control

The fixed phase shift between the main and peaking PAs is set by a varactor added at the gates of the input stage. The varactor allows tuning the resonant input match between the two branches. This results in a change in phase between the two PA branches. Figure 5.19 shows the adjustable phase shift between the main and peaking PAs.

Another varactor is added at the gates of the pre-driver. This varactor is here to compensate for the variation in C_{GS} due to the large change in bias voltage. The effect on the capacitance and Q-factor is previously shown in Fig. 5.12. The reduction in Q-factor will lead to a reduction in the gain between the input stage and the pre-driver. Because of this penalty in gain and power transfer, the varactor is only placed in front of the pre-driver and overcompensates the phase to compensate for the following driver and output stage.

5.3.1.4 PA Simulation Results

The PA is fully simulated, all EM interconnects are simulated together to capture their mutual coupling [31], and the parasitics of the transistors are captured in their parasitic extraction. The small-signal S-parameter simulations for the 3 operation modes are shown in Fig. 5.20. For small signals, the peaking PA is not turned on in both low power and dynamic modes; therefore, they show the same S-parameters. The S11 is below -10 dB over the band of operation, and a good input match for the low power mode is provided because the input stage is always on. The large signal gain and PAE of the 3 modes are shown in Fig. 5.21. The dynamic mode bridges the gap between the low and full power modes and extends the OP1dB, therefore

Fig. 5.20 The simulated S-parameters for the 3 operation modes

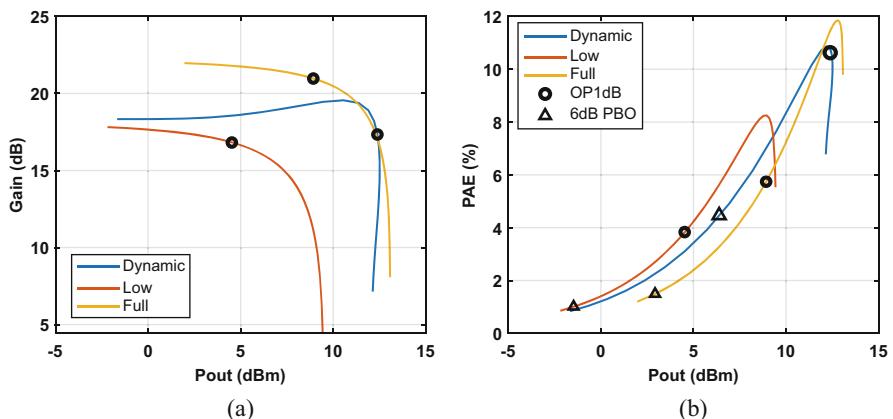
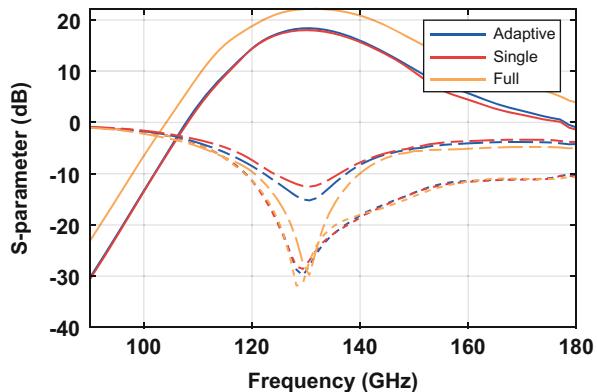


Fig. 5.21 The simulated (a) large signal gain and (b) PAE of the 3 PA modes

reducing the effect of soft compression seen in both low and full power modes. The dynamic mode also shows a PAE between both modes and therefore improves the PAE at back-off. Furthermore, because of the improved linearity, the OP1dB and the 6 dB back-off points are further at a higher power level, which further improves the relative PAE.

5.3.2 The I/Q Modulator and LO Generation

The power amplifier consumes most of the power budget and is a big part of the transmitter, but without a modulator to upconvert the IF-data signals and a good quadrature LO signal, no data can be transmitted. To reduce the complexity of the LO generation, an external LO at 1/3 the frequency is brought on-chip and

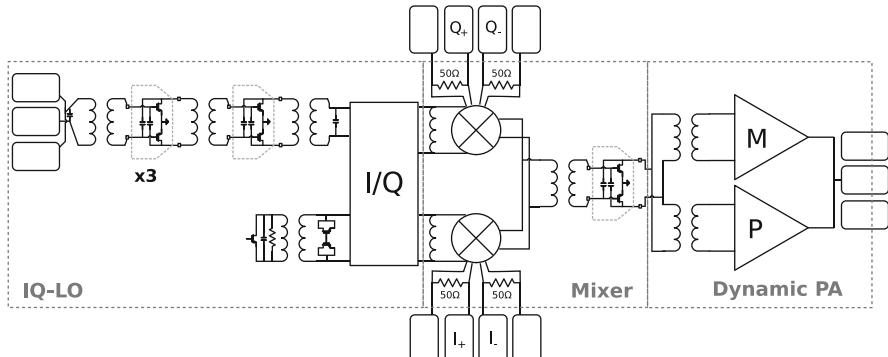


Fig. 5.22 The schematic of the transmitter

multiplied with a tripler to create the required carrier. The multiplied carrier is used to generate the required quadrature LO signal. The quadrature LO signal is fed to two balanced mixers to upconvert the data, which is fed to the PA. Figure 5.22 shows the schematic of the entire transmitter.

5.3.2.1 Quadrature LO Circuit

The quadrature LO is generated from an LO signal at 1/3 the operating frequency brought on-chip through a GSG pad. The signal is directly upconverted by a tripler designed as a differential common-source amplifier consisting of 2×10 UTs. Two neutralization capacitors of 26.4 fF are added to improve gain and stability. The upconverted signal is first buffered by an equally sized amplifier before being fed to a differential coupler for quadrature generation. The layout of the LO circuit is shown in Fig. 5.23. Due to the decreased gain at the fundamental frequency compared to the provided 1/3 LO frequency, sufficient rejection of the lower harmonic is necessary. For this reason, the matching transformer between the buffer and coupler is designed with a low coupling factor. Compared to high-k matching, the rejection is increased by 5 dB as shown in Fig. 5.24. The two quadrature signals are created with a differential coupler [32]. The two coupled lines have a cross-over in the middle to balance the I and Q paths, and they see the same amount of vias and end in the same metal layer improving the symmetry and thus the image rejection ratio (IMRR). To compensate for possible fabrication tolerances and further improve the IMRR, an adjustable dummy load is added at the isolation side. The dummy load consists of a varactor, transformer, and NMOS. Tuning the varactor mainly affects the phase of the coupler, while the control of the NMOS changes both phase and magnitude. Figure 5.25 shows the effect and range of the dummy load. In theory, a perfect quadrature signal can be created. The total harmonic rejections after the coupler are 35 dBc and 42 dBc for the I and Q paths, respectively. The Q-LO enjoys

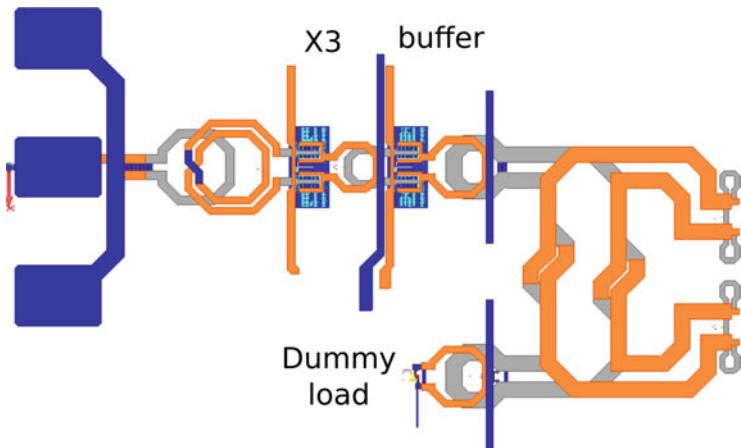


Fig. 5.23 The layout of the quadrature LO circuit

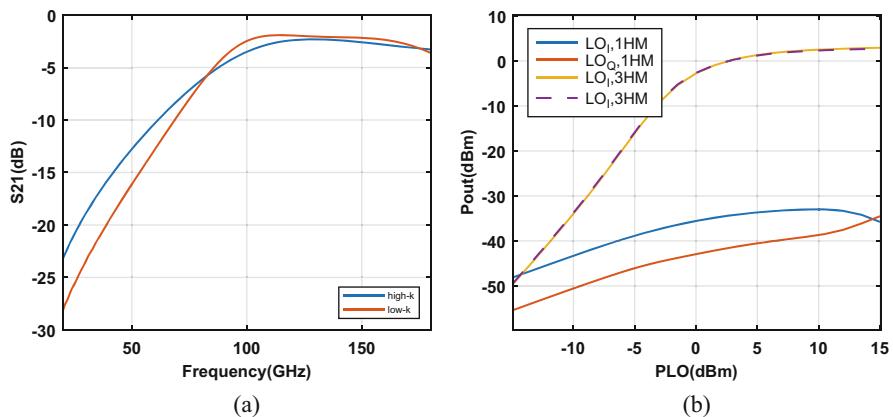


Fig. 5.24 (a) A comparison of harmonic rejection between high-k and low-k matching and (b) the harmonic rejection of the quadrature signals

a higher rejection compared to I-LO because of the coupler. At a lower frequency, the through and coupled paths are not balanced as less coupling is happening.

5.3.2.2 The Double Balanced IQ-Mixer

The IQ-mixer consists of two Gilbert cell mixers, with the LO mixing quad consisting of 4×3 UTs and 2 current steering IF BB transistors of 3 UTs. The differential BB-input has two 50Ω resistors to provide the input match and ESD for protection. The schematic and layout of the full IQ-mixer are shown in Fig. 5.26, a symmetrical layout minimizes LO-feedthrough and gain balance. A

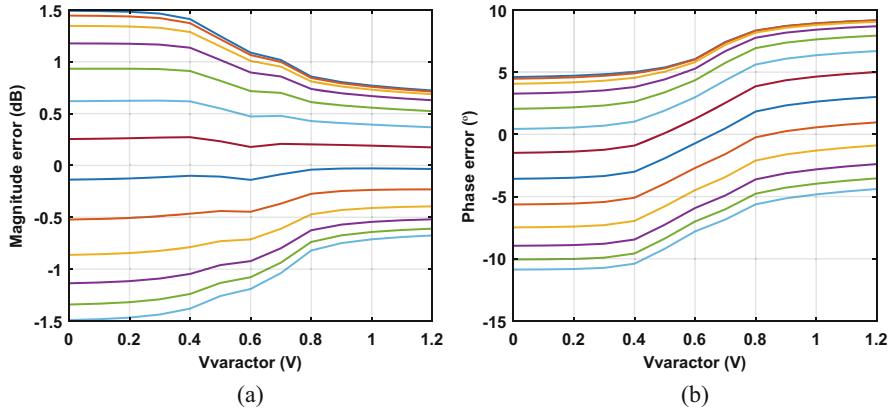


Fig. 5.25 The simulated quadrature, (a) magnitude, and (b) phase error for the 2 control voltages

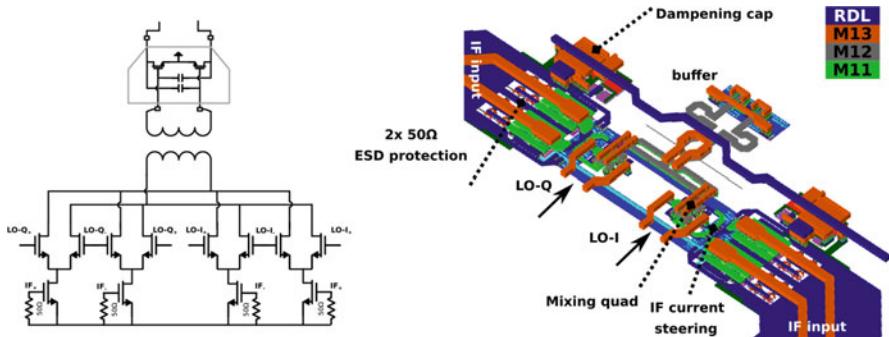


Fig. 5.26 The schematic and layout of the mixer

damped capacitor is placed symmetrically on the supply lines on both sides of the output transformer. This equalizes the impedances seen into both legs of the output transformer by the mixer [33]. This improves both LO-feedthrough and stability. The modulated output signal is then buffered by another capacitive neutralized common-source amplifier with 2×4 UTs. The buffer is matched to the PA with a parallel power splitter.

The simulated IF-bandwidth and power sweep of the full TX are shown in Fig. 5.27. The simulated transmitter shows a bandwidth of 20 GHz, with a good tunable IMRR and low LO-feedthrough. Furthermore, the dynamic PA improves the PAE and linearity.

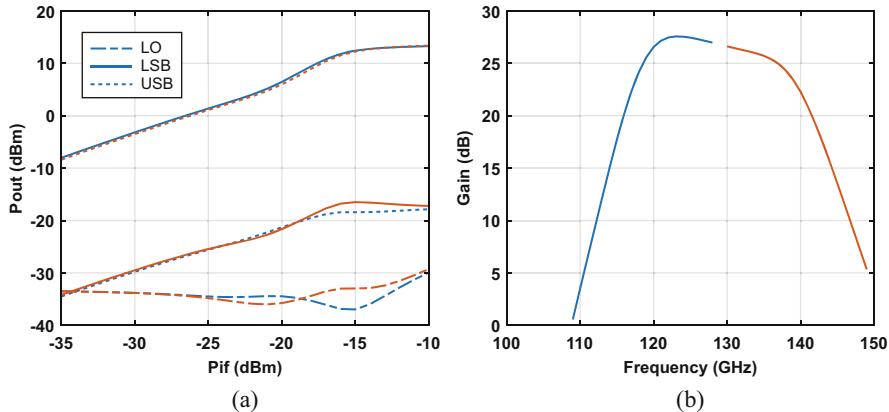


Fig. 5.27 The simulated (a) 1 GHz IF Pin–Pout characteristic and (b) the TX conversion gain for a swept IF frequency, in LSB (blue) and USB (red) mode

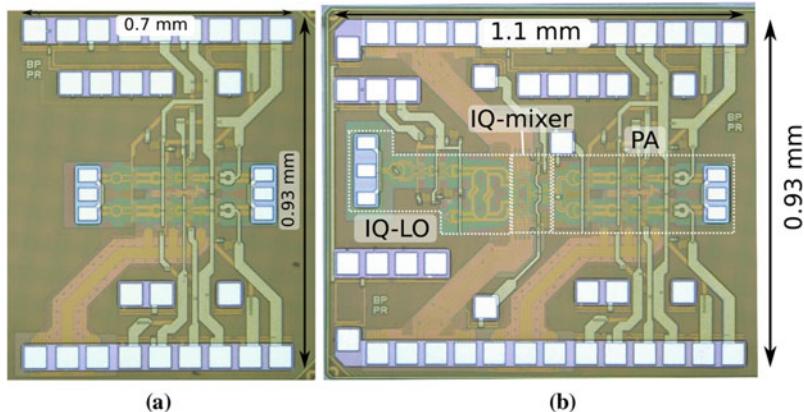


Fig. 5.28 The die micrograph of the fabricated (a) PA and (b) transmitter in 16 nm FinFET process

5.4 Measurements and Discussion

For verification, both the PA with RF input and the full transmitter are taped separately. Both dies are bonded to their respective PCBs and are probed for measurement.

The fabricated 16 nm FinFET PA and transmitter are shown in Fig. 5.28. The PA and transmitter measure $1.1 \text{ mm} \times 0.7 \text{ mm}$ and $1.1 \text{ mm} \times 0.93 \text{ mm}$, respectively, including all the pads.

5.4.1 Power Amplifier Measurements

The PA is gold-bonded to a carrier PCB for supply and bias connections. The RF input and output are probed for characterization. The R&S ZNA is used to do both small-signal and large signal measurements. For small-signal measurements, the setup is calibrated to use WINCAL software to perform a through-reflect-reflect-match (TRRM) calibration up to the probe tips using a calibration substrate as shown in Fig. 5.29. Calibration of the large signal measurement setup is a bit more elaborate. The setup and calibration steps are shown in Fig. 5.30. First, the VNA is calibrated, using the waveguide calibration kit, up to the end of the S-bends. In

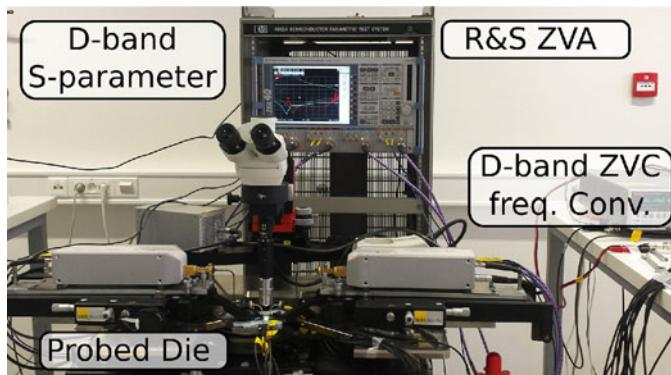
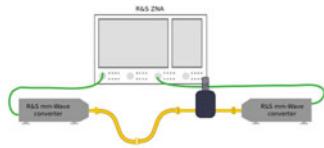
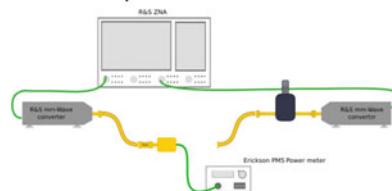


Fig. 5.29 Small-signal measurement setup with D-band probes, extender, and VNA

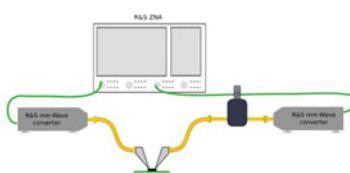
1. ZNA calibration



2. ZNA power calibration



3. Probe calibration



4. Chip measurement

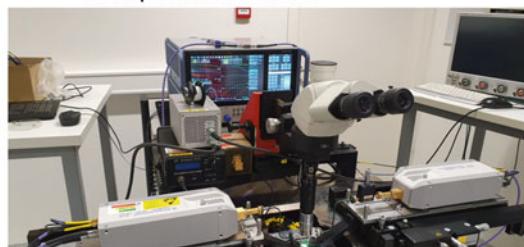


Fig. 5.30 Large signal measurement setup and calibration routine

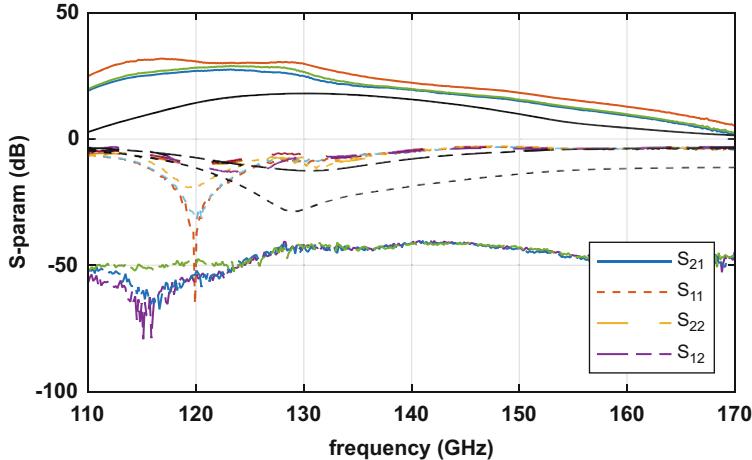


Fig. 5.31 The measured S-parameters of PA for the three modes next to the simulation

the next step, the VNA is power calibrated with the PM5b power meter, using the R&S leveling tool. Because of the waveguide connection of the power meter, the reference plane has to be put at the end of the S-bend. In the next step, the probes are attached and the through of the calibration substrate is measured to get the loss of the probes. Both probes are measured simultaneously and are assumed to have equal performance. Because the PA can deliver more power than the extender itself, an attenuator is added to ensure the extender does not add any compressive behavior as this is not taken into account by the leveling tool.

5.4.1.1 Small-Signal Measurement

The measured S-parameters of the PA for the three modes next to the simulation are shown in Fig. 5.31. Measurement and simulation show a large difference, and the gain difference is about 16 dB. This large difference is probably related to the special double-poly pitch transistors used in this design. The benefit of these models is most likely underestimated at higher frequencies together with a small frequency shift resulting in a larger gain. This is combined with a 4-stage amplifier exaggerating the effect. The transistors still exhibit a gain boost of almost 4 dB per stage.

The gain of the PA can be easily lowered by reducing the bias voltage and operating the PA more toward class-B. For operation in the full or low power mode, the gain does not need to be lowered but is necessary for operating the dynamic mode. Although the gain is increased, the power characteristics are not, and a larger gain output stage will therefore mean a lower power is necessary at its input to drive the stage into compression. This lower power will unfortunately also mean a

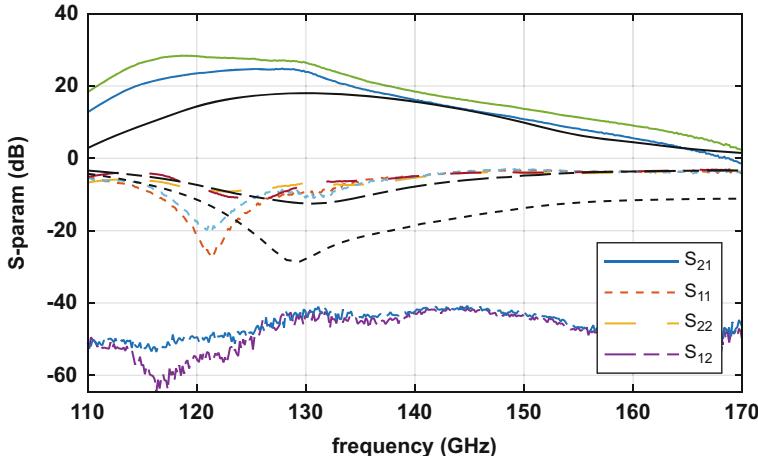


Fig. 5.32 The measured S-parameters of PA for the three modes with a reduced bias voltage of 350 mV

lower input power is provided to the rectifier, and the rectifier cannot provide enough conversion gain to turn the peaking PA on fast enough.

For the following large signal measurements, the PA is thus operated from a 350 mV bias voltage. The measured S-parameters for the reduced bias voltage are therefore shown in Fig. 5.32.

5.4.1.2 Large Signal Measurement

Figure 5.33 shows the measured gain and normalized output phase for the three modes. The dynamic mode demonstrates the expected transition between the low and full power modes and can therefore increase the linear region. Furthermore, the dynamic mode also demonstrates excellent phase characteristics with a maximum phase distortion of 2° . The dynamic mode should not only improve the linearity but also improve the efficiency at back-off. The PAE and DC-power consumption of the PA are shown in Fig. 5.34. It shows the expected efficiency curves next to the ideal class-A curve; although the difference between full power and dynamic power is less than expected, some benefit is seen. Due to the reduced bias voltage, the PA will operate more toward class-AB and will already show some efficiency improvement compared to a class-A PA. The class-AB behavior of the full power mode can also be seen in the DC-power consumption curves. Furthermore, the dynamic mode saves about a quarter of power at back-off compared to the full power mode, but some overhead does not let it go down as far as the low power mode.

The rectifier at the input of the main output stage controls how the peaking PA turns on. The rectifier is controlled by 3 voltages: the bias voltage of the rectifier transistors (V_{rect}), the current through or the conversion gain of the rectifier is

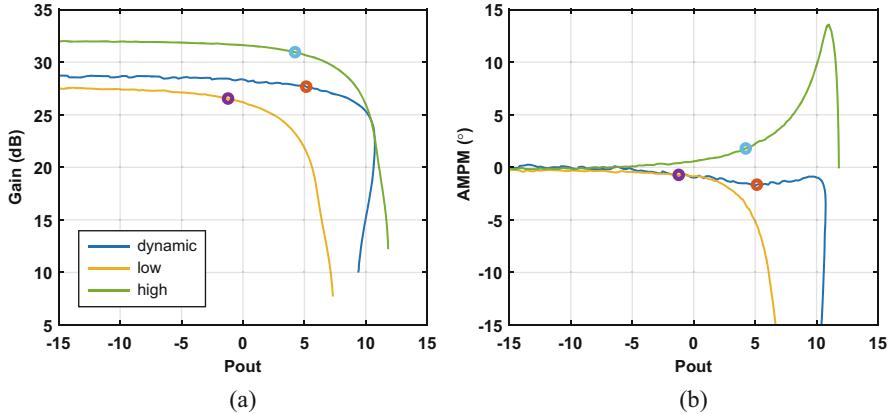


Fig. 5.33 Measured large signal gain (a) and AM-PM (b) over output power for the three modes

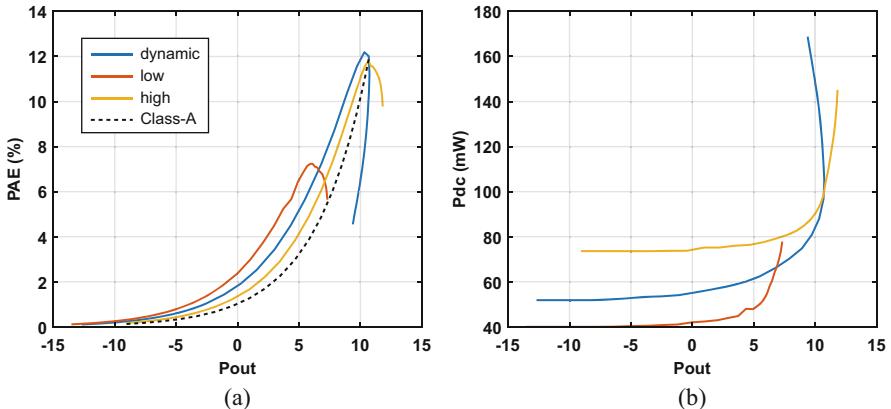


Fig. 5.34 Measured PAE (a) and DC-power consumption (b) over output power for the three modes

controlled by (V_{gain}), and the base offset voltage (V_{offset}). The effects of these control voltages, when one is changed and the other 2 are kept constant, are shown in Fig. 5.35 with the black dot representing the OP1dB.

Next to the control of the AM-AM, also AM-PM can be controlled. Figure 5.36 shows the effect of the two varactors on the gain (a,b) and phase (c,d). Figure 5.36b shows the AM-PM curves for a changing fixed phase shift between the main and peaking PA. It shows the ability to compensate for the AM-PM compression of the main PA, while little penalty is paid in the gain as shown in Fig. 5.36a. Figure 5.36d shows the AM-PM curves for the varactor seeing the changing bias voltage of the rectifier. It also demonstrates this can be used to compensate AM-PM if a changing bias is present in the PA, for little penalty to the gain as shown in Fig. 5.36c.

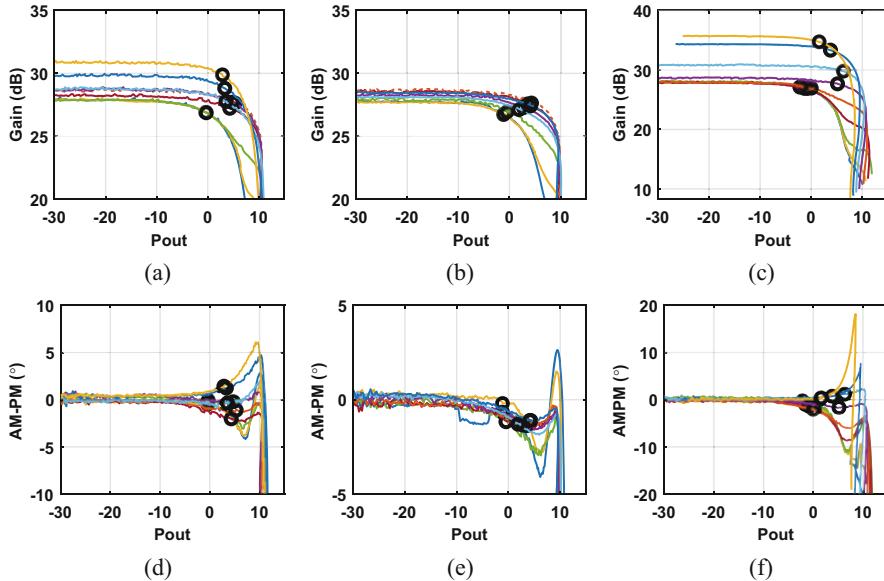


Fig. 5.35 Measured gain (a)–(c) and phase (d)–(f) characteristic for the 3 rectifier control voltages, change of V_{rect} , V_{gain} , and V_{offset}

5.4.1.3 Conclusion PA Measurements

The measured PA in dynamic mode shows a saturated output power of 11 dBm with a peak efficiency of 12.2%, at the OP1dB back-off; the efficiency drops down to 5.1% but still shows a 60% improvement compared to a class-A reference PA and the full power mode PA that has a lower OP1dB. The dynamic mode has proven to improve the linearity for both AM–AM and AM–PM. Although the applied bias reduction has allowed for the rectifier to show operation, the gain is still larger compared to simulation. Therefore, the rectifier operated on the edge and no gain expansion has been shown.

5.4.2 Transmitter Measurements

5.4.2.1 Continuous Wave Measurements and Characterization

To characterize multiple facets of the transmitter, multiple measurements are possible each with a different measurement setup. Three different setups are possible as shown in Fig. 5.37.

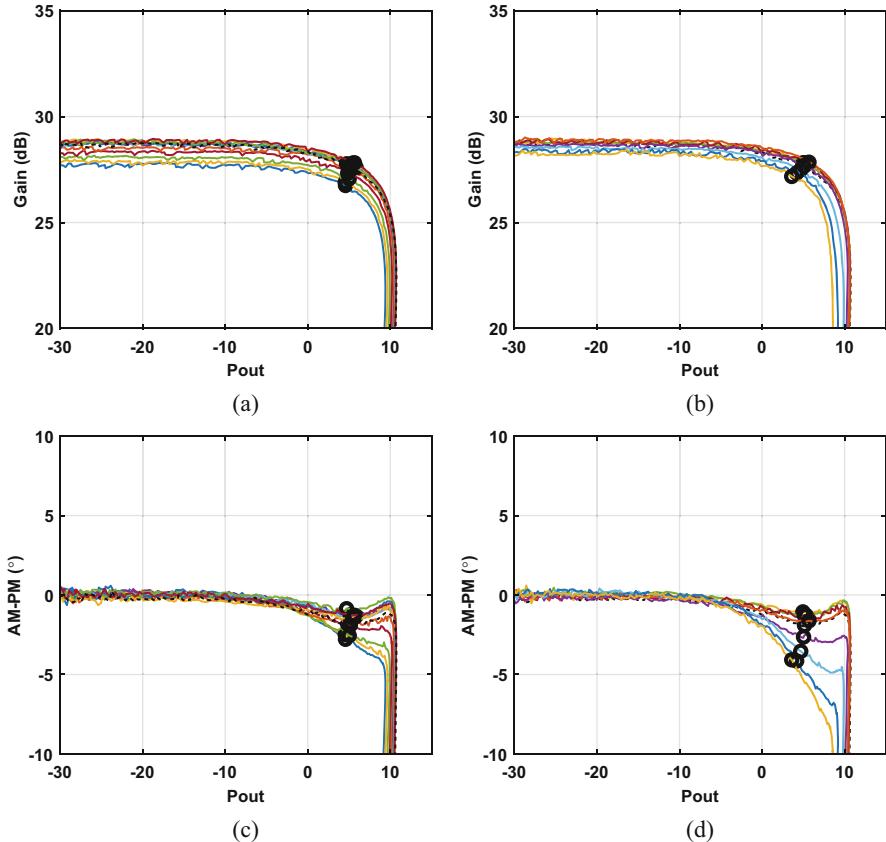


Fig. 5.36 Measured gain and phase for changing varactor of fixed and dynamic phase control, with the dashed black line demonstrating the nominal case

The first measurement bypasses the mixer and IF path by applying a DC offset on one of the mixers. This measures the performance of the LO path in combination with the PA. Figure 5.38 shows the output power at 3 times the input frequency against the power at the input of the LO. A saturated output power level is reached for an input power of -5 dBm , this demonstrates no excess input LO power is required to operate the transmitter. For large LO powers, the transmitter can be driven into saturation, and the maximum output power can be measured. They show the expected slope of 3 dB/dB for a tripler circuit [13]. The calibrated saturated output power over frequency for the 3 operating modes is shown in Fig. 5.39. The transmitter demonstrates a saturated output power of 12 dBm for the full and dynamic modes, while the low power mode shows a reduced output power of 8 dBm . Furthermore, the loading of the turned-off peaking amplifier reduces the output power at higher frequencies.

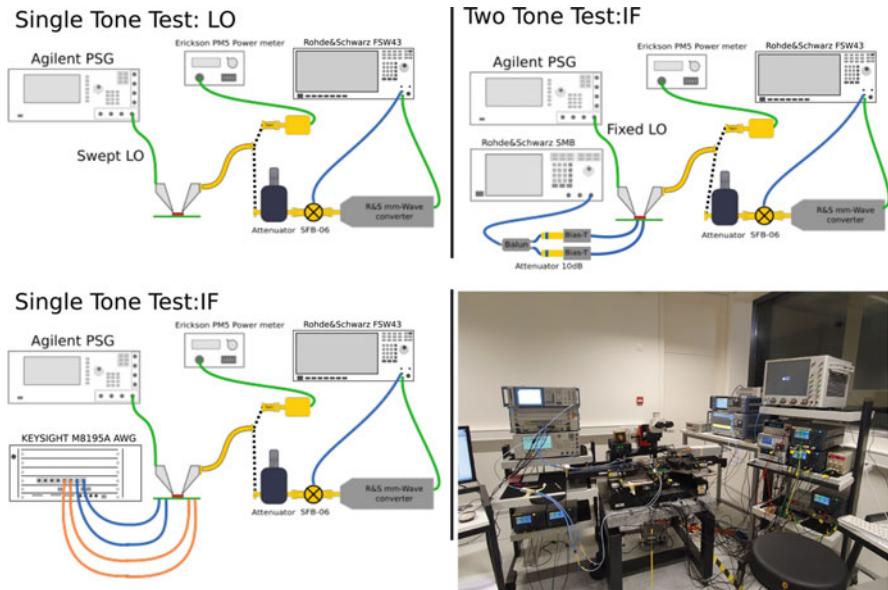


Fig. 5.37 The measurement setups used for characterization of the transmitter: single tone LO sweep, single tone IQ IF sweep, and dual tone single mixer IF sweep

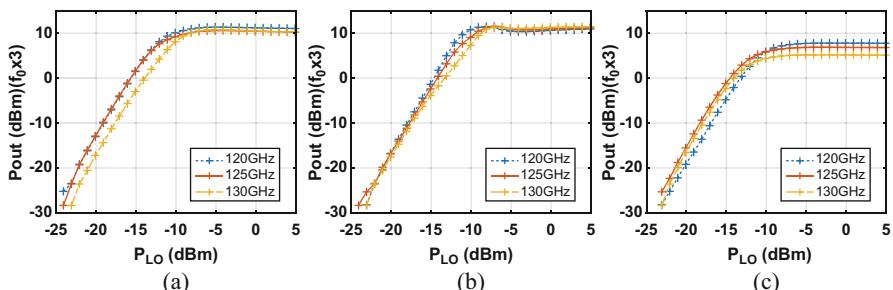


Fig. 5.38 The measured output power for an increasing LO power in (a) full power mode, (b) dynamic-power mode, and (c) low power mode

A second measurement applies a single tone f_{if} to only one of the IF inputs using a signal generator, balun, and bias-T. Only one mixer is used resulting in a two-tone test at $f_{lo} - f_{if}$ and $f_{lo} + f_{if}$. The output is probed, and an external mixer downconverts the signal that is analyzed on the spectrum analyzer. Figure 5.40 shows the two tones, the third-order intermodulation product, and LO-feedthrough (LOFT) for the three modes of operations using a 1 GHz of input signal. The hand-off between low and full power modes by the dynamic bias is again visible with also a slightly reduced IM3. Figure 5.41 shows measured two tones, IM3 and LOFT in

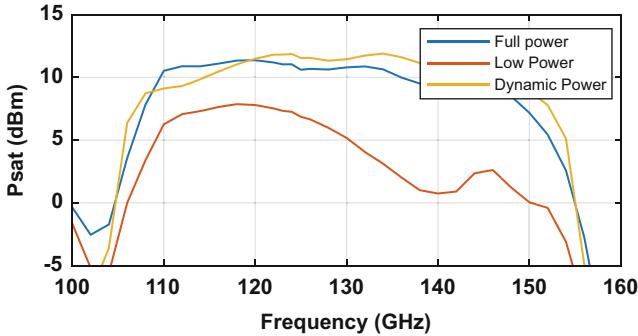


Fig. 5.39 The measured saturated output power of the transmitter in the three modes

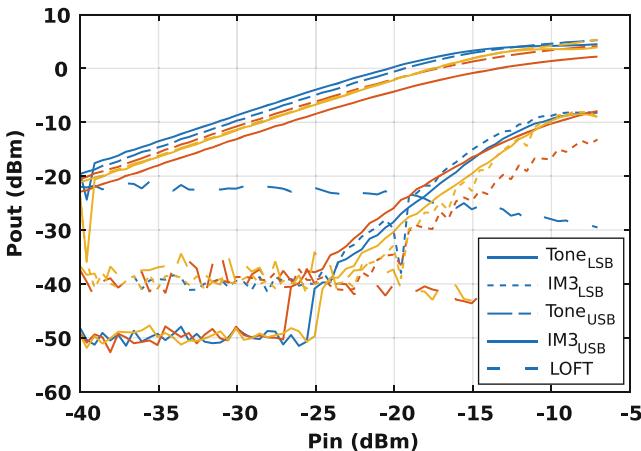


Fig. 5.40 The measured power for the two tones, the IM3 and LO for an IF of 1 GHz, in three power modes (high, dynamic, and low)

full power mode for three different LO input powers. The increased LO-input power allows the increase or decrease of the conversion gain of the modulator.

The final characterization measurement applies two f_{if} quadrature tones to generate a single tone and its suppressed image. The two quadrature signals are generated by a Keysight AWG, and the full setup is shown in Fig. 5.37. The first two measurements capture the quadrature mixer's performance given by the LO-feedthrough and the IMRR as shown in Fig. 5.42. The left-hand side shows the LO-feedthrough compared to the transmitted signal expressed as dBc. For no applied offset voltage between the I and Q channels, we can achieve the lowest LO-feedthrough demonstration low inherent feedthrough by the mixer. The right-hand side shows the IMRR at 120, 125, and 130 GHz, and the IMRR can at least be

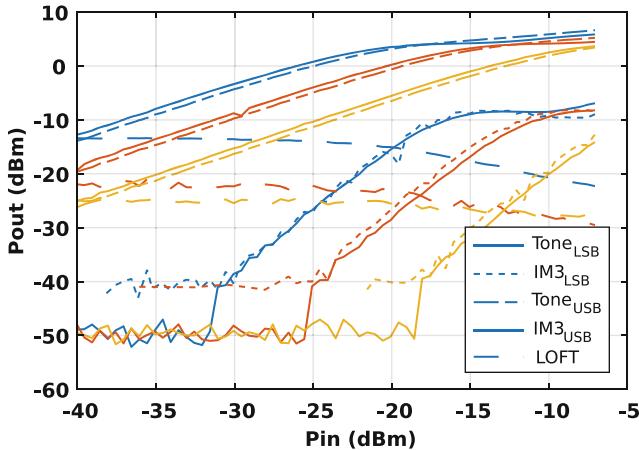


Fig. 5.41 The measured power for the two tones, the IM3 and LO for an IF of 1 GHz in full power mode for 3 LO powers: -10,-5,0 dBm

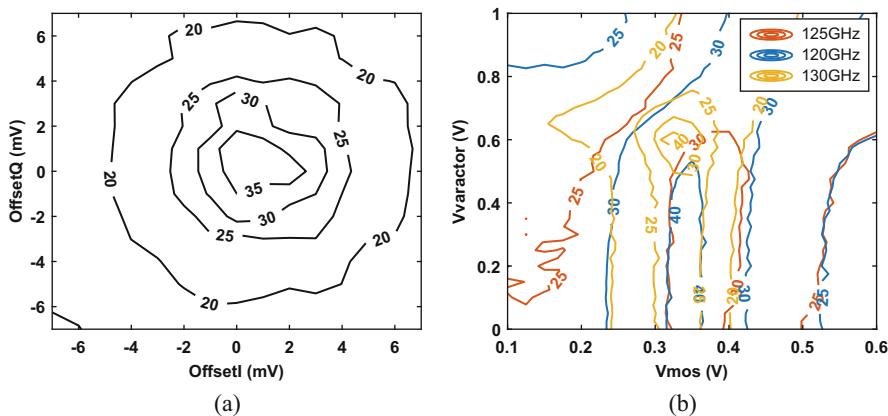


Fig. 5.42 The measured (a) LO-feedthrough compared to the carrier for DC offset voltages and (b) the IMRR for the 2 control voltages of the NMOS and varactor

tuned down to 30 dB using the adjustable voltage of the NMOS and varactor of the quadrature circuit. Figure 5.43 shows the conversion gain at different LO powers by sweeping the IF frequency from 0.1 GHz up to 10 GHz, for a fixed LO of 125 GHz at both sidebands.

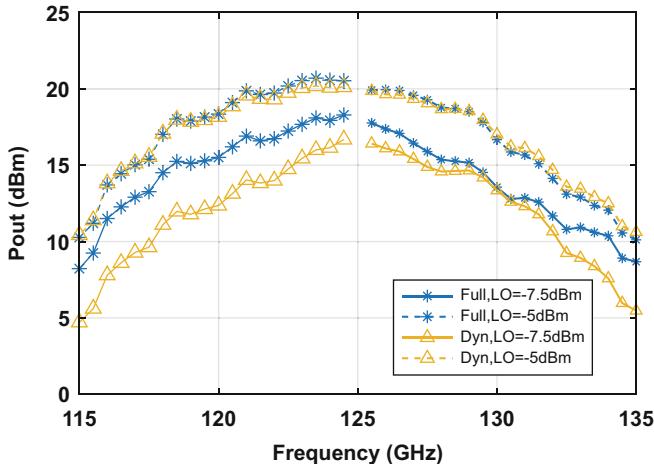


Fig. 5.43 The measured IF-to-RF conversion gain for swept IF frequency with an LO of 125 GHz

5.4.2.2 Modulated Data Measurements

To verify the full transmitter capability, modulated data measurements are performed by transmitting data and demodulating using measurement equipment. Figure 5.44 shows the measurement setup; on the transmit side, a Keysight AWG is used to create zero-IF differential I and Q signals that are brought on-chip through a bond wire connection. A PSG provides the 125/3 GHz LO and is brought on-chip with a GSG probe. The modulated RF signal is received by probing the output of the transmitter. The modulated signal first goes through a D-band attenuator, and this attenuator ensures the mixer in the next stage is not driven into compression. The SFB-06 downconverting mixer, with an 134.4 GHz LO generated by A R&S signal generator in combination with a R&S frequency converter, downconverts the signal to a 9.4 GHz carrier. The downconverted signal is amplified by a preamplifier before being connected to a Keysight oscilloscope. The data on the received IF signal is demodulated with the VSA software on the scope. The output power of the modulated data measurements is measured by integration on the scope and with a PM5b power meter.

The data rate and average output power for a 16-QAM signal in the 3 modes are shown in Fig. 5.45 for an EVM of at most -17 dB. Furthermore, a non-equalized measurement is performed, which allows data rates up to 12 Gbps, and static equalization is used, which boosts the maximum data rate up to 50 Gbps. The equalization will compensate for the amplitude and phase response of the input cables, bond wire transition, transmitter, and the whole receive chain. Figure 5.46 shows the measured EVM and efficiency against the output power for a 16-QAM 1 Gbaud signal by changing the IF-input power. The EVM varies from -16 dB, when the constellation clearly shows compression at the high output powers, down to -23 dB at back-off where the noise is the main cause for error. The expected

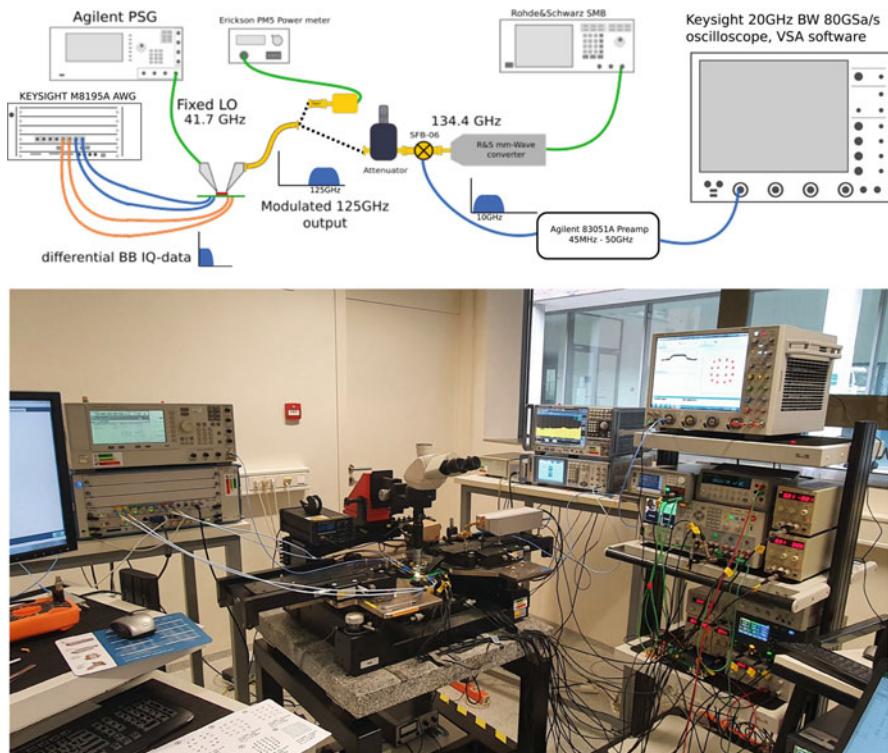


Fig. 5.44 The setup for modulated data measurements

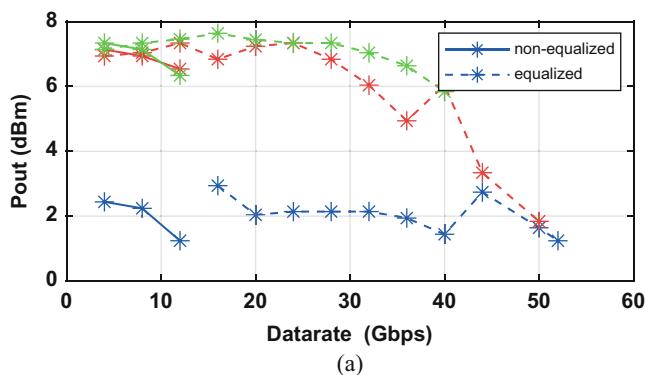


Fig. 5.45 Summary of 16-QAM data measurements for the 3 PA modes with the maximal output power for each data rate for a minimum EVM of 17 dB

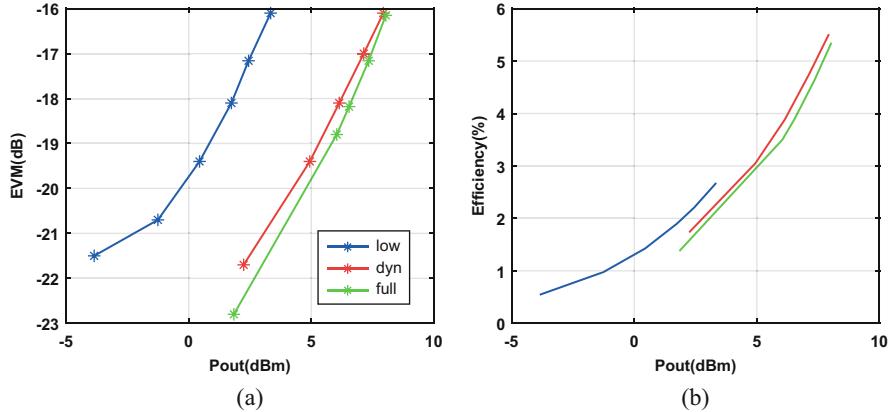


Fig. 5.46 The measured (a) EVM and (b) efficiency against the output power for a 16-QAM 1GBaud signal

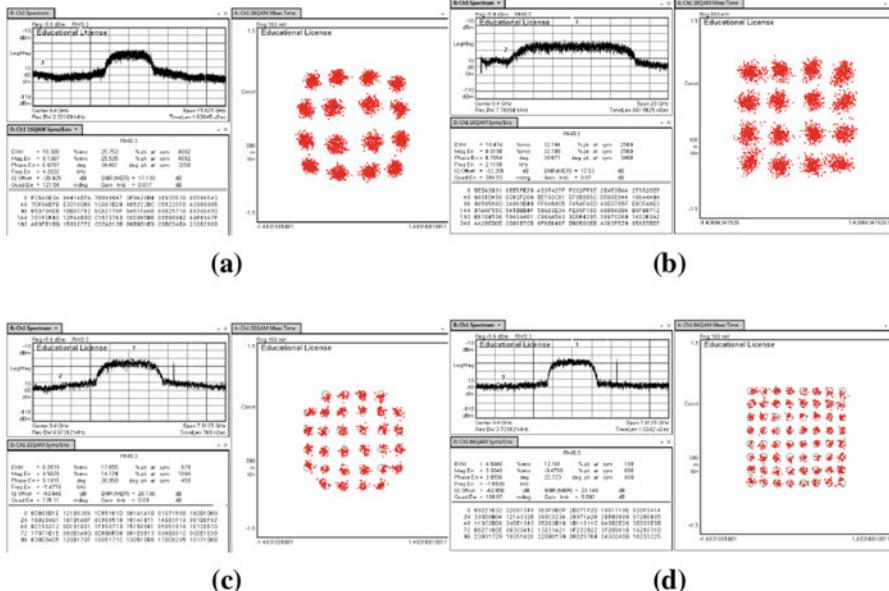


Fig. 5.47 Measured constellation, spectrum, and EVM summary of the transmitter in full power mode for (a) 16-QAM 3GBaud, (b) 16-QAM 10Gbaud equalized, (c) 32-QAM 2Gbaud, and (d) 64-QAM 1.5Gbaud constellation

difference in efficiency is similar to the PA. However, a good efficiency of 4.5% is still achieved for the overall system.

The constellation diagram, EVM summary, and spectrum of 4 data points are shown for full and dynamic-power modes in Figs. 5.47 and 5.48, respectively. It

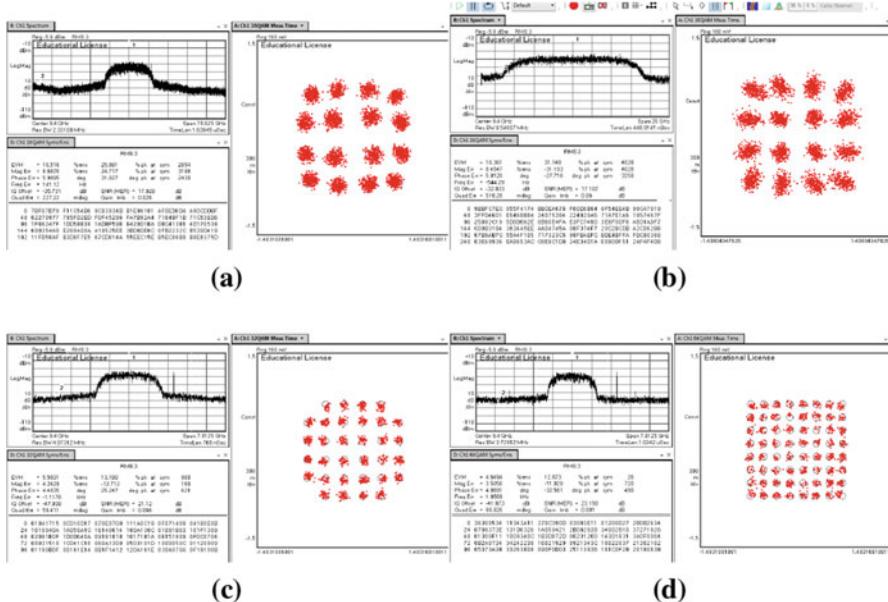


Fig. 5.48 Measured constellation, spectrum, and EVM summary of the transmitter in dynamic-power mode for (a) 16-QAM 3GBaud, (b) 16-QAM 10Gbaud equalized, (c) 32-QAM 2Gbaud, and (d) 64-QAM 1.5Gbaud constellation

shows the constellation of the 16-QAM highest data rate for both unequalized and equalized measurements. Furthermore, higher spectral efficient modulation schemes such as 32-and 64-QAM are also shown. Although the constellation diagram looks good, these modulation types require a higher SNR that is limited by the receiver of the measurement setup.

5.5 Conclusion

This chapter discussed the issues of using classical power amplifier efficiency enhancement techniques above 100 GHz. It is followed by the proposed design of a sequential amplifier and demonstrated that IL needs to be balanced for each technique leading to the final power amplifier. The amplifier allows being operated in three modes: low power, full power, and dynamic mode improving efficiency and linearity of the PA. The PA is further extended with a full IQ modulator to make an IQ transmitter. The measured transmitter shows a saturated output power of 12 dBm and a maximal data rate of 44 Gbps at an EVM of -17 dB using a 16-QAM signal.

References

1. E. McCune, *Practical Digital Wireless Signals* (Cambridge University, Cambridge, 2010)
2. M. Vigilante, E. McCune, P. Reynaert, To EVM or two EVMs?: An answer to the question. *IEEE Solid-State Circuits Mag.* **9**(3), 36–39 (2017)
3. W.H. Doherty, A new high efficiency power amplifier for modulated waves. *Proceedings of the Institute of Radio Engineers* **24**(9), 1163–1182 (1936)
4. E. Kaymaksut, P. Reynaert, Transformer-based uneven Doherty power amplifier in 90 nm CMOS for WLAN applications. *IEEE J. Solid-State Circuits* **47**(7), 1659–1671 (2012)
5. E. Kaymaksut, D. Zhao, P. Reynaert, Transformer-based Doherty power amplifiers for mm-Wave applications in 40-nm CMOS. *IEEE Trans. Microwave Theory Tech.* **63**(4), 1186–1192 (2015)
6. P. Indirayanti, P. Reynaert, A 32 GHz 20 dBm-P_{SAT} transformer-based Doherty power amplifier for multi-Gb/s 5G applications in 28 nm bulk CMOS, in *2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)* (2017), pp. 45–48
7. S. Hu, F. Wang, H. Wang, A 28-/37-/39-GHz linear Doherty power amplifier in silicon for 5G applications. *IEEE J. Solid-State Circuits* **54**(6), 1586–1599 (2019)
8. F. Wang, T. Li, S. Hu, H. Wang, A super-resolution mixed-signal Doherty power amplifier for simultaneous linearity and efficiency enhancement. *IEEE J. Solid-State Circuits* **54**(12), 3421–3436 (2019)
9. H.T. Nguyen, H. Wang, A coupler-based differential mm-wave Doherty power amplifier with impedance inverting and scaling baluns. *IEEE J. Solid-State Circuits* **55**(5), 1212–1223 (2020)
10. D. Zhao, S. Kulkarni, P. Reynaert, A 60-GHz outphasing transmitter in 40-nm CMOS. *IEEE J. Solid-State Circuits* **47**(12), 3172–3183 (2012)
11. S. Li, M.Y. Huang, D. Jung, T.Y. Huang, H. Wang, A mm-wave current-mode inverse outphasing transmitter front-end: A circuit duality of conventional voltage-mode outphasing. *IEEE J. Solid-State Circuits* **56**(6), 1732–1744 (2020)
12. S. Li, T. Chi, J. Park, H.T. Nguyen, H. Wang, A 28-GHz flip-chip packaged Chireix transmitter with on-antenna outphasing active load modulation. *IEEE J. Solid-State Circuits* **54**(5), 1243–1253 (2019)
13. A. Standaert, P. Reynaert, A 390-GHz outphasing transmitter in 28-nm CMOS. *IEEE J. Solid-State Circuits* **55**(10), 2703–2713 (2020)
14. I. Abdo, C.d. Gomez, C. Wang, K. Hatano, Q. Li, C. Liu, K. Yanagisawa, A.A. Fadila, J. Pang, H. Hamada, H. Nosaka, A. Shirane, K. Okada, 22.2 a 300GHz-band phased-array transceiver using bi-directional outphasing and Hartley architecture in 65nm CMOS, in *2021 IEEE International Solid- State Circuits Conference (ISSCC)*, vol. 64 (2021), pp. 316–318
15. R.P. Dixian Z., *CMOS 60-GHz and E-band Power Amplifiers and Transmitters* (Springer, Berlin, 2015)
16. D. Zhao, P. Reynaert, A 60-GHz dual-mode class ab power amplifier in 40-nm CMOS. *IEEE J. Solid-State Circuits* **48**(10), 2323–2337 (2013)
17. S.C. Cripps, *RF Power Amplifiers for Wireless Communications*. Artech House, Inc., Norwood, (2006)
18. T. Lehmann, R. Knöchel, Design and performance of sequential power amplifiers, in *2008 IEEE MTT-S International Microwave Symposium Digest* (2008), pp. 767–770
19. H.T. Nguyen, T. Chi, S. Li, H. Wang, A 62-to-68GHz linear 6Gb/s 64QAM CMOS Doherty radiator with 27.5 %/20.1% PAE at peak/6dB-back-off output power leveraging high-efficiency multi-feed antenna-based active load modulation, in *2018 IEEE International Solid - State Circuits Conference—(ISSCC)* (2018), pp. 402–404
20. H.T. Nguyen, T. Chi, S. Li, and H. Wang, A linear high-efficiency millimeter-wave CMOS Doherty radiator leveraging multi-feed on-antenna active load modulation. *IEEE J. Solid-State Circuits* **53**(12), 3587–3598 (2018)

21. H.T. Nguyen, S. Li, H. Wang, 4.6 a mm-wave 3-way linear Doherty radiator with multi antenna coupling and on-antenna current-scaling series combiner for deep power back-off efficiency enhancement, in *2019 IEEE International Solid-State Circuits Conference—(ISSCC)* (2019), pp. 84–86
22. B. Razavi, *RF Microelectronics*. Prentice Hall Press, London (2011)
23. S. Kulkarni, P. Reynaert, A 60-GHz power amplifier with AM–PM distortion cancellation in 40-nm CMOS. *IEEE Trans. Microwave Theory Tech.* **64**(7), 2284–2291 (2016)
24. S. Golara, S. Moloudi, A.A. Abidi, Processes of AM-PM distortion in large-signal single-FET amplifiers. *IEEE Trans. Circuits Syst. I Regul. Pap.* **64**(2), 245–260 (2017)
25. M. Vigilante, P. Reynaert, A wideband class-AB power amplifier with 29–57-GHz AM–PM compensation in 0.9-V 28-nm bulk CMOS. *IEEE J. Solid-State Circuits* **53**(5), 1288–1301 (2018)
26. Y. Zhang, P. Reynaert, A high-efficiency linear power amplifier for 28 GHz mobile communications in 40nm CMOS, in *2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)* (2017), pp. 33–36
27. S.N. Ali, P. Agarwal, J. Baylon, S. Gopal, L. Renaud, D. Heo, A 28GHz 41%-PAE linear CMOS power amplifier using a transformer-based AM-PM distortion-correction technique for 5G phased arrays, in *2018 IEEE International Solid-State Circuits Conference—(ISSCC)* (2018), pp. 406–408
28. U. Çelik, P. Reynaert, An E-band compact power amplifier for future array-based Backhaul networks in 22nm FD-SOI, in *2019 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)* (2019), pp. 187–190
29. U. Celik, P. Reynaert, An NP matrix-based E-band power amplifier achieving 24-Gbit/s data rate using 256 QAM in 22-nm FD-SOI. *IEEE Trans. Microwave Theory Tech.* **69**(11), 4667–4677 (2021)
30. S. Shakib, H. Park, J. Dunworth, V. Aparin, K. Entesari, A highly efficient and linear power amplifier for 28-GHz 5G phased array radios in 28-nm CMOS. *IEEE J. Solid-State Circuits* **51**(12), 3020–3036 (2016)
31. M. Vigilante, P. Reynaert, 20.10 a 68.1-to-96.4GHz variable-gain low-noise amplifier in 28 nm CMOS, in *2016 IEEE International Solid-State Circuits Conference (ISSCC)* (2016), pp. 360–362
32. B. Philippe, P. Reynaert, A Quadrature Phase Detector in 28nm CMOS for Differential mm-Wave Sensing Applications Using Dielectric Waveguides, in *Proceedings of the IEEE European Solid State Circuits Conference (ESSCIRC)* (IEEE, New York, 2018), pp. 114–117
33. S. Ooms, P. Reynaert, A 120-GHz wireless link using 3D-printed lens with flexible dielectric fiber feed and 28-nm CMOS transceiver. *IEEE Solid-State Circuits Lett.* **3**, 142–145 (2020)

Index

A

Accumulation-mode MOS varactor, 40, 106, 107, 113, 115, 122, 127
Active device, 7, 44, 50, 78–80, 83–85, 111
AM-AM distortion, 72–74, 77, 90, 91, 101, 105, 106, 114, 122, 124
AM-PM distortion, 73, 74, 77, 106–109, 113, 122, 124
Antenna gain, 3

B

The back end of line (BEOL), 7, 17, 18, 20, 21, 83
Bandwidth, 2, 29

C

Capacitive neutralization, 26, 27, 58, 76, 85, 87, 107, 115
Capacitor, 18, 19, 40, 41
Cascode amplifier, 81
Characteristic impedance, 25
Conjugate match, 28, 29, 69, 85

D

Digital pre-distortion (DPD), 106
Directional coupler, 101
Doherty power amplifier, 97, 98
Drain efficiency (DE), 71, 72

E

Efficiency, 39, 71, 76, 100
Error vector magnitude (EVM), 73, 105, 130

F

Figure of merit (FoM), 8, 43, 44, 46
Figure of merit tuning (FoMt), 44
Fin field-effect transistor (FinFET), 3, 11
Free space path loss (FSPL), 3
Frequency multiplier, 110, 115

G

Gain match, 70, 81, 85, 100
Gate resistance, 9, 12, 15, 45, 46

H

Harmonic signal, 43, 55, 74, 115, 116
Hybrid coupler, 85, 104

I

Image rejection ratio (IMRR), 115, 126, 127
Impedance matching, 9, 29, 30, 48, 86
Impulse sensitivity function (ISF), 38
Inductor, 20–22, 42
Insertion loss (IL), 32, 33, 82, 85, 102–104, 110

L

Layout effects, 9
LC-resonator, 37, 38, 47, 50
Load-line theory, 70
Load-modulation, 100, 101

M

- Magnetic coupling, 22, 23, 31, 47, 56, 113, 115, 116
 Main PA, 97, 101, 106, 122
 Maximal oscillation frequency (f_{max}), 4, 9, 13
 Metal-oxide-metal (MOM) capacitor, 18, 19, 52
 Millimeter-wave, 1, 2, 32
 Mixer, 54, 110, 116

O

- Outphasing power amplifier (Outphasing PA), 99

P

- Peaking PA, 97, 101, 102, 105–108, 113, 122
 Peak to average power ratio (PAPR), 97
 Phase locked loop (PLL), 37, 42, 43
 Phase noise (PN), 38, 39, 44
 Power added efficiency (PAE), 69, 72, 77, 78, 90, 91, 102–104, 108, 113, 114
 Power amplifier (PA), 69, 71, 91, 97–99, 101, 102, 104, 110
 Power combining, 83, 85, 86, 100, 101, 104, 108, 110
 Power match, 70, 81, 85, 100

Q

- Quadrature amplitude modulation(QAM), 97, 98, 105, 129–131
 Quality factor (Q), 2, 18, 20, 22, 23, 29–32, 37, 39–41, 52, 53, 58, 100, 106, 107

R

- Rectifier, 102, 112, 121
 Reflection coefficient, 28, 29
 Ripple, 30

S

- Sequential power amplifier (SPA), 101, 102, 104
 Signal-to-noise ratio (SNR), 2, 69
 Sixth-generation (6G), 1
 Source degeneration, 106
 Stability, 26, 27, 88, 90
 Switched capacitor, 41, 57
 Switched inductor, 57

T

- Transformer, 22, 30, 47, 52, 56, 86, 87, 115, 116
 Transition frequency (f_T), 4, 9
 Transmission line, 25, 26, 85, 98, 104
 Transmitter, 69, 71, 99, 110, 115, 118, 128
 Tuning range (TR), 39–41, 43–46, 50, 52, 54, 56, 58, 62, 63