

Analog Circuit Design

Sensors, Actuators and Power Drivers;
Integrated Power Amplifiers from Wireline
to RF; Very High Frequency Front Ends

Edited by

Herman Casier

Michiel Steyaert

Arthur H.M. van Roermund



Springer

Analog Circuit Design

Herman Casier • Michiel Steyaert
• Arthur H.M. van Roermund
Editors

Analog Circuit Design

Sensors, Actuators and Power Drivers;
Integrated Power Amplifiers from Wireline
to RF; Very High Frequency Front Ends



Springer

Editors:

Herman Casier
AMI Semiconductor
Oudenaarde
Belgium

Michiel Steyaert
KU Leuven
Belgium

Arthur H.M. van Roermund
Technical University of Eindhoven
The Netherlands

ISBN 978-1-4020-8262-7

e-ISBN 978-1-4020-8263-4

Library of Congress Control Number: 2008924617

© 2008 Springer Science + Business Media B.V.

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Contents

Preface	vii
Part I: Sensors, Actuators and Power Drivers for the Automotive and Industrial Environment	1
Heterogeneous Integration of Passive Components for the Realization of RF-System-in-Packages	3
<i>Eric Beyne, Walter De Raedt, Geert Carchon, and Philippe Soussan</i>	
The Eye-RIS CMOS Vision System.....	15
<i>Ángel Rodríguez-Vázquez, Rafael Domínguez-Castro, Francisco Jiménez-Garrido, Sergio Morillas, Juan Listán, Luis Alba, Cayetana Utrera, Servando Espejo and Rafael Romay</i>	
An Inductive Position Sensor ASIC	33
<i>Petr Kamenicky, Pavel Horsky</i>	
CMOS Single-Chip Electronic Compass with Microcontroller	55
<i>Christian Schott, Robert Racz, Samuel Huber, Angelo Manco, Markus Gloor, Nicolas Simonne</i>	
Protection and Diagnosis of Smart Power High-Side Switches in Automotive Applications.....	71
<i>Andreas Kucher</i>	
Part II: Integrated PA's: from Wireline to RF	91
Integrated CMOS Power Amplifiers for Highly Linear Broadband Communication	93
<i>K. Mertens, M. Unterweissacher, M. Tiebout, C. Sandner</i>	
Power Combining Techniques for RF and mm-Wave CMOS Power Amplifiers.....	115
<i>Patrick Reynaert, M. Bohsali, D. Chowdhury and A. M. Niknejad</i>	

Switched RF Transmitters	145
<i>Willem Lafleure, Michiel Steyaert and Jan Craninckx</i>	
High-Speed Serial Wired Interface for Mobile Applications	163
<i>Gerrit W. den Besten</i>	
High Voltage xDSL Line Drivers in Nanometer Technologies.....	179
<i>Bert Serneels, Michiel Steyaert, Wim Dehaene</i>	
VoIP SLIC Open Platform: The Wideband Subscriber Line Interface Circuit for Voice over IP (VoIP) Applications	205
<i>Luc D'Haeze, Jan Sevenhuijsen, Herman Casier, Damien Macq, Stefan van Roeyen, Stef Servaes, Geert De Pril, Koen Geirnaert, Hedi Hakim</i>	
Part III: Very High Frequency Front Ends	235
Systems and Architectures for Very High Frequency Radio Links.....	237
<i>Peter Baltus, Peter Smulders, Yikun Yu</i>	
Key Building Blocks for Millimeter-Wave IC Design in Baseline CMOS	259
<i>Mihai A.T. Sanduleanu, Eduardo Alarcon, Hammad M. Cheema, Maja Vidojkovic, Reza Mahmoudi and Arthur van Roermund</i>	
Analog/RF Design Concepts for High-Power Silicon Based mmWave and THz Applications	283
<i>Ullrich R. Pfeiffer</i>	
SiGe BiCMOS and CMOS Transceiver Blocks for Automotive Radar and Imaging Applications in the 80-160 GHz Range	303
<i>S.P. Voinigescu, S. Nicolson, E. Laskin, K. Tang and P. Chevalier</i>	
A Comparison of CMOS and BiCMOS mm-Wave Receiver Circuits for Applications at 60GHz and Beyond	327
<i>Sharon Malevsky and John R. Long</i>	
Integrated Frontends for Millimeterwave Applications Using III-V Technologies	343
<i>Herbert Zirath, Sten E. Gunnarsson, Camilla Kärnfelt, Toru Masuda, Mattias Ferndahl, Rumen Kozuharov, Arne Alping</i>	

Preface

This book is part of the *Analog Circuit Design* series and contains the revised contributions of all speakers of the 16th AACD Workshop, which was organized by Jan Sevenhans of AMI Semiconductor and held in Oostende, Belgium on March 27-29, 2007. The book comprises 17 tutorial papers, divided in three chapters, each discussing a very relevant topic in present days analog design.

1. Sensors, Actuators and Power Drivers for the Automotive and Industrial Environment
2. Integrated PA's: from Wireline to RF
3. Very High Frequency Front Ends

These papers were presented by experts in the field during the workshop. They were selected by the organizer and the program committee consisting of Herman Casier of AMI Semiconductor Belgium, Prof. Michiel Steyaert from Katholieke Universiteit Leuven, Belgium and Prof. Arthur van Roermund from Eindhoven University of Technology, The Netherlands, who are also the editors of this book.

The aim of the AACD workshop is to bring together a group of expert designers to study and discuss new possibilities and future developments in the area of analog circuit design. Each AACD workshop has given rise to the publication of a book by Springer in their successful series of Analog Circuit Design. For the previous books and topics in the series, see next page.

The series provides a valuable overview of analog circuit design and related CAD, mainly in the fields of basic analog modules, mixed-signal electronics, AD and DA converters, RF systems and automotive electronics. It is a reference for whoever is engaged in these disciplines and wishes to keep abreast of the latest developments in the field.

We sincerely hope that this 16th book continues the tradition and provides a valuable contribution to our Analog Design Community.

Herman Casier

Previous Books in *Analog Circuit Design*

2006	Maastricht (The Netherlands)	High-Speed AD Converters Automotive Electronics: EMC issues Ultra Low Power Wireless
2005	Limerick (Ireland)	RF Circuits: Wide Band, Front-Ends, DAC's Design Methodology and Verification of RF and Mixed-Signal Systems Low Power and Low Voltage
2004	Montreux (Swiss)	Sensor and Actuator Interface Electronics Integrated High-Voltage Electronics and Power Management Low-Power and High-Resolution ADCs
2003	Graz (Austria)	Fractional-N Synthesizers Design for Robustness Line and Bus drivers
2002	Spa (Belgium)	Structured Mixed-Mode Design Multi-Bit Sigma-Delta Converters Short-Range RF Circuits
2001	Noordwijk (The Netherlands)	Scalable Analog Circuit Design High-Speed D/A Converters RF Power Amplifiers
2000	Munich (Germany)	High-Speed A/D Converters Mixed-Signal Design PLLs and Synthesizers
1999	Nice (France)	(X)DSL and other Communication Systems RF-MOST Models Integrated Filters and Oscillators
1998	Copenhagen (Denmark)	1-Volt Electronics Design and Implementation of Mixed-Mode Systems Low-Noise and RF Power Amplifier for Communications

1997	Como (Italy)	RF Analog to Digital Converters Sensor and Actuator Interfaces Low-Noise Oscillators, PLLs and Synthesizers
1996	Lausanne (Swiss)	MOST RF Circuit Design Bandpass Delta-Sigma and Other Data Converters Translinear Circuits
1995	Villach (Austria)	Low-Noise, Low-Power, Low-Voltage Mixed-Mode design with CAD tools Voltage, Current and Time References
1994	Eindhoven (The Netherlands)	Low-Power Low-Voltage Integrated Filters Smart Power
1993	Leuven (Belgium)	Mixed Analogue-Digital Circuit Design Sensor Interface Circuits Communication Circuits
1992	Scheveningen (The Netherlands)	Operational Amplifiers Analog to Digital Conversion Analog Computer Aided Design

Part I: Sensors, Actuators and Power Drivers for the Automotive and Industrial Environment

Man and machine perceive and control the physical world through physical parameters such as force and pressure, speed and acceleration, temperature, gas composition, electromagnetic fields, light... These parameters are, with a few exceptions, not electrical and an interface is required to measure and control them by an electrical system. Sensors translate the physical parameter in an electrical current or voltage and actuators do the opposite.

In the past, sensors and actuators were fabricated as discrete components and only their electrical interface was put on the chip. They were optimized for sensitivity and stability. Nowadays, more sensors and higher power actuator drivers are integrated on the controller chip. Due to the resulting technology limitations, the on-chip sensor has a lower sensitivity but since controllability and calibration flexibility improves and since interference from the environment on the sensitive connections between sensor and control chip is greatly reduced, the final system sensitivity comes close or even exceeds the sensitivity of the discrete sensor system.

The first paper describes the integration of high-quality passive devices on active wafers (RF-SOC) or on an intermediate glass or high resistivity silicon substrate (RF-SIP). Multilayer thin film technology allows the realization of passives with relevant values and quality for use in RF-applications. Resistors, capacitors, inductors and the device platform are described.

A smart CMOS camera, where image acquisition and processing are truly intermingled, is the subject of the next paper. The signal processing is realized in two steps and resembles natural vision systems. At the first step the data rate of the parallel vision signals is reduced by analog processing. At the second step, intelligent processing is realized on digitally-coded information data by means of digital processors.

The third paper describes an inductive contact-less sensor system for high resolution angular or linear position sensing, which is well suited for automotive applications. The sensor is a cheap PCB pattern and the ASIC integrates the actuator driver, the sensor interface and the analog and digital signal processing. Besides the circuits, also the special automotive and safety issues are detailed.

2 Part I: Sensors, Actuators and Power Drivers

Currently, discrete, high mobility compound semiconductor Hall devices are used for the measurement of low magnetic fields such as the earth magnetic field. The fourth paper shows how these discrete devices can be replaced by an integrated, low sensitivity Hall effect sensor on a low voltage CMOS technology with an integrated magnetic concentrator post-processing. Extensive analog and digital signal processing and calibration is used.

The last paper focuses on a high-side power switch for automotive applications. It describes the functional and diagnosis requirements such as on-resistance and current sensing. Elaborate over-temperature, over-voltage, over-current, inductive clamping and loss-of-ground protections are described. The integration of these diagnosis and protection circuits allows the power switches to function in the harsh automotive and industrial environment.

Heterogeneous Integration of Passive Components for the Realization of RF-System-in-Packages

Eric Beyne, Walter De Raedt, Geert Carchon, and Philippe Soussan
IMEC, Kapeldreef 75, Leuven, 3001, Belgium

Abstract

Applications using rf radios operating at frequencies above 1 GHz are proliferating. The highest operating frequencies continue to increase and applications above 10 GHz and up to 77 GHz are already emerging. Systems become more complex and devices need to operate at several different frequency bands using different wireless standards. The rf-front end sections of these devices are characterized by a high diversity of components, in particular high precision passive components. In order to be produced cost-effectively, these elements need to be integrated along with the semiconductor devices. This paper describes the requirements for successful integration of rf-passive devices and proposes multilayer thin film technology as an effective rf-integration technology.

1. Introduction

As wireless communication devices are becoming ever more abundant in numbers and variety, high density system integration is becoming an increasingly important requirement. High density integration of rf-radio devices not only requires the integration of the active devices (rf-system-on-chip, rf-SOC), it also requires the integration of a large number of passive devices, such as transmission lines, resistors, capacitors and inductors, as well as functional blocs such as filters and baluns. In order to reduce the system size, as well as the system cost, a higher degree of miniaturisation is required. These components do not scale as well active IC-technology, making it difficult to integrate all these devices on-chip. Therefore; a proper portioning of the rf system is required. The active devices may be integrated in one or two SOC devices and the external passive devices should be integrated in the SOC package, effectively realizing an rf-System-in-a-Package, rf-SIP.

A key enabling technology for the realization of these rf-SIP ‘interposer’ substrates with integrated passives is the multilayer thin film technology as used for wafer-level-packaging, WLP, of device wafers (redistribution and bumping). A key feature of this technology is the use of photolithographic technology for

the definition of the various passive circuit components, resulting in a high degree in miniaturization and high patterning accuracy, with tolerances in the μm and sub- μm ranges. This results in an excellent circuit repeatability and predictability, key ingredients for the realization of first-time right and high manufacturing yield devices.

As transistor dimensions scale down and CMOS and Si-based semiconductors are increasingly replacing GaAs for microwave and mm-wave applications, circuit performance becomes increasingly determined by the on-chip passive component quality. However, in the attempt to pace up with this evolution, thinner on-chip metals and dielectrics have a troubling effect on the Q factor of on-chip passives. A cost-effective and attractive solution is to realize on-chip inductors using thin-film WLP techniques, similar to those used for realising the rf-SIP interposer substrates.

2. Multilayer Thin Film

A technology used to integrate passive components for rf-applications above 1GHz must allow for the realization of passive components with values relevant to those applications and with a high degree of precision and repeatability. Also, to allow for the integration, a high degree of scaling is required to fit the complex circuits in a small area.

These requirements strongly favor the use of a photolithographic defined technology, where a large number of devices are collectively realized on large substrates or wafers. We have proposed [1,2,3] the use of multilayer thin film for this purpose. The infrastructure for this technology was developed for the wafer-level-packaging, WLP, or silicon back-end-of-line processing. High volume manufacturing equipment with automatic handling is now available for the common Si-wafer sizes.

The basic elements of this technology are a thin-film, high density metallization technology and a thin-film, dielectric deposition technique, capable of realizing very small via holes in the isolation layers to allow for high density interconnects between the different layers in the structure.

Thin-film technology is well suited for the integration and miniaturization of passive components. Complex materials can be deposited with high repeatability to form the highest quality resistor or capacitor layers. The thin-film lithography assures a high dimensional accuracy, enabling small tolerances and increased miniaturization and, therefore, avoiding the need for “trimming” of resistor or capacitor values. The electroplated copper lines, described above,

are ideally suited for realizing high quality inductors, particularly those required for high frequency applications.

3. Integration of Passive Devices – Requirements

3.1 Resistors

The resistance of an integrated thin film resistor is given by:

$$R \approx \frac{\rho}{h} \times \frac{l}{w} \quad (1)$$

Where ρ is bulk resistance of the resistor material, h the film thickness and l and w , respectively, the resistor length and width.

From (1) it is clear that resistance scales well with reducing dimensions as it is basically proportional to l/w , commonly referred to as “the number of squares”. The ratio ρ/h is referred to as the sheet resistance ρ_{sq} . This value is defined by the material deposition and processing technology and can be accurately controlled during the production process. Thin film deposition techniques such as magnetron sputtering or physical vapor deposition (PVD) may result in highly repeatable and predictable resistive films.

The limits of resistor scaling are mainly defined by the maximum allowable current density. At high current densities self-heating of the resistor, non-ohmic behavior and several reliability problems may occur. Another important limit is the loss of resistance accuracy when scaling down resistors to too small dimensions. Resistance process tolerance is given by:

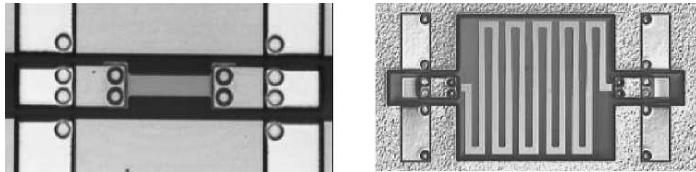
$$(\delta R)^2 = (\delta \rho_{sq})^2 + \Delta z^2 \left(\frac{1}{w^2} + \frac{1}{l^2} \right) \quad (2)$$

Where Δz is the patterning accuracy of the lateral dimensions l and w . When using photolithography, this value is typically smaller than 1 μm . The resistance accuracy for high precision applications is therefore dominated by the control of the resistance material composition and its uniform deposition process as sufficiently large resistor dimensions can be chosen.

For rf-front end applications, resistors are mainly used for matching and terminating transmission lines (range 10 – 100 Ohm) as well as for low frequency bias resistors (kOhm range). Bias resistors typically require relative rather than absolute precision. Therefore it is important to integrate a material with a relatively low sheet resistance. This will result in small rf-resistors with excellent rf-performance. Large value resistors are then realized as long,

meandering structures. This is more cost effective than integrating a second resistance material with a higher sheet resistance.

We choose to use a PVD TaN film with a sheet resistance 25 Ohm/ \square . This is a low resistance for such a film, but allows the use of a relatively thick layer, effectively improving the tolerance to deposition thickness variations. This material also exhibits a low temperature coefficient of resistance of about -100 ppm/ $^{\circ}\text{C}$. Examples of thin film resistors integrated in multilayer thin film technology are shown in figure 1.



*Figure 1 : Integrated TaN resistors for rf integration,
left 100 Ohm , right 4.8 kOhm.*

3.2 Capacitors

The capacitance of an integrated parallel plate thin film capacitor (metal-insulator-metal or MIM capacitor) is given by:

$$C \approx \frac{\varepsilon}{h} \times (l \times w) \quad (3)$$

Where ε is the insulator dielectric constant, h the dielectric thickness and l and w respectively the MIM capacitor plate length and width.

From (3) it is clear that MIM capacitors do not scale at all with reducing lithographic dimensions. The only scaling options are to increase the dielectric constant of the material or to decrease the thickness of the insulator layer. Reducing the layer thickness is limited by the electric breakdown of the insulating layer and increasing leakage currents through thin dielectric layers. These properties are highly material dependent and influenced by the choice of contact materials and surface roughness. Other important properties for rf-capacitors are there voltage linearity and temperature coefficient of capacitance. The capacitance process tolerance is given by:

$$(\delta C)^2 = \left(\delta \frac{\varepsilon}{h} \right)^2 + \Delta z^2 \left(\frac{1}{w^2} + \frac{1}{l^2} \right) \quad (4)$$

Where Δz is the patterning accuracy of lateral dimensions l and w . When using photolithography, this value is typically smaller than 1 μm . In practice, the

accuracy of rf-capacitors is defined by both the lithography precision and the dielectric deposition accuracy.

The high frequency characteristics of a capacitor are strongly influenced by its parasitic conductance (dielectric loss) as well as its parasitic resistance and inductance. Figure 2 illustrates the rf-characteristics of a typical MIM capacitor. In order to limit the impact of the series resistance and inductance on capacitor behavior, the size of the capacitor plates should be minimized. If they become too small, the patterning accuracy will dominate the capacitor precision (4).

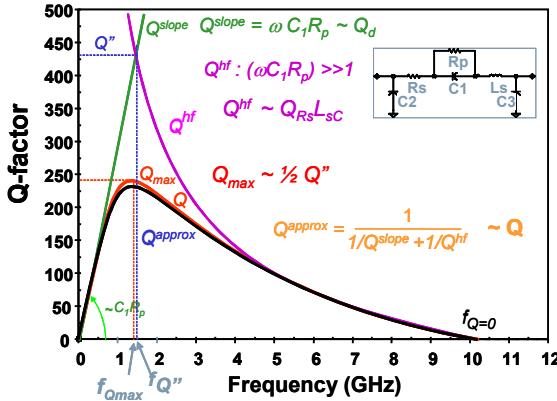


Figure 2: RF-quality factor of a typical MIM capacitor, illustrating the impact of capacitor parasitics.

Which capacitance density ε/h should be integrated for realizing rf-circuits? Capacitors are used for a large variety of applications. For low frequency decoupling $\sim nF$ value capacitors are required, for passive rf circuits, such as filters and couplers, much smaller values are required. These capacitors are used as reactive $Z_C=1/j2\pi f C$ impedances. The required values scale therefore down with increasing frequency. For 1 to 100 Ohm reactive impedance, the required C values range from 1 to a few 100 pF for frequencies between 1 and 10 GHz. When used in filter structures, even lower capacitor values are required.

It is clear that such wide range of requirements can not be met with a single material. When using a high density capacitor, realizing a small valued capacitor would require an extremely small capacitor size, resulting in a poor capacitance tolerance (4). When using a low capacitance density, the size of the capacitor may become very large, resulting in a high parasitic resistance and inductance, resulting in poor rf-performance. Therefore a compromise is required.

We use basically three types of capacitors. For the larger value capacitors of 1 pF to 1 nF, anodized Ta, Ta_2O_5 , is used as dielectric to realize capacitors with a density of 720 pF/mm². This technology also allows to realize capacitors with much higher density, up to e.g; 5 nF/mm² [4], however for rf-applications this results in limited applicability. For capacitors with smaller values of 0.5 pF – 2 pF, the low dielectric constant (2.7) multilayer dielectric BCB is used. The density of these capacitors is 6.5 pF/mm². The smaller capacitors, < 500 fF, are realized as interdigitated metal fingers.

3.3 Inductors –solenoid type

The inductance of a solenoid type inductor is approximately given by :

$$L \approx \mu \times N^2 \times \frac{A_c}{l_c} \approx \mu \times \frac{V_c}{P_w^2} \quad (5)$$

Where μ is the permittivity of the magnetic material, N is the number of windings, and A_c , l_c and V_c are, respectively, the cross section, length and volume of the magnetic core. P_w is the pitch of the windings around the magnetic core.

From (5) it is clear that, in first order, the coil inductance is proportional to the volume of magnetic material. Scaling down inductor sizes (size reduction while maintaining the nominal value) is therefore even more difficult than scaling capacitors. Integrating smaller sized inductors requires high permittivity materials and narrow conductor pitches. For rf-applications, this requires the use of ferrites with very low electric conductivity. These materials are difficult to integrate in thin film technology, especially as large magnetic volumes are still required. Another problem with these materials is the predictability, accuracy and non-linearity of the magnetic properties. Furthermore, the series resistance and parasitic capacitances of solenoid type inductors typically limit their application to low frequencies [5, 6].

3.4 Inductors –spiral type

Spiral-type inductors are generally used in rf-integration. The inductance value is a complex function of the geometry of the spiral inductor [5, 7]. Some simple analytical approximations can be found in literature. The general behavior of inductors is approximately given by Schieber's formula [8]:

$$L \approx 3.596 \cdot 10^{-5} \mu_0 N^2 \phi \quad (6)$$

Where ϕ is the diameter of the spiral inductor and N the number of windings.

Spiral inductors do not scale readily with area, but they do scale with the number of turns, which is basically limited by the trace-pitch. In order to realize high quality, low resistance inductors the conductor width should be maximized. Optimum inductors are therefore realized by minimizing the conductor spacing. The inductance process tolerance is only a weak function of the line width precision. Therefore, the precision and repeatability of spiral inductors is typically very high.

The high frequency characteristics of an inductor are strongly influenced by its parasitic resistance and capacitance. Figure 3 illustrates the quality factor behavior of a typical rf-inductor. At low frequency, the inductance is limited by the coil inductance. At high frequencies, the characteristics are dominated by the parasitic capacitance.

Large inductor values are typically not required for realizing rf-circuits. Inductors are typically used for impedance matching and resonating circuits. Typical values vary from 0.1 to 5 nH. Examples of inductors realized in multilayer thin film technology are given in figure 4.

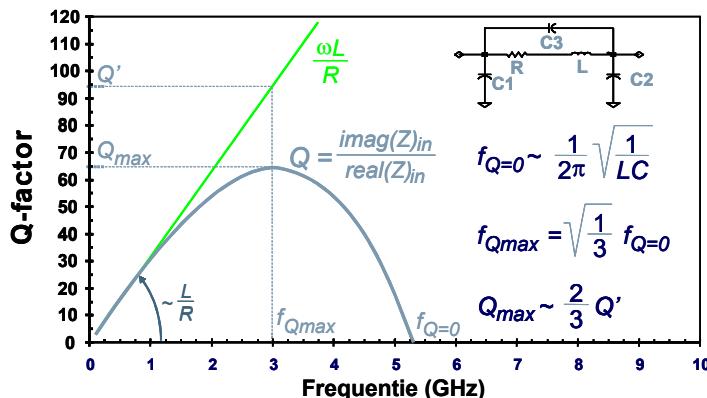


Figure 3: Inductor quality factor versus frequency for rf spiral inductors on high resistivity substrates.

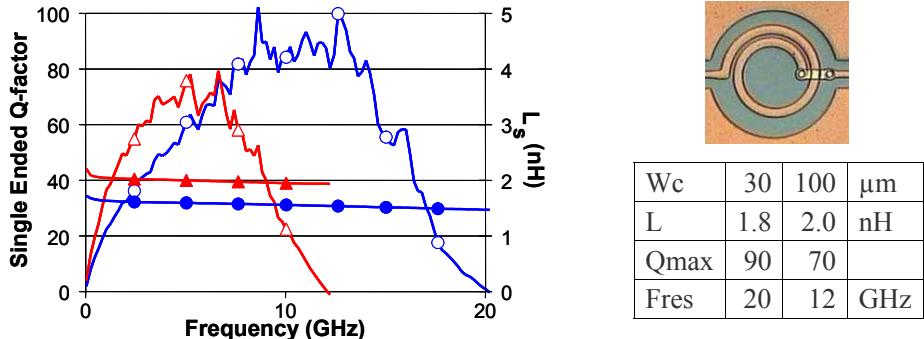


Figure 4: Measured inductance and quality factor versus frequency. Multilayer thin film, 1.5 turn inductors, realized on high resistivity Si substrates, with a line spacing of 10 μm , and a gap of 150 μm between the outer trace and the ground plane. The inner diameter of the 30 μm trace width inductor is 250 μm and that of the 100 μm trace width inductor is 300 μm .

4. Integrated Passive Device Platform

The R-L-C passives described above, as well as interconnect lines and other passives can be realized in a single multilayer thin film technology. An example of an rf-SIP build-up topology, developed at IMEC, is shown in figure 5. On a low loss rf-substrate (glass or high resistivity silicon) integrated resistors, capacitors and inductors are integrated. In order to allow for the practical use of this technology, a design library was developed. This library consists of electrically equivalent circuit models for all the relevant rf-passive circuits, as well as models for the interconnect lines, discontinuities and wire bond or flip-chip connections. These models are parametric with the main geometric dimensions, allowing for a flexible optimization of the rf-design. The library also automatically generates the mask layout for the circuit, further improving the predictability of the designed circuit. Complex filters, coupling structures and filter functions can be realized effectively using this approach. Some application examples are shown in figure 6.

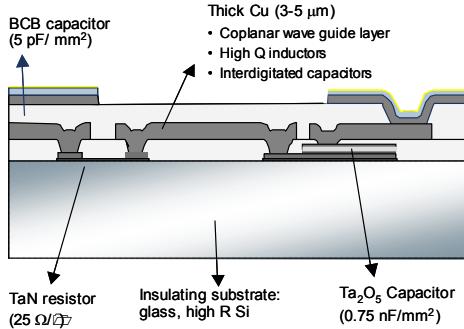


Figure 5: Schematic cross-section of IMEC's integrated passives platform technology.

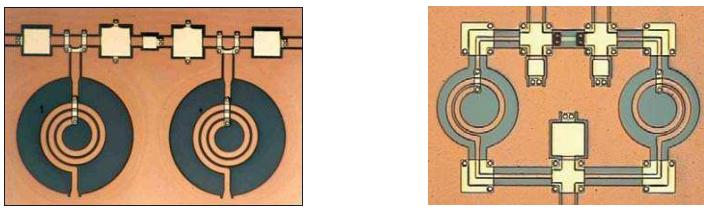


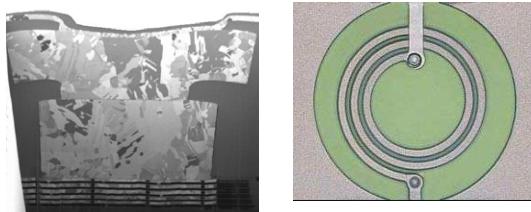
Figure 6: Examples of integrated rf functions in the rf-interposer technology.
Left: 2.45 GHz Bandpass Filter. Right: 7 GHz Wilkinson power splitter integrated on a high resistance Si substrate, rf-SIP.
Characteristics: Return loss < -20 dB, Isolation < -20 dB, Loss 0.5 dB, Size: 1.65×1.05mm².

5. “Above-IC” rf-SOC

A characteristic feature of rf front-end integrated circuits is the relatively large area occupied by on-chip inductors. The large size of these inductors is due to the physical limitations of scaling inductors while maintaining performance (Q-factor). Realizing these inductors on the rf-SIP interposer substrate, as described above, is generally not an option as the interconnect parasitics, even a small flip-chip bump, would degrade the performance improvement gained from using high Q off-chip inductors. However, the integrated passives technologies can also be applied directly on the device wafer, “above-IC” [9].

For many high frequency rf-IC’s, the poor quality factors of regular on-chip inductors is a limiting factor. This is due to the relatively high sheet resistance of

the on-chip metallization and the losses in the semi-conducting silicon substrate. By placing the spiral inductor in the thin film layer, “above-IC”, the distance between the spiral and the lossy substrate is greatly increased. By using a thicker, electroplated Cu conductor, a much lower track resistance is obtained. A FIB cross-section of such an inductor process is shown in figure 7. In this case a 10 μm thick copper layer and a 12 μm thick dielectric is used. Inductors with Q-factors above 30 up to 5 GHz were obtained over 10 $\mu\Omega\text{cm}$ Si CMOS wafers. The Q factor can even be increased by applying a ground shield on the silicon substrate. Also differential inductors with high quality factors and very high resonance frequencies can be realized, as shown in figure 8.



*Figure 7: High-Q, 10 μm thick Cu inductor processed on top of a 10 $\mu\Omega\text{cm}$ CMOS wafer.
Left : cross-section contact inductor; Right : top view.*

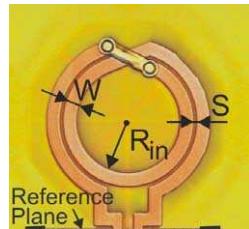
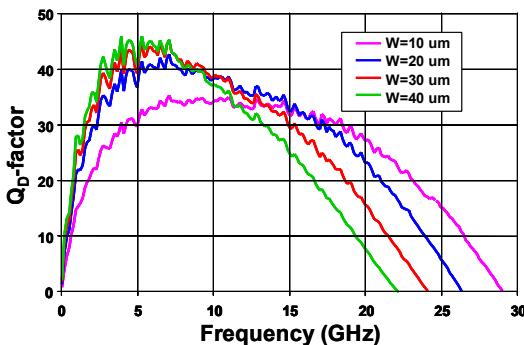


Figure 8: Differential 1.6nH “above-IC” inductors integrated on a 10 $\mu\Omega\text{cm}$ CMOS wafer.

Inductor parameters: 2 turns, line thickness 10 μm , line spacing of 10 μm , line widths ranging from 10 to 40 μm and inner diameter of 250 μm .

The post-processing is compatible with both Cu and Al back-end. The technology is cost-effective and consumes no additional Si real estate. Measurements performed on MOS transistors and back-end interconnects show no important performance shifts after post-processing. The WLP inductors have

increased performance and resonance frequency as compared to back-end versions enabling the design of high-performance low-power circuits such as VCOs.

IMEC applied this technology to realize a 5 GHz and 15 GHz low-power VCO in 90 nm CMOS [10]. The 5 GHz and 15 GHz VCOs use respectively a 3nH and a 0.6nH WLP inductor without ground shielding resulting in a differential Q factor of respectively 40 and 55. The 5 GHz and 15 GHz VCOs show a low core power consumption of respectively 0.33 mW and 2.76 mW, a phase-noise of –115 and –105 dBc/Hz (at 1 MHz offset) and a tuning range of 148 MHz and 469MHz. For comparison, a 6.3 GHz, 90nm VCO using a back-end inductor with patterned ground shield has a core power consumption of 5.9 mW with a phase noise of –118 dBc/Hz at 1 MHz.

6. Conclusion

Multilayer thin-film WLP technology with integrated passives can be applied both on active wafers (rf-SOC) as on an intermediate glass or high-resistivity Si substrates (rf-SIP). The results prove that thin-film technology allows integrating high-Q passives for wireless telecommunication applications covering a very broad frequency range, from the 1 to 5 GHz mobile-phone standards up to 77 GHz automotive radar.

References

- [1] E. Beyne, R. Van Hoof, A. Achen, “The use of BCB and Photo-BCB Dielectrics for High Speed Digital and Microwave applications”, Proceedings of the 1995 International Conference on Multichip Modules, Denver, Colorado, April 19-21, 1995; pp. 513-518.
- [2] Ph. Pieters, S. Brebels, E. Beyne, “Integrated Microwave filters in MCM-D”, Proceedings IEEE-MultiChip Module Conference MCMC-96, Santa Cruz, California, February 6-7, 1996.
- [3] G. Carchon, K. Vaesen, S. Brebels, W. De Raedt, E. Beyne, B. Nauwelaers, “Multilayer thin-film MCM-D for the integration of high-performance RF and microwave circuits”, IEEE Trans. Components and Packaging Technologies, Vol. **24** (3), 2001, pp. 510-519.
- [4] P. Soussan, L. Goux, M. Dehan, H.V. Meeren, G. Potoms, D. Wouters, E. Beyne, “Low Temperature Technology Options for Integrated Hight Density Capacitors”, Proceedings of the 2006 ECTC Conference, May 30 – June 2, 2006, pp.515-519.

- [5] S.F. Mahmoud, E. Beyne, “Inductance and quality-factor evaluation of planar lumped inductors in a multilayer configuration”, IEEE Transactions on Microwave Theory and Techniques, Vol. **45**, 6, June 1997, pp. 918 – 923.
- [6] W. Ruythooren, E. Beyne, J.-P. Celis, J. De Boeck, “Integrated high-frequency inductors using amorphous electrodeposited Co-P core”, IEEE Transactions on Magnetics, Vol. **38**, 5, Sept. 2002, pp. 3498 – 3500.
- [7] P. Pieters, K. Vaesen, S. Brebels, S. Mahmoud, W. De Raedt, E. Beyne, R. Mertens, “Accurate modelling of high Q-inductors in thin-film multilayer technology for wireless telecommunication applications”, IEEE Trans. MMT-S, Vol. **49**, (4), pp. 489-599, 2001.
- [8] D. Schieber, “On the inductance of printed spiral coils”, Archiv für Elektrotechnik, **68**, 1985, pp. 155-159.
- [9] G. Carchon, et al. “High-Q RF Inductors on Standard Silicon Realised using Wafer-level Packaging Techniques”, MMTS conference, Philadelphia, June 9-12 2003.
- [10] X. Sun, O. Dupuis, D. Linten, G. Carchon, P. Soussan, S. Decoutere, E. Beyne, “High-Q above-IC inductors using thin film wafer-level packaging, demonstrated on 90nm rf-CMOS 5 GHz VCO and 24 GHz LNA”, IEEE Transactions on Advanced Packaging Technology, part B, Vol. **29**, 4, Nov. 2006, pp. 810-817.

The Eye-RIS CMOS Vision System

Ángel Rodríguez-Vázquez, Rafael Domínguez-Castro, Francisco Jiménez-Garrido, Sergio Morillas, Juan Listán, Luis Alba, Cayetana Utrera, Servando Espejo and Rafael Romay

AnaFocus (Innovaciones Microelectrónicas S.L.)
Avda Isaac Newton, Pabellón de Italia, Planta 4
Parque Tecnológico Isla de la Cartuja
41092-Sevilla (SPAIN)
angel.rodriguez-vazquez@anafocus.com

Abstract

Eye-RIS is the name of a family of *vision systems* which are conceived for single-chip integration using CMOS technologies. The Eye-RIS systems employ a *bio-inspired* architecture where image acquisition and processing are truly intermingled and the processing itself is realized in two steps. At the first step processing is fully *parallel* owing to the concourse of dedicated circuit structures which are integrated close to the sensors. These circuit structures handle basically *analog information*. At the second step, processing is realized on digitally-coded information data by means of *digital processors*. Overall, the processing architecture resembles that of *natural* vision systems, where parallel processing is made at the *retina* (first layer) and significant reduction of the information happens as the signal travels from the retina up to the visual cortex. This chapter outlines the concept of the Eye-RIS system and its main components and presents experimental data to illustrate its practical operation.

1. Introduction

CMOS technologies enable *on-chip* embedding of optical sensors with data conversion and processing circuitry, thus making possible to incorporate *intelligence* into optical imagers and eventually to build *vision*¹ systems in the form of CMOS chips.

1. *Vision* is defined as the set of tasks to interpret the environment from the information contained in the light reflected by the objects in such environment. It involves signal acquisition, signal conditioning, and information extraction and processing. *Sensor intelligence* refers to the incorporation of processing capabilities into the sensor itself.

During the last years many companies have devised different types of CMOS sensors with different sensory-pixel types and different levels of intelligence. These CMOS devices are either targeted to replace CCDs in applications where smartness is important or to make optical sensing and eventually vision feasible for applications where compactness, power consumption and cost are important. Most of these smart CMOS optical sensors follow a conventional architecture where *sensing* is physically separated from *processing* and processing is realized by using either PCs or DSPs (see Figure 1). Some of the intelligence attributes embedded by these sensors are related to *signal conditioning* and typically include:

- Electronic shutter and exposure time control,
- Electronic image windowing,
- Black calibration and white balance,
- Fixed Pattern Noise (FPN) cancellation,
- Etc.

Other intelligence attributes are related to *information processing* itself and are supported by libraries and software to cover typical image processing functions and to allow users implement image processing algorithms.

Note that in the conventional architecture of Figure 1 most of the intelligence is far from the sensor. Hence all input data, most of which are *useless*, must be codified in digital form and processed. On the one hand, this fact stresses the system requirements regarding memory, computing resources, etc.; on the other, it causes a significant bottleneck in the data-flow.

Such way of processing is actually quite different from what is observed in *natural vision systems*, where processing happens already at the sensor (*the retina*), and the data are largely compressed as they travel from the retina up to

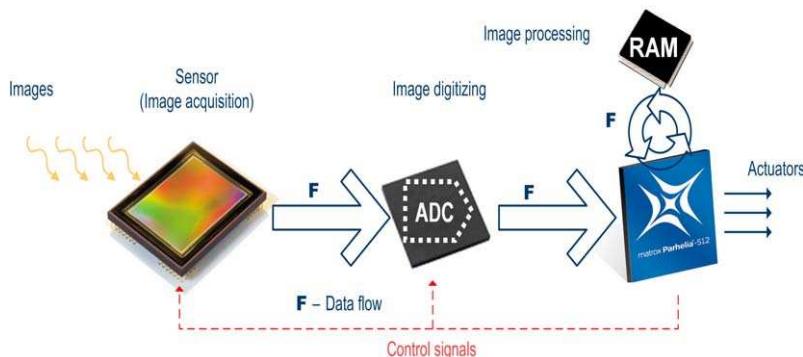


Figure 1: Conventional Smart Image Sensor Concept

the visual cortex. Also, processing in retinas is realized in *topographic* manner; i.e. through the concourse of structures which are *spatially distributed* into arrangements similar to those of the sensors and which operate *concurrently* with the sensors themselves.

These architectural concepts borrowed from nature, namely:

- realization of the processing tasks by an *early processing* step followed by a *post processing* step,
- incorporation of the early processing structures right at the sensor layer,
- concurrent sensing and processing operations through the usage of either topographic or quasi-topographic early-processing architectures.

define basic attributes of the Eye-RIS vision system.

Some of these attributes, particularly the splitting of processing into pre-processing and post-processing, are also encountered in the smart camera Inca 311 from Philips. This camera embeds a *digital pre-processing* stage based on the so-called Xetal processor [1]. As a main difference to this approach, in the Eye-RIS vision system pre-processing is realized by using *mixed-signal* circuits distributed in a pixel-wise area arrangement and embedded with the optical sensors. Because pre-processing operations are realized in truly parallel manner in the analog domain, power efficiency and processing speed of the Eye-RIS system are both very large. Specifically, the last generation sensor/pre-processor used at the Eye-RIS, the so called Q-Eye, exhibits a computational power of 250GOps¹ with a power consumption of 4mW per GOps.

2. The Eye-RIS Vision System Concept

Eye-RIS is a generic name used to denote the bio-inspired vision systems from AnaFocus. These systems are conceived for on-chip integration of all the structures needed for:

- *Capturing* (sensing) images
- *Enhancing* sensor operation, such as to enable high dynamic range acquisition
- Performing *spatial-temporal* processing
- *Extracting* and *interpreting* the information contained into images
- Supporting *decision-making* based on the outcome of that interpretation.

The Eye-RIS are *general-purpose*, fully-programmable hardware-software vision systems. They are complemented with a software layer and furnished with

1. GOps = Giga operations per second

a library of image processing functions which are the basic instructions for algorithm development.

Three generations of these system have been already devised (Eye-RIS v1.0, v1.1 and v1.2) in a road-map towards single chip implementation (Eye-RIS v2). All these generations follow the architectural concepts depicted in Figure 2.

The main difference between the concept in Figure 2 and the conventional one depicted in Figure 1 comes from the “*retina-like*” structure placed at the front-end in Figure 2. This “*retina-like*” front-end stage is conceptually depicted as a *multi-layer* one. In practice it is a *multi-functional* structure where all the conceptual layers depicted in Figure 2 are actually realized on a common semiconductor substrate. These functions include:

- 2-D image sensing.
- 2-D image processing. Programmable tasks in space (across the spatial pixel distribution) as well as in time are contemplated.
- 2-D memorization of both analog and digital data.
- 2-D data-dependent task scheduling.
- Control and timing.
- Addressing and buffering the core cells.
- Input/output.
- Storage of user-selectable instructions (programs) to control the execution of operation sequences.
- Storage of user-selectable programming parameter configurations.

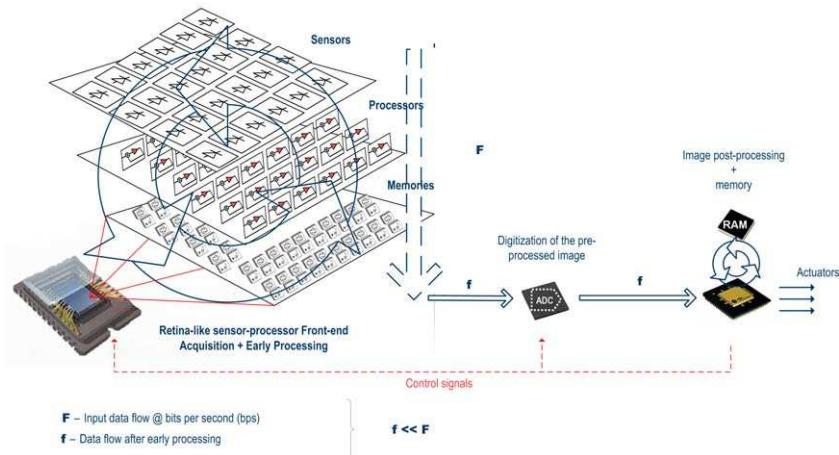


Figure 2: Eye-RIS system conceptual architecture with multi-functional retina-like front-end

Note from Figure 2 that the front-end largely reduces the amount of data (from **F** to **f**) which must first be codified into digital representations and then processed. At this early processing stage many useless data are hence discarded through processing and only the relevant ones are kept for ulterior processing. Quite on the contrary, in the conventional architecture of Figure 1 the whole data amount **F** must be codified and processed. This reduction of data supports the rationale for advantages of the Eye-RIS vision system architecture.

In order to quantify the advantages let us calculate the *latency time* needed for the system to react in response to an event happening in the image. In the case of Figure 1,

$$t|_{\text{LAT}}^{\text{conv}} = t_{\text{acq}} + N \times R \times t_{\text{A/D}} + N \times t_{\text{comp}} + t_{\text{proc}}$$

Where N is the number of pixels in the image, R is the number of bits employed for coding each pixel value, t_{acq} is the time required for the sensors to acquire the input scene, $t_{\text{A/D}}$ is the per-bit conversion time, t_{comp} is the time needed to compare the new image with a previous one stored in memory as to detect any change, and t_{proc} is the time needed to understand the nature of the change and hence prompt a reaction. Although the exact value of the latency time above may significantly change from one case to another, let us consider for reference purposes that most conventional systems produce values in the range of 1/30 to 1/50 sec.

Similar calculations for Figure 2 yield,

$$t|_{\text{LAT}}^{\text{Eye-RIS}} = t_{\text{acq}} + M \times R \times t_{\text{A/D}} + t_{\text{comp}} + t_{\text{proc}}$$

Where M denotes the reduced number of data obtained after early processing. Comparing the latency times for figures yield,

$$t|_{\text{LAT}}^{\text{conv}} - t|_{\text{LAT}}^{\text{Eye-RIS}} = (N - M) \times R \times t_{\text{A/D}} + (N - 1) \times t_{\text{comp}} .$$

Since typically it is $N \gg 1$, the equation above can be simplified as,

$$t|_{\text{LAT}}^{\text{conv}} - t|_{\text{LAT}}^{\text{Eye-RIS}} \approx (N - M) \times R \times t_{\text{A/D}} + N \times t_{\text{comp}}$$

Also, since in most cases the number of changes to be tracked will be small, we can assume that $M \ll N$ and hence further simplify the equation above into,

$$t|_{\text{LAT}}^{\text{conv}} - t|_{\text{LAT}}^{\text{Eye-RIS}} \approx N \times (R \times t_{\text{A/D}} + t_{\text{comp}})$$

It shows that the time saving reported by the Eye-RIS system approximately equals the sum of the time needed for conversion and the time needed in the conventional architecture for pixel-wise comparison. This time saving enables Eye-RIS system be employed for applications that require on-line operation with rapidly changing scenes¹.

Besides on-line operation, the data reduction featured by the Eye-RIS system relaxes the computational demands on the post-processing structures and hence the complexity and power consumption of these structures.

Figure 3 shows a conceptual block diagram of one instance of the so-called Eye-RIS system family, namely the Eye-RIS v1.2. Its basic functional features include:

- **Early processing:** Front-end, retina-like sensor-processor. Particularly, current Eye-RIS systems employ the so-called Q-Eye front-end sensor-processor which will be described in the next section.
- **Post processing:**
 - 32-bit RISC uP at 70MHz – realized on a FPGA.
 - 32Mb SDRAM for program and image/data storage and 4Mb Flash for FPGA configuration.
- **Interfaces:**
 - 2 Serial Programming Interface (SPI).
 - UART: General purpose RS232 port.

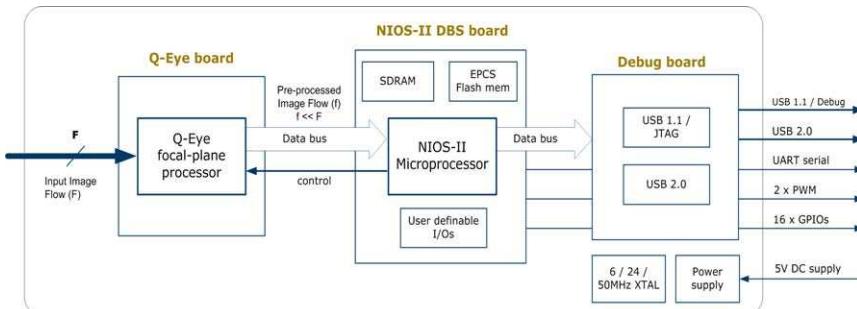


Figure 3: Conceptual block diagram of the Eye-RIS v1.2

-
1. In a traffic collision at 80Km/h, the time lag for a standard driver to hit the steering wheel is 12,5m/s. In order to properly estimate the distance from the driver to the wheel as to control the triggering of the air-bag, vision systems must acquire-and-process at rates higher than 500 frames/sec which is difficult to achieve with conventional systems but which are intrinsic to the operation of the Eye-RIS vision systems.

- 16 general-purpose 3.3V TTL I/Os.
- USB 1.1 Interface for JTAG control.
- USB 2.0 Interface for high-speed image I/O.

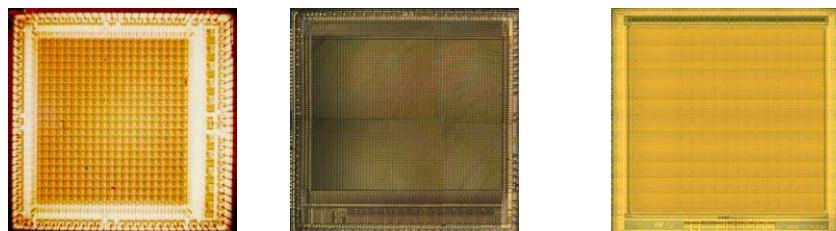
▪ **System tools:**

- Eye-RIS ADK (Application Development Kit), an Eclipse-based development environment including all the tools needed for programming Eye-RIS, namely: project manager, code editor, C/C++ compiler, assembler and linker, source-level debugger and etc.
- Image-processing library including basic routines such as: point-to-point operations, spatio-temporal filtering operations, morphological operations, statistical operations, blob analysis, etc.

3. The Retina-Like Front-End: From ACE Chips to the Q-Eye

A key component of the Eye-RIS vision system is the *retina-like front-end* which combines signal acquisition and processing embedded on the same physical structure. Since the late nineties several versions of this key component have been devised as stand-alone chips by researchers at the Institute of Microelectronics of Seville belonging to the Spanish Council of Research – called ACE and CACE chips. The principles, concepts and circuits of these chips are very well described in a couple of monographs from Dr. Ricardo Carmona [2] and Dr. Gustavo Liñán [3], respectively.

The most representative among ACE chips instances are collected in Figure 4.



ACE400 [4]

22 x 20 B&W cells

27.5 cells/mm²

Pixel Memory: 4 B&W

Program Memory: 8 Templates

15.8 GOPs

IO: 22Lines Binary

ACE4K [5]

64 x 64 Gray-Scale cells

82 cells/mm²

Pixel Memory: 4 B&W + 4 G

Prog. Mem.: 32 Templates +

64 Dig. Instructions

40 GOPs

IO: 16bit B&W Bus+ 16lines G Bus

ACE16K [6]

128 x 128 Gray-Scale cells

180 cells/mm²

Pixel Memory: 2 B&W + 8 G

Prog. Mem.: 32 Templates +

4096 Dig. Inst.

190 GOPs

IO: 32bit Digital Bus

Embedded Data Converters per Column

Figure 4: The three generations of ACE chips

There three micro-photographs are included and data pertaining to each of the corresponding chips are presented. These data show a progressive increase of spatial resolution, cell density and embedded functionality from the so-called ACE400 chip [4] to the so-called ACE16k [6] chip.

Figure 5(a) shows the block diagram of a sensory-processing pixel in the ACE16k chip [6]. It is seen that this pixel embeds much more functions than those found in conventional imagers. These latter contain only the optical module (the structure embedding the optical sensors) and the I/O module. However, the pixels of retina-like vision front-ends contain also structures to *store* information locally, to support *interactions* among pixels and to perform different types of *processing tasks*.

A consequence of the multi-functional pixel is that pixels have large area (i.e. the number of pixels per mm^2 is small – see Figure 4) with only a portion of the total area available for light sensing. Hence:

- On the one hand, the *fill factor* decreases. In practice the impact of this can be largely attenuated by using *microlenses* on top of the chip, since most of the light received by a pixel is focused onto its active sensing area.
- On the other hand, since the total silicon area is constrained due to reliability considerations, the *spatial resolution* is constrained as well.

However, limited spatial resolution may not be a problem if the system is intended for high-level tasks such as segmentation, object detection, etc. This is illustrated in Figure 6.

The top pictures in Figure 6 shows the faces of *Lena* with decreasing resolutions. Even with the lowest resolution, *Lena* will most probably be recognized by

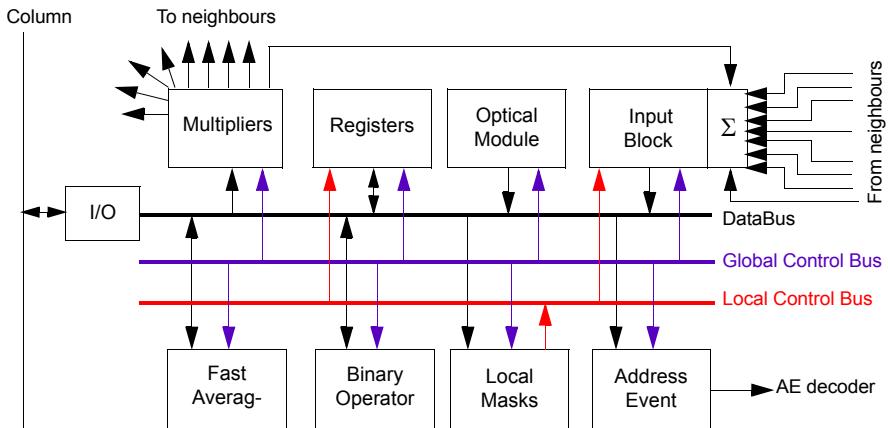


Figure 5: ACE16k cell block diagrams [6]

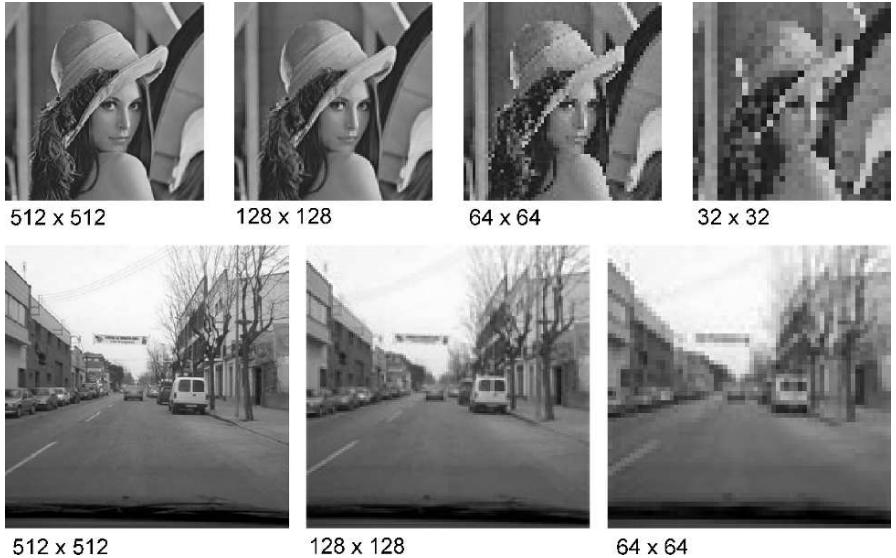


Figure 6: Illustrating the impact of resolution decrease on recognition

readers familiar with image processing. The bottom figure in Figure 6 shows a picture corresponding to a traffic scenario with different spatial resolution values. It is seen that main attributes of the scenario can be recognized even with low pixel counts. In both cases the information needed to recognize relevant attributes emerges from the geometrical features of the image, most of which are kept when the number of pixels decrease.

This qualitative reasoning suggests that there are many applications where spatial resolution can be traded by system efficiency in terms of operation speed, power consumption and system compactness.

As Figure 5 shows the processing tasks realized by retina-like front-ends and in particular by the ACE chips, involve interactions among pixels. Asides from the optical module, the most relevant analog design challenges for ACE chips, and in general for retina-like chips, are related to storage and to the realization of *linear convolution* operations with programmable kernels.

In a linear convolution operation a kernel of weights is multiplied by each image spatial sample and its neighbourhood in a small region, the results summed, and the outcome used to change the sample [7]. In the case of ACE and CACE chips two type of kernels per pixel are employed in accordance with the so-called CNN paradigm [8]. One kernel is applied to *input* data (they can be either input image samples or data stored in the registers at the pixels) while the other is applied to

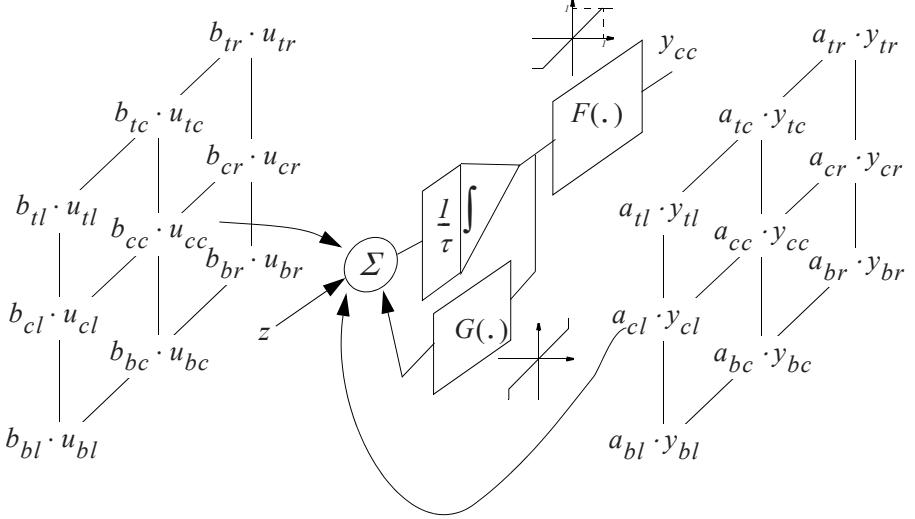


Figure 7: Block diagram of a cell belonging to two coupled CNN layers

state variables. The convolution operations realized in ACE and CACE chips are illustrated at conceptual level in Figure 7(a) where inputs and state variables are denoted as x and y respectively.

Convolution multiplications in ACE and CACE chips are realized by using tunable *transconductors*. To save silicon area, each transconductor employs a single transistor operating in ohmic regime [9]. Calibration and robust control of the operation of these transconductors define the most significant analog design challenges of ACE and CACE chips [2] [3] and cause the most significant limitations to their practical operation. These limitations are overcame with the so-called Q-Eye chip.

4. The Q-Eye Chip

A redesign of the ACE16k chip, the so-called ACE16K-v2, was employed to build the first two generations of the Eye-RIS system. These first two generations are referred to as Eye-RIS v1.0, v1.1 respectively. During the last two years a new chip, called Q-Eye, was devised which is employed at the front-end of the last generation Eye-RIS, namely the so-called Eye-RIS v1.2.

The Q-Eye chip significantly differs from the ACE chips. Main drawbacks of the ACE chips are lack of robustness, large power consumption and reduced cell den-

sity. The Q-Eye chip overcomes these drawbacks by making significant changes at both architectural and circuital design level. At the outcome, the cell density is increased by about 6.5 times and the power consumption is largely reduced. Also, these improvements are not detrimental of the functionality embedded per pixel; rather on the contrary, the Q-Eye includes at pixel level co-processing structures which are not found in the ACE chips.

Figure 8 is a block diagram of a pixel of the Q-Eye chip. As for the ACE chips basic analog processing operations among pixels are *linear convolutions* with programmable masks. However, the Q-Eye does not employ transconductance

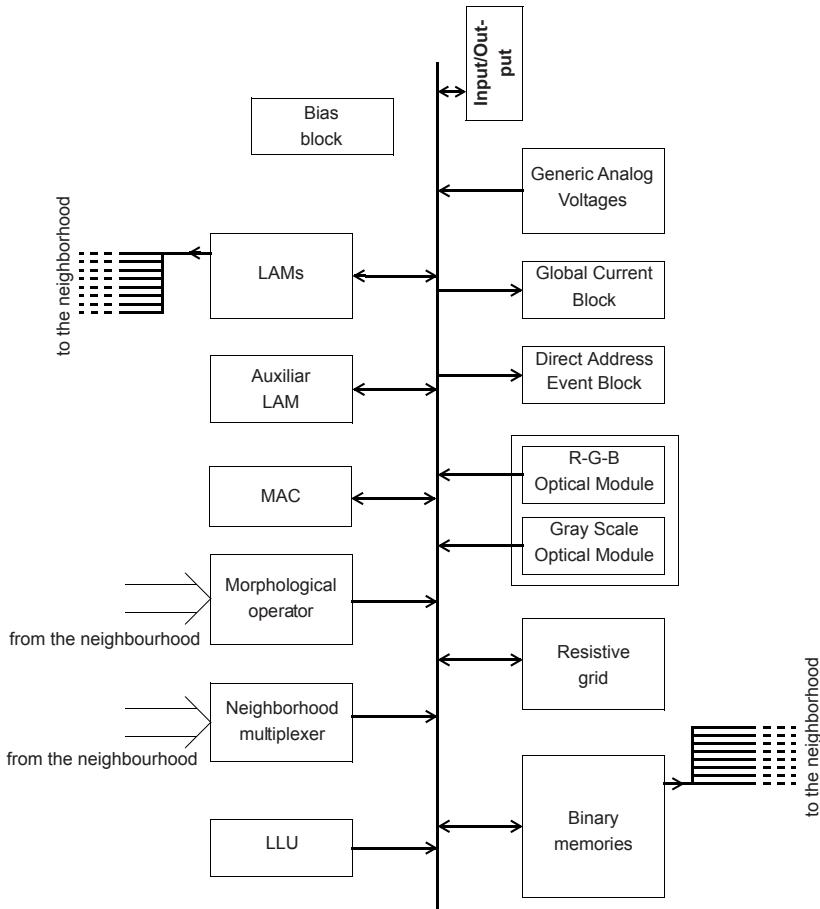


Figure 8: Block diagram for Q-eye pixel

multipliers (as it happens in the ACE chips) but a Multiplier-Accumulator Circuit unit (MAC) which processes neighbour pixels into an *algorithmic sequence*. Despite this sequential operation, computation times are similar to those obtained for ACE chips since no calibration of the transconductors is needed [9].

The area saving reported by the absence of spatially replicated structures (i.e. the transconductance multipliers employed for the linear convolutions) enables the incorporation at the Q-Eye pixel of functions which are not found in the ACE chips. These include:

- A pattern matching block to perform fast *morphological* operations on binary images and which complements the Local Logic Unit already found in the ACE chips.
- An additional bank of analog memories to allow: 1) shifting of grey-scale and binary images through an analog multiplexer and 2) swapping between analog memories.
- A circuitry for analog thresholding.
- Three additional sensors for *colour* RGB sensing.

Besides these functions the Q-Eye array includes a *resistive grid* with programmable diffusion time.

Figure 9 shows a floor-plan of the Q-Eye chip. The external interface of the Q-Eye is completely digital and synchronous. It is composed of a 32-bit data bus for image I/O and two additional buses, namely a 10-bits data bus and 12-bits address bus. These latter buses are employed to program a digital control system which contains 256 control words of 60-bit and individual register for analog references and miscellaneous configuration. This system controls the array of processing-sensing cells, on the one hand, and the I/O control unit which handles all basic I/O process, on the other. The I/O interface can operate in three modes:

- loading-downloading of grey-scale images,
- loading-downloading of binary images and
- address-event mode.

Grey-scale values are coded into digital form by on chip 8-bits AD Flash converters, and decoded by on chip 8-bits DA resistors string converters.

To the purpose of improved *power management*, and hence reduced power consumption, most of the processing blocks in the Q-Eye and the analog reference buffers used for biasing have independent power up/down signals. Also, the operation speed of most blocks is programmable. Thus, the chip can be tuned to process either very high frame rates or low frame rates with optimum power consumption for each configuration.

Robustness enhancement is achieved through improved calibration techniques. In the ACE16k chip, offsets were stored into analog memories which experienced

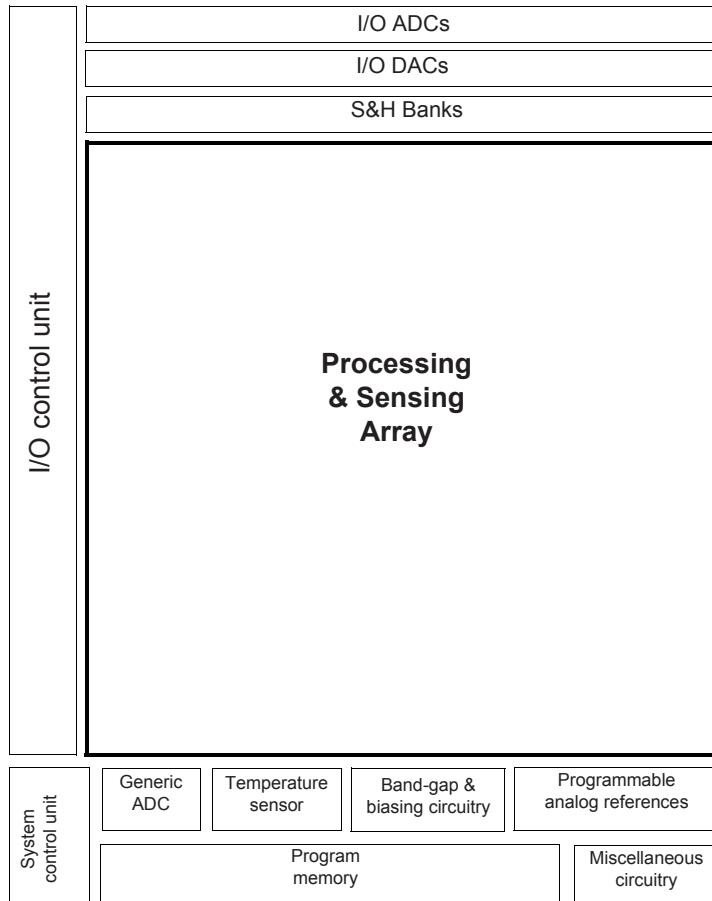


Figure 9: Block diagram of the Q-Eye chip

significant degradation specially at high temperatures. Instead, offsets in the Q-Eye chip are stored in static (non volatile) digital memories. Automatic calibration in the Q-Eye is performed by dedicated state machines which control in-loop A/D converters. Also, a temperature sensor and a temperature controlled correction loop are embedded in the Q-Eye to preclude the impact of junction temperature increases onto optical sensors and analog memories.

Main features of the Q-Eye chip are listed below:

- 176 x 144 cell array with $29.1\mu\text{m} \times 29.1\mu\text{m}$ area per cell (cell density of 1,180 cells/mm²).

- Monochrome or color RGB (1 multi-mode grey-scale pixel plus 3 RGB pixels per cell).
- High-speed non-rolling electronic shutter. Programmable exposure time (controlling step-down to 20ns).
- Sensitivity above 1.0V/lux-sec at 550nm (with microlenses).
- Fill factor above 50% (with microlenses).
- Frame rate above 10,000 fps.
- 4 + 1 (two banks) high-retention analog and 4 binary memories.
- Analog multiplexer for image shifting.
- Analog MAC unit.
- Programmable, 3×3 neighborhood pattern matching with 1/0/d.n.c. pattern definition (fast morphological functions).
- Programmable local logical unit (LUT table).
- Resistive grid for controllable image smoothing.
- 37.5mm² die area.
- 0.18μm 1.8V (core), 3.3V (I/O) CMOS technology.
- 50MHz clock frequency.
- < 100mW typical power consumption with 300mW peak during grey-scale image I/Os.
- Binary and analog image I/Os.
- On-chip bank of 4 ADCs and 4 DACs (8-bit@50MHz) for grey-scale image I/Os.

5. The Eye-RIS System in Operation

The Eye-RIS vision systems are conceived to enable vision for applications where *compactness*, *cost*, energy *consumption efficiency* and *operation speed* define major targets. Since the Eye-RIS systems are general-purpose platforms they can be software-programmed for a large variety of applications including:

- **Smart Surveillance.** Eye-RIS systems can be used to create distributed-intelligence camera networks. By moving the image processing functions towards each individual camera, scalability problems found in surveillance networks can be avoided. The Eye-RIS system can be programmed to perform: people tracking and counting, abandoned objects detection, forbidden areas monitoring, etc. This functionality will reduce the communication bandwidth and video storage demands, since only pre-specified activity patterns will get into the data stream.
- **Industrial Inspection.** The Eye-RIS systems take full advantage of the retina-like image processing to implement ultra-high speed quality-control and classification tasks that will enhance the speed of industrial production lines.

- **Automotive.** On-chip releases of AnaFocus Eye-RIS vision systems can be used to create ultra-compact intelligent vision systems. These vision systems will boost the efficiency of security systems in the automotive industry, matching in- and off-cabin applications, such as detecting the use of security belts, sleepiness alert, collision threat warning, etc.
- **Military.** Eye-RIS systems can be programmed to capture High Dynamic Range (HDR) images. Such HDR images can be afterwards processed on-chip to recognize and track high-speed moving objects in the scene under very aggressive illumination conditions.
- **Toy Industry.** Eye-RIS vision systems on-chip can be programmed to develop a variety of consumer-oriented entertainment applications. One example is the development of sophisticated human-computer interfaces based on gesture recognition that can be inexpensively incorporated to electronic toys and other consumer electronic devices.

Demonstrations representative of the use of the Eye-RIS systems for these application are found at the AnaFocus web page (www.anafocus.com). Figure 10 and Figure 11 illustrate the operation of the retina-like front-end, the Q-Eye chip. The two set of pictures at the top in Figure 10 show two different high dynamic range (HDR) images acquired in linear integration mode and HDR mode, respectively. In the latter case, HDR acquisition is achieved by processing right at the pixel level utilizing an algorithm based on the *well capacity adjustment* technique [12].

Images (c) to (h) in Figure 10 show inputs and outcome, respectively for different linear and nonlinear diffusion processes realized by using the on-chip embedded resistive grid whose parameters (mainly the spatial band-width of the diffusion process) can be controlled by the user. The results of performing low-pass, high-pass and band-pass spatial filtering on the input image Figure 10(c) are shown in Figure 10(d) (e) and (f), respectively. Figure 10(g) shows the output of a masked diffusion process (bottom figure) the mask being the binary figure at the top.

Figure 11 illustrate on the multi-functional capabilities of the Q-Eye chip by showing the input and output sequence of a Sobel filtering process (Figure 11(a)) and the input and outputs of the extraction of geometrical features from a binary image. Specifically, from left to right: the input binary image, the result of eliminating the single, isolated points, the borders of the latter image and the centroids of the same image. Note that all the described operations are realized directly at the pixel level, and simultaneously in all the pixels of an image, leading to extremely low processing times and consumed power.

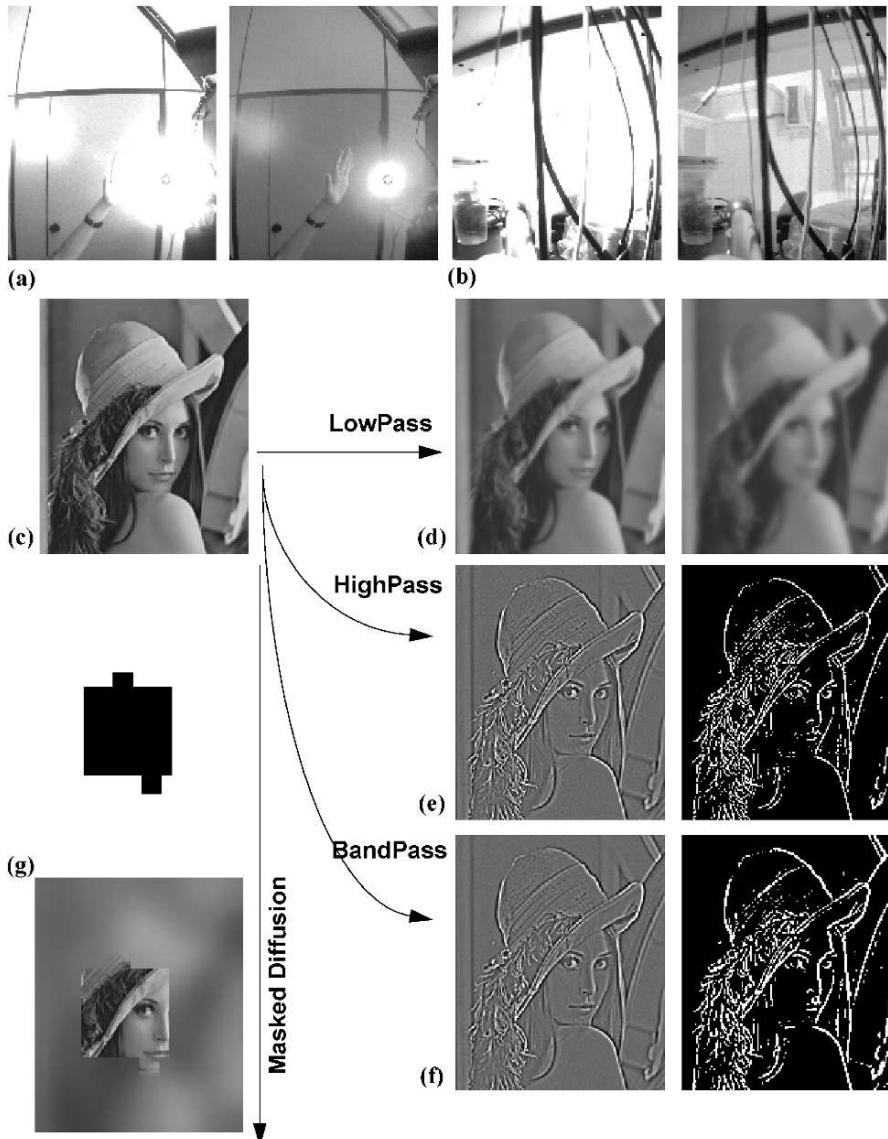


Figure 10: The Q-Eye chip in operation: (a) (b) Capturing HDR images in linear and HDR modes; (c) Input image for diffusive spatial filtering; (d) Outcome of lowpass filtering; (e) Outcome of highpass filtering; (f) Outcome of bandpass filtering; (g) Binary mask and outcome of a masked diffusion

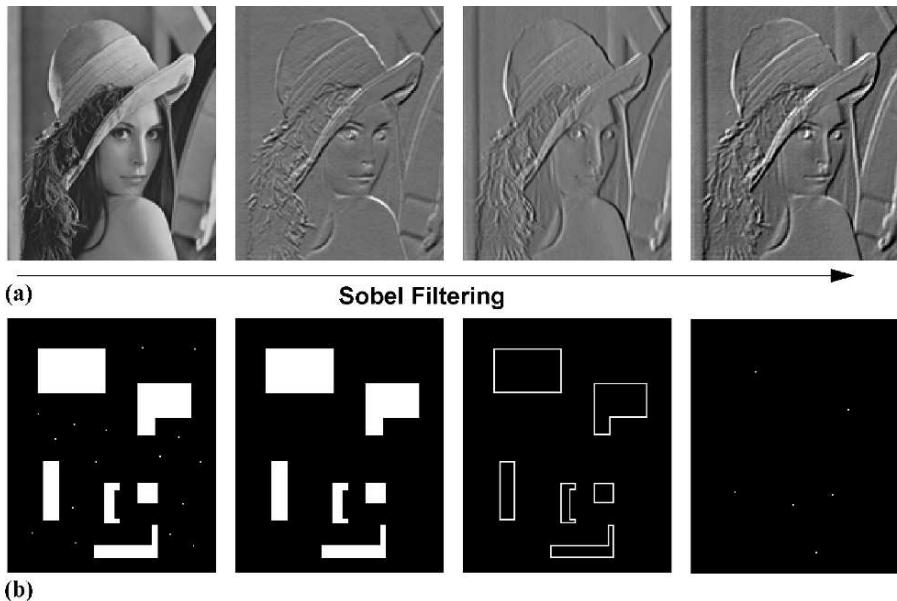


Figure 11: The Q-Eye chip in operation: (a) Input and outcome of a Sobel filtering process; (b) Extraction of geometrical features from a binary image

6. Conclusion

While vision in living beings handles a significant percentage of the information needed to interact with the environment, the use of vision in machines is limited due to very high cost/performance ratio. The Eye-RIS vision system overcomes this drawback by employing a bio-inspired architecture which can be realized at low cost in single chip form and which is capable of high-speed on-line operation.

7. Acknowledgements

Authors would like to acknowledge the many constructive discussions with Dr. Gustavo Liñán and Dr. Ricardo Carmona from IMSE-CNM.

References

- [1] R.P. Kleihorst, A.A. Abbo, A. van der Avoird, M.J.R. Op de Beeck, L. Sevat, P. Wielage, R. van Veen, H. van Herten, “Xetal: A Low-power High-

- Performance Smart Camera Processor". *Proc. of the 2001 IEEE Int. Symposium on Circuits and Systems*, Vol. 5, pp. 215 - 218, 2001.
- [2] R. Carmona Galán, "Analysis and Design of Mixed-Signal Chips for Real-Time Image Processing". PhD dissertation, University of Seville, 2002 (in english).
- [3] G. Liñán Cembrano, "Design of Programmable Mixed-Signal Low-Power Consumption Chips for Vision Systems". PhD dissertation, University of Seville, 2002 (in english).
- [4] R. Domínguez-Castro, S. Espejo, A. Rodríguez-Vázquez, R. Carmona, Péter Földesy, Ákos Zarányi, Péter Szolgay, Tamás Szirányi and Tamás Roska, "A 0.8 μ m CMOS 2-D Programmable Mixed-Signal Focal-Plane Array Processor with On-Chip Binary Imaging and Instructions Storage". *IEEE J. Solid-State Circuits*, pp. 1013-1026, IEEE July 1997.
- [5] G. Liñán, S. Espejo, R. Domínguez-Castro and A. Rodríguez-Vázquez, "ACE4k: An Analog I/O 64x64 Visual Microprocessor Chip with 7-bit Analog Accuracy". *Int. Journal of Circuit Theory and Applications*, Vol. 30, pp. 89-116, Wiley March-June 2002.
- [6] G. Liñán, A. Rodríguez-Vázquez, R. Carmona, F. Jiménez-Garrido, S. Espejo and R. Domínguez-Castro, "A 1000 FPS at 128 x 128 Vision Processor UIT 8-Bit Digitized I/O". *IEEE Journal of Solid-State Circuits*, Vol. 39, pp. 1044-1055, IEEE July 2004.
- [7] J.C. Russ, *The Image Processing Handbook*. CRC Press 1992.
- [8] L.O. Chua and T. Roska, *Cellular Neural Networks and Visual Computing*. Cambridge University Press 2002.
- [9] A. Rodríguez-Vázquez, G. Liñán, S. Espejo and R. Domínguez-Castro, "Mismatch-Induced Trade-Offs and Scalability of Analog Preprocessing Visual Microprocessor". *Analog Integrated Circuits and Signal Processing*, Vol. 37, pp. 73-83, Kluwer Academics 2003.
- [10] Servando Espejo, Rafael Domínguez-Castro and Angel Rodríguez-Vázquez, "Vision System on a Chip". *Elektronik*, pp. 54-59, May 2005.
- [11] A. Rodríguez-Vázquez, *From Photons to Decisions: The CMOS Challenge*. Internal Notes, August 2006.
- [12] S.J. Decker, R.D. McGrath, K. Brehmer, and C.G. Sodini, "A 256x256 CMOS Imaging Array with Wide Dynamic Range Pixels and Column-parallel Digital Output". *IEEE J. of Solid State Circuits*, Vol. 33, pp. 2081-2091, Dec. 1998.

AN INDUCTIVE POSITION SENSOR ASIC

Petr Kamenicky, Pavel Horsky

AMI Semiconductor, Czech Republic

petr_kamenicky@amis.com, pavel_horsky@amis.com

Co:
Henning Irle

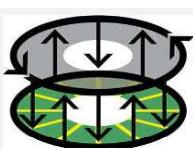
Hella KG Hueck & Co, Germany

Abstract

A principle and realization of an ASIC for a fast, precise, and high resolution Contact-less Inductive Position Sensor (CIPOS), is described.

The inductive contact-less sensor is suitable for angular or linear position sensing. Advantage of used measurement principle is, that it is strictly ratio metric independent of absolute values. The mechanical robustness, insensitivity to temperature variation, electrical or magnetic fields as well as its mechanical tolerances make the sensor ideal for harsh automotive environment. The sensor fulfils all the requirements for safety relevant applications. The presented ASIC includes sensor excitation, precise input signal analog and digital processing as well as the output signal transmitter. The signal path and main blocks are described.

1. Introduction



CIPOS logo

Today the automotive industry demands position sensors on many different places, e.g. accelerator pedal sensors, turbocharger actuator systems, steering angle sensors, throttle body position sensors, gear box control position sensors, head lamp position sensors, etc. Traditionally these sensors were equipped by potentiometers, which have all the reliability disadvantages of mechanical contact sensors (wear out, open contact etc.). In the

modern cars these potentiometers are replaced by electronic contact-less sensors which are based on different principles: Hall sensors, magneto-resistive sensors or inductive sensors.

The main advantages of inductive sensors for automotive applications are their mechanical variability (linear, rotational, arc segment, axis can go across the sensor,...), high temperature range, simplicity and robustness.

This paper describes an inductive positioning sensor system, consisting of a sensor, developed by Hella KG Hueck & Co, and a sensor ASIC, developed by AMI Semiconductor.

2. Inductive sensor principle

The inductive contact-less sensor consists of a stator and a rotor or cursor. (Fig. 2, 3, 4) The stator contains an excitation coil (which is part of an LC resonance network), receiver coils and electronics for signal processing. The rotor or cursor is a passive element with one closed winding which is designed in a special geometry (Fig. 2, 3, 4).

The inductive coupling between the excitation coil, the rotor or cursor and the three receiver coils leads to 3 ratio-metric signals that are dependent on the angular rotor or linear cursor position (Fig. 3 & 4). By the means of analog and digital signal processing, a signal, linearly dependent on the rotor/cursor position is extracted.

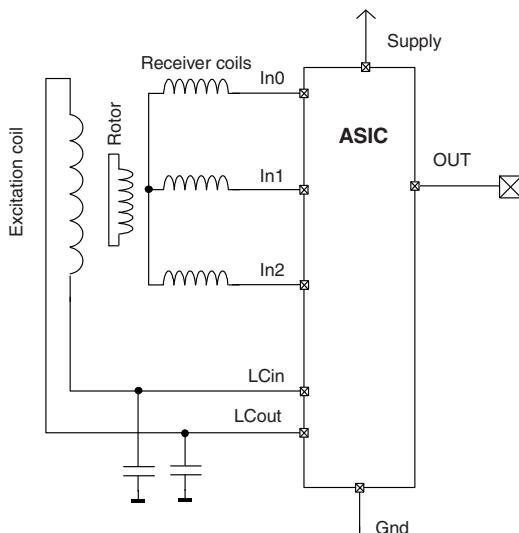


Fig. 1 Simplified application schematic

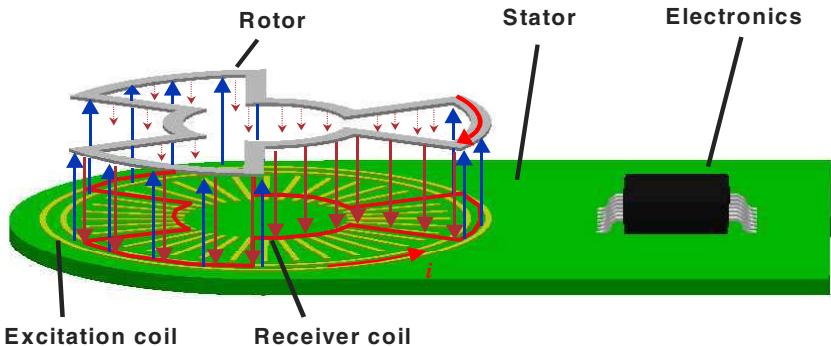


Fig. 2 Rotational sensor principle

An application schematic is shown in Fig. 1 and inductive rotational sensor principle is shown in Fig. 2.

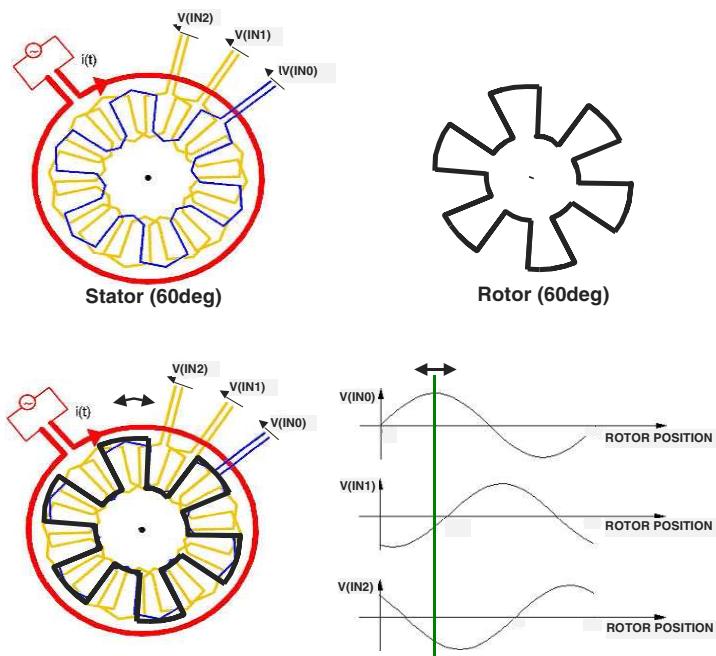


Fig. 3 Rotation sensor - output signals

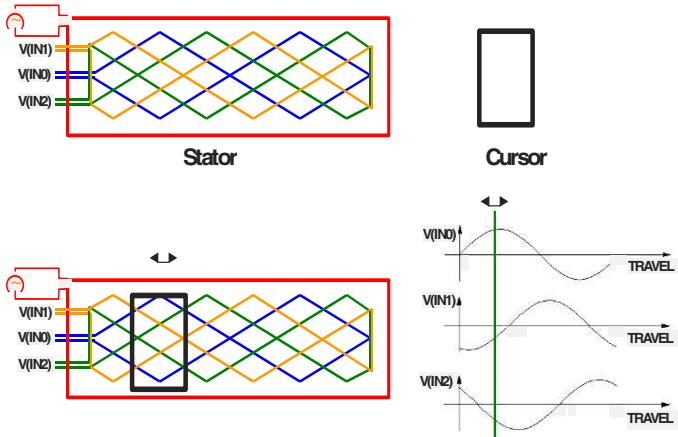


Fig. 4 Linear sensor - output signals

Different rotor/stator constructions allow to build 360, 180, 120, 90, 60,... degree rotational sensors to adjust the sensor to the required angle range.

3. System description

The block diagram of the inductive sensor positioning ASIC, responsible for all the necessary signal processing, is shown in Fig. 5. The signal path is highlighted.

The LC oscillator driver [4] is used to drive the excitation coil of the sensor. The coil has to be driven with a harmonic current and to minimize the current consumption of the coil driver, the coil is a part of the LC resonance circuit and the driver has only to deliver the losses in the resonance circuit. The LC oscillator is further described in chapter “Oscillator driver”.

Outputs of the sensor coils are IN0, IN1 and IN2 signals. Their amplitude depends on the rotor position. The IN0, IN1 and IN2 signals are first band limited by the Low Pass EMC filters to limit the high superposed EMC voltage on all INx pins to ensure low EMC susceptibility. Next, controlled by the digital, the proper combinations of input signals INx are selected in the analog multiplexer.

To create a DC signal from the AC input signal, a synchronous rectifier is used. The gain of the rectifier is a cosine function of the phase shift. Hence the phase shift between the Clock and the signal on the rectifier input is crucial as well as the Clock duty cycle. A very fast clock comparator with short, symmetrical delays is used.

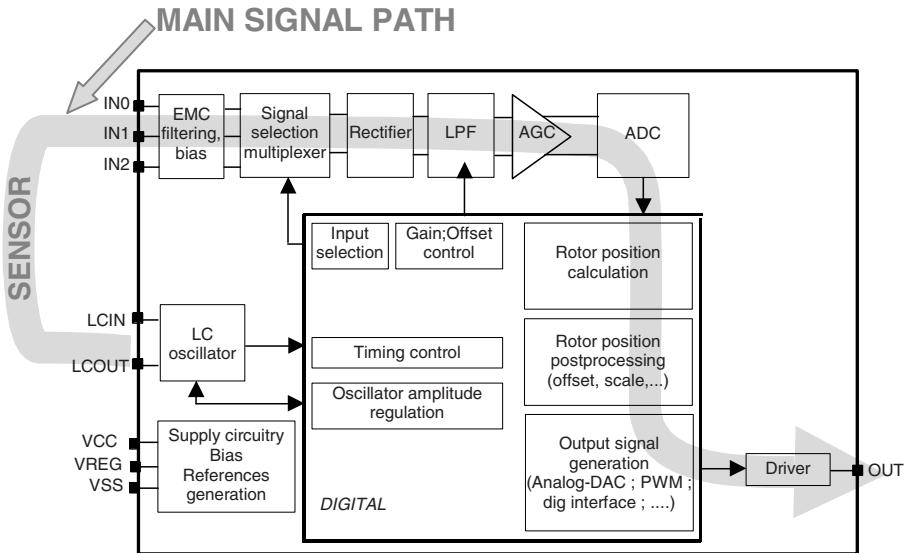


Fig. 5 ASIC block diagram

The second and higher order harmonics from the rectifier and remaining AC common mode signal are filtered by a combined differential/common-mode Low Pass Filter (LPF). The resulting DC signal is in the order of a few mV. Its amplitude varies significantly due to mechanical construction and mechanical tolerances (rotor to stator air gap).

To handle such a signal a high performance automatic gain control block is required [5]. A three stage Automatic Gain Control (AGC) block is implemented as shown in Fig. 6. The AGC gain is continuously adjusted by a finite state machine (FSM) at the end of each rotor position calculation to fully utilize the ADC input voltage range.

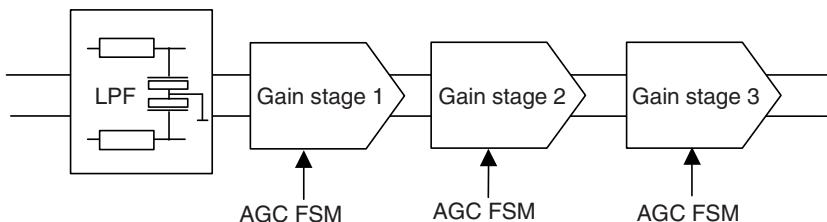


Fig. 6 AGC- gain control stage

Single ended input signal processing typically limits the resolution to 10 bits or 60dB Signal to Noise Ratio (SNR) due to the noise penetrating into the analog circuitry from the digital. This system is based on differential signal processing and achieves a resolution >12 bits.

The presence of a high amplitude common mode signal on the input results in the requirement for a high Common Mode Rejection Ratio (CMRR). This is achieved by full symmetry of the input amplifier (both schematic and layout). To meet the application requirements, the gain stage has to be fully differential, with very small gain non-linearity, high input and output signal dynamic range, high (preferably infinite) input impedance, very high CMRR and small offset. Because the gain stage is also used as a input stage of the connected ADC, high Slew Rate (SR) and gain bandwidth is required. The performance of the gain stage is key for the performance of the whole system and is further discussed in chapter “Automatic Gain Control”.

The analog output signal of the AGC stage is converted into a digital signal by a 12bit second order Sigma Delta converter. Further signal processing to calculate the rotor position is done by a custom DSP block.

The digital position signal is transmitted as an analog signal ratio metric to supply voltage or as a PWM signal or as a digital signal.

3.1 Redundancy requirement for safety relevant applications

For safety relevant applications, a redundant configuration with two independent systems is required. The inductive sensor can be configured as fully redundant. In this configuration, each system has its own fully galvanic separated excitation coil and its own set of receiving coils. The rotor structure is shared. Each system provides independent output signal. The redundant system application with two excitation coils is shown in Fig. 7.

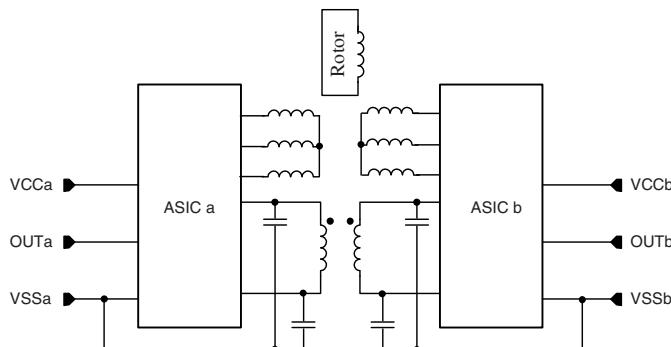


Fig. 7 Redundant application with two excitation coils

In redundant systems with two excitation coils, the mutual inductance is not negligible (as indicated by the coil coupling in Fig. 7). In the case one of the two systems is not supplied e.g. due to failure on the supply lines, the not supplied ASIC may not overload the other ASIC oscillator. A special oscillator design is required to fulfill this requirement (chapter “Oscillator driver”).

For diagnostic purposes, the oscillator amplitude has to be measured. For this measurement the available Sigma Delta converter is used and the oscillator signals are connected to the ADC inputs with CMOS switches. As seen in Fig. 8, normal CMOS switches have parasitic diodes to ground and supply. These diodes can clip the oscillation amplitude for signal amplitudes >2VBE and short circuits the excitation coil of a system with a failed supply. In this case, the electro-magnetic field induced into the excitation coil of the failing system by the other working ASIC is likely to have an amplitude higher than two diode voltages due to the high mutual inductance between the two excitation coils. The short-circuit on the failing systems coil will disturb/overload the oscillator of the still supplied ASIC.

A special, over supply voltage tolerant pass-gate has to be used in this case. This is described in the next point.

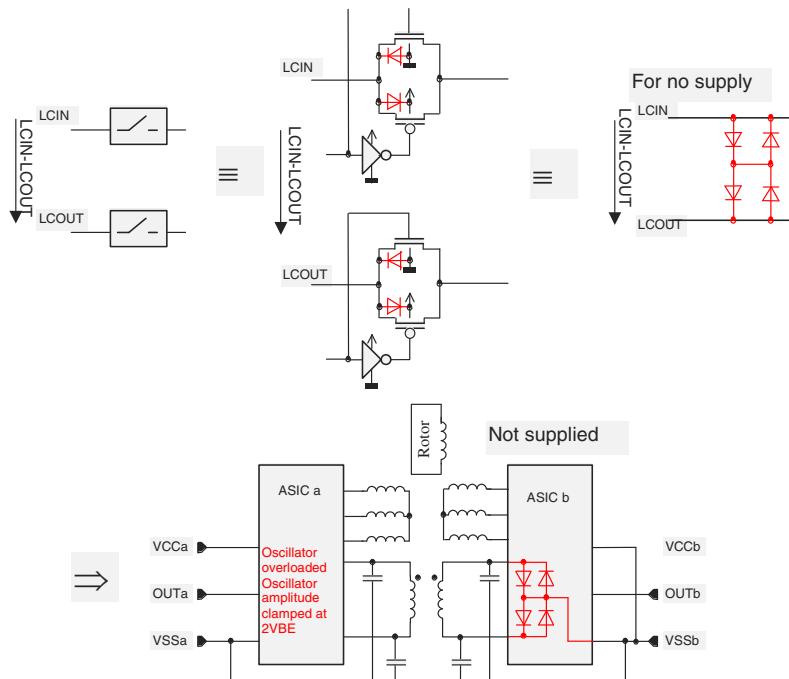


Fig. 8 Oscillator load of not supplied ASIC

The measurement path is always high impeded and does not cause any additional oscillator clipping in the redundant system.

3.2 Over-supply voltage tolerant pass-gate

To avoid interference when the supply is missing, over supply voltage tolerant pass-gate has to be used. Fig. 9 depicts a possible version of the over supply voltage tolerant pass-gate, that prevents undesired interaction through mutual inductance for negative DC shifts.

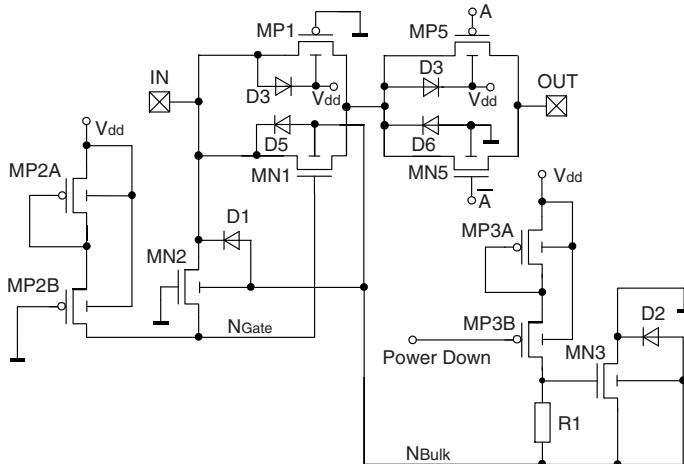


Fig. 9 Negative DC shift tolerant pass-gate

The NMOS of the CMOS pass-gate (MN1 and MP1) has a separate gate and bulk driving circuit, which prevents the build-in bulk-source diode D1 to conduct if the supply of the circuit is failing. If the supply is present ($V_{DD} > 2 V_{Th}$), the node NGate is biased to $V_{DD} - V_{Th}$ through the transistors MP2A and MP2B. If Power Down is off, the node NBulk is forced to the ground level by transistor MN3 since the gate of transistor MN3 is forced above one threshold voltage thanks to resistor R1 and transistors MP3A and MP3B. The signal Power Down can be replaced by VSS if power down is not required. The switch MN1 is ON with its gate high and its bulk connected to ground. MP3A and MP2A are diode connected and are used to increase the voltage on V_{DD} for which the pass-gate is negative voltage tolerant. On the other hand, it decreases the dynamic range on the input by lowering MN1 gate voltage on node NGate. MP3A and MP2A are optional and can be removed.

If the supply is missing, (i.e. $V_{DD} < 2 V_{Th}$) both MP2B and MN3 are switched off. The node NBulk (in the considered technology, the n type transistors are in a floating p-well which is not connected to the substrate or ground level) will

follow V_{in} (the parasitic capacitor at node N_{Bulk} will be discharged by D_1). The voltage at node N_{Bulk} will be limited to $\min(V_{in} + V_{BE1}, V_{BE2})$ where V_{BE1} and V_{BE2} are the built-in voltages of D_1 and D_2 . Indeed, if $V(N_{Bulk}) > V_{BE2}$, the diode D_2 is forward biased and N_{Bulk} is discharged through D_2 . The same happens through D_1 if $V(N_{Bulk}) > V_{in} + V_{BE1}$.

If V_{in} decreases below the ground level by more than one threshold voltage, the transistor MN_2 is activated and the node N_{Gate} is forced to V_{in} , MN_1 remains OFF. Node N_{Bulk} is floating and only capacitive discharging current flows through diode D_1 .

In summary, in case of missing supply, the input voltage will be allowed to drop below ground by more than 1 V_{Th} without creating a conductive path.

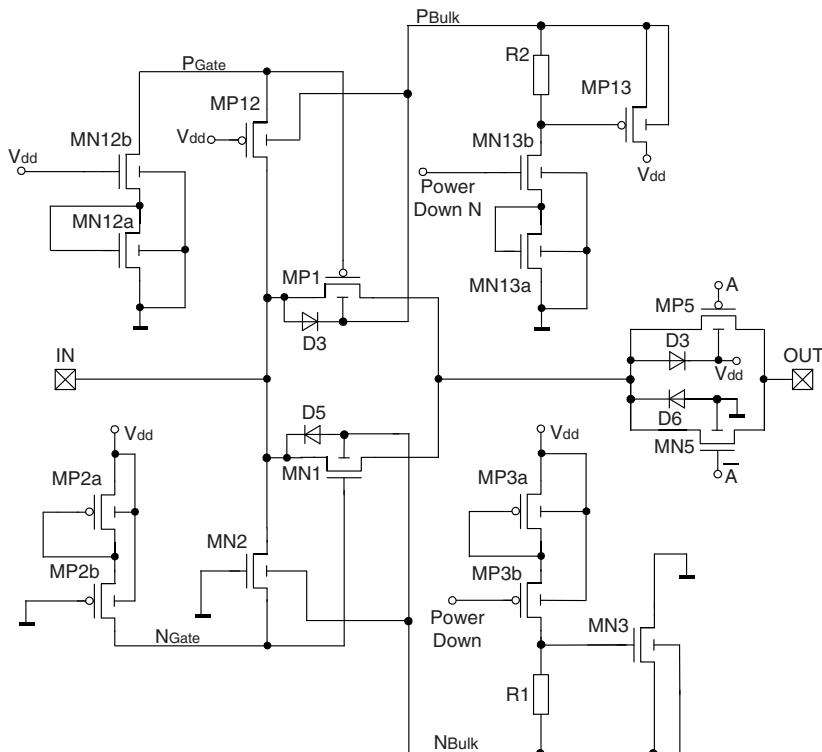


Fig. 10 Both positive and negative DC shift tolerant pass-gate

Out of the scope of the application, the described principle can be equally applied to a positive DC shift tolerant pass-gate (complementary structure to the negative DC shift tolerant pass-gate).

Combining both structures, a pass gate tolerant to both positive and negative DC voltages on the input in case of missing supply can be constructed as shown in Fig. 10.

3.3 Oscillator driver

A classical on chip oscillator driver with two transconductance Gm stages and off-chip RLC network, where all losses are represented by a serial resistance R_s , is shown in Fig. 11.

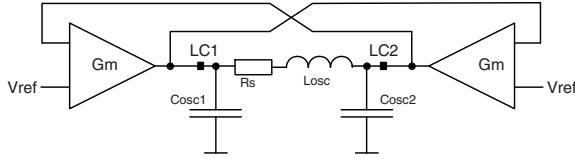


Fig. 11 Oscillator driver

Amplitude control

When we take for simplicity $C = C_{osc1} = C_{osc2}$, then the oscillation condition (to keep stable oscillations) is

$$G_{m0} = R_s \cdot \frac{C}{L_{osc}} = 2 \cdot \frac{R_s}{\omega^2 \cdot L_{osc}^2} = \frac{1}{2} \cdot R_s \cdot \omega^2 \cdot C^2 \quad (1)$$

To regulate the oscillation amplitude, non-linearity has to be inserted into the circuit. It can be created by limiting the output current of the drivers. The easiest approximation of the driver output current is shown in Fig. 12

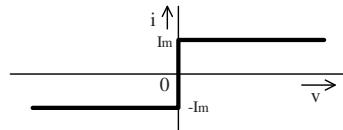


Fig. 12 Driver current (static)

Voltage of steady state oscillations is given by the formula (2)

$$V = 2k \cdot \frac{L_{osc}}{C \cdot R_s} \cdot I_M = 2k \frac{I_M}{G_{m0}} \quad (2)$$

Amplitude-voltage regulation is done by a digital control loop, which is controlling the maximum driver current.

$$\Delta V = 2k \cdot \frac{L_{osc}}{C \cdot R_s} \cdot \Delta I_M = V \cdot \frac{\Delta I_M}{I_M} = V \cdot \delta I \quad (3)$$

A linear voltage step requires an exponential current control. The required exponential function is approximated by a PWL function.

Driver realization

The main challenges for the driver are a wide dynamic range of the output current and high speed (to limit losses the driver must be much faster than the oscillation frequency, which is typically 4MHz) [4].

The wide dynamic range of the output current (0:1984) is obtained by a current DAC and parallel switching of Gm output stages in the driver. Since the relative resolution is limited, instead of using linear 11-bit DAC an exponential type 7-bit DAC has been used. The full 7-bit scale of the DAC is divided into 8 ranges and in each of these ranges the output current step is constant (piecewise linear approximation of exponential function) – see Fig. 13.

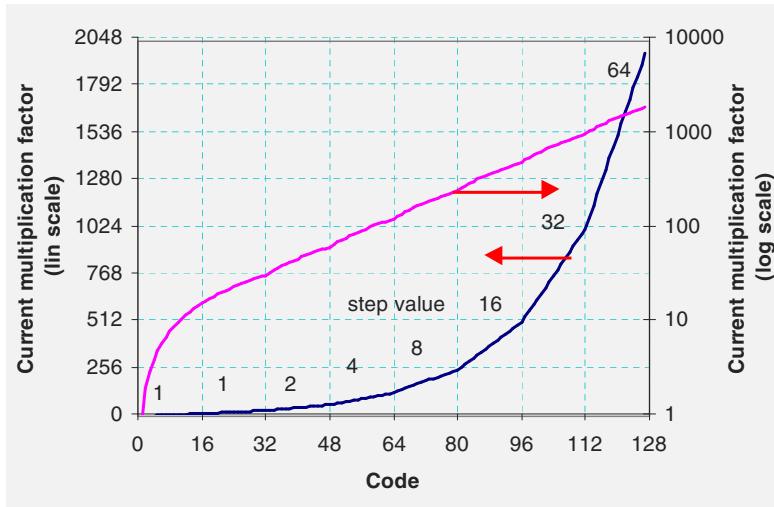


Fig. 13 Current multiplication for 7-bit PWL approximated exponential DAC

Simplified block diagram of the oscillator driver is shown in Fig. 14. It consists of a prescale block (with scaling factors 1, 2, 4 and 8) delivering current I_{ref2} into two complementary current mirrors. Both top and bottom current mirrors consist of two parts – one with fixed current outputs of 16, 16, 32 and 64 times I_{ref2} and a second part with a 7-bit binary weighted current DAC with output current from 0 to 127 I_{ref2} . Oscillator Gm block integrates two functions. When

increasing the code, which increases current, it switches more fixed currents from top and bottom current mirrors to the output. The second function activates more output stages in parallel.

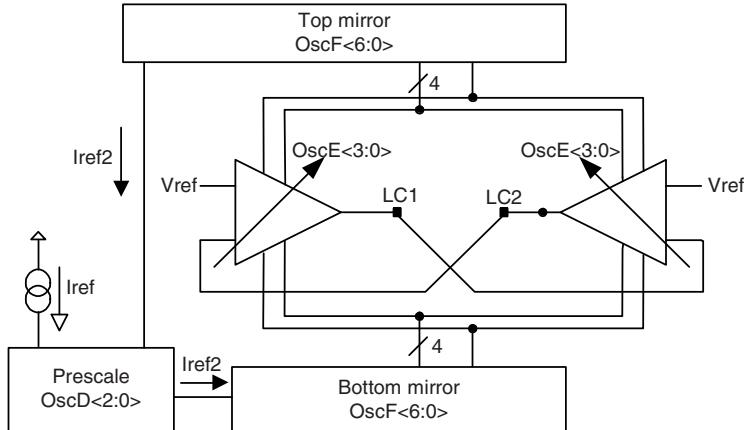


Fig. 14 Simplified diagram of oscillator current limitation

Table 1 Coding of driver control signals

DAC Segment	7 bit input data $B<6:0>$		Prescaler output $Iref2$ units	Active G_m stages	Control Signal Codes		
	MSBs	LSBs			$Iref$	$Ibias$ OTA	DAC
-				-	$OscD<2:0>$	$OscE<3:0>$	$OscF<6:0>$
0	000	B3 B2 B1 B0	1	1	000	0000	0 0 0 B3 B2 B1 B0
1	001	B3 B2 B1 B0	1	2	000	0001	0 0 0 B3 B2 B1 B0
2	010	B3 B2 B1 B0	2	2	001	0001	0 0 0 B3 B2 B1 B0
3	011	B3 B2 B1 B0	2	3	001	0011	0 0 B3 B2 B1 B0 0
4	100	B3 B2 B1 B0	4	3	011	0011	0 0 B3 B2 B1 B0 0
5	101	B3 B2 B1 B0	4	5	011	0111	0 B3 B2 B1 B0 0 0
6	110	B3 B2 B1 B0	8	5	111	0111	0 B3 B2 B1 B0 0 0
7	111	B3 B2 B1 B0	8	9	111	1111	B3 B2 B1 B0 0 0 0

Setting of the current limitation is done using 3 independent control busses: prescaler bus $OscD<2:0>$, G_m -switching bus $OscE<3:0>$ and current mirror bus $OscF<6:0>$. Simple logic is used to decode these oscillator control busses from the 7 bit input digital word. Table 1 shows how to generate the control signals. The output current can be calculated using following formula

$$I_{out} = I_{unit} \cdot (1 + OscD) \cdot \left[OscF + 16 \cdot \left\{ OscE \bmod 2 + \text{int} \frac{OscE}{2} \right\} \right] \quad (4)$$

Overdriving output without supply

In redundant systems with two drivers and two mutually coupled excitation coils if one of the systems looses supply or ground connection, it cannot load the other system, which must remain working.

The standard CMOS output driver (Fig. 15a) has intrinsically built in diodes, which will cause loading of the other oscillator if one system looses the supply voltage. Because the oscillator coil is floating, connected only to LCIN and LCOUT pins, it is sufficient to remove only one of the two diodes to prevent loading of the other system. In case an additional PMOS is used (Fig. 15b), the LCx pin can go negative, but the voltage range of the driver is limited (due to voltage needed to open MP1d).

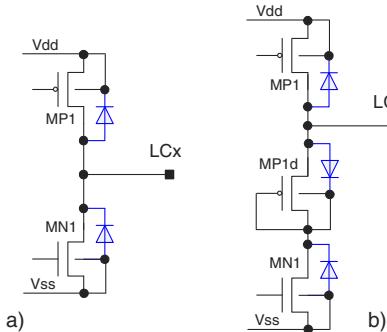


Fig. 15 CMOS driver topology

To allow the LCx pin to go negative when the supply voltage is lost and to not change the voltage range of the driver we use the circuit in Fig. 16. When the chip is powered on, transistor MP6 is on and node NBulk is shorted by MN6 to ground. To enable the driver, signals Ena and EnaN (inverted) are used.

Without supply, the voltage on Vdd is lower than 2 PMOS V_t needed to switch on MP6. MN6 is then off. For negative voltage on the LCx pin (below NMOS V_t) transistors MN5 and MN3 are on connecting Nbulk and Ng1 to LCx potential. This ensures that MN1 is off. All PMOSes connected to LCx potential are also off (MP1, MP3, MP4) and there is no current flowing through the LCx pin. For positive overdrive on LCx bulk diode of MP1 is activated. MP3 is used to increase potential on node Ng2 to cancel the current path through MP1 (I_{top} is common for both LC1 and LC2 drivers).

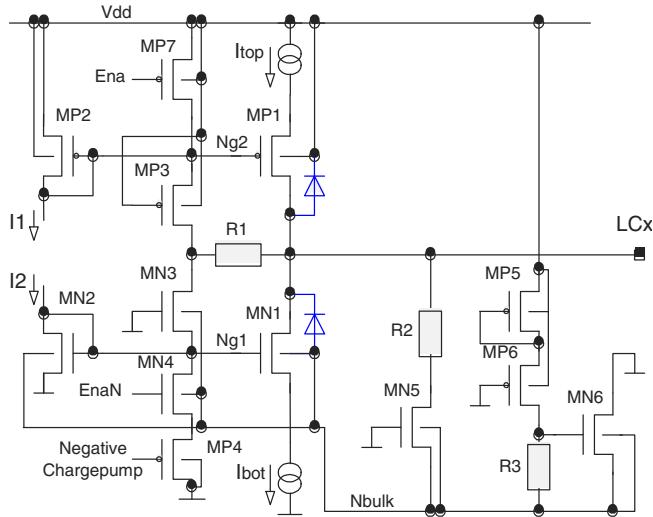


Fig. 16 Output driver topology

3.4 Automatic gain control stage

To accomplish the demanding requirements mentioned at the beginning of chapter three, a fully differential Current Feedback Amplifier (CFA) was used [5]. It provides symmetrical high impedance inputs, accurate and temperature stable gain, simple gain control and wide input dynamic range. Intrinsic input offset is comparable to common Voltage Feedback Amplifier (VFA) structures and is further reduced by build in offset compensation circuit. Typical VFA structures with resistor feedback either suffer from non-zero input current, high noise, limited SR or high input offset.

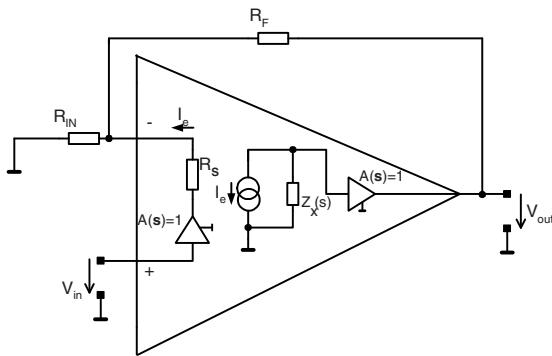


Fig. 17 Standard CFA structure

The standard CFA approach is shown in Fig. 17. The formula describing the gain of this structure is

$$A_{CL}(\mathbf{s}) = \frac{1 + \frac{R_F}{R_{IN}}}{1 + \left(R_F + R_S \left(1 + \frac{R_F}{R_{IN}} \right) \right) Z_x(\mathbf{s})} \quad (5)$$

Assuming $R_S = 0$ and $R_F \ll Z_x(s)$, we can simplify (5) to a form

$$A_{CL}(\mathbf{s}) = 1 + \frac{R_F}{R_{IN}} \quad (6)$$

The requirements of symmetrical high impedance inputs don't allow the direct application of the standard CFA structure. This can be solved by the use of differential CFA amplifier as described in Fig. 18.

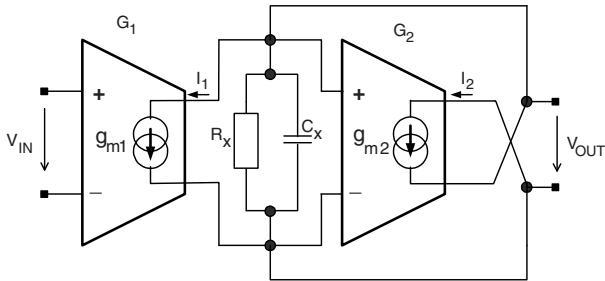


Fig. 18 Differential CFA model

The core is composed of two transconductances gm_1 and gm_2 . The CFA gain frequency response is given by

$$A(s) = \frac{-g_{m1}R_X}{1 + g_{m2}R_X + sR_XC_X} \quad (7)$$

By proper design we can easily guarantee that $g_{m2}R_X \gg 1$ and simplify (7) into form

$$A(s) = \frac{-g_{m1}R_X}{g_{m2}R_X + sR_XC_X} \quad (8)$$

From (7) is clear that the DC gain is

$$A_0 = -\frac{g_{m1}}{g_{m2}} \quad (9)$$

The main limitation of this topology for gain $A_0 \neq 1$ is the transconductance dependency of the degenerated differential pair on the input voltage amplitude. By a slight modification of the described topology (Fig. 18) we can obtain a gain stage for which the input voltage seen by g_{m1} and g_{m2} ; for $g_{m1} = g_{m2}$ are equal even for $A_0 \neq 1$ and thus the transconductance dependences are canceled out. The topology is shown in Fig. 19.

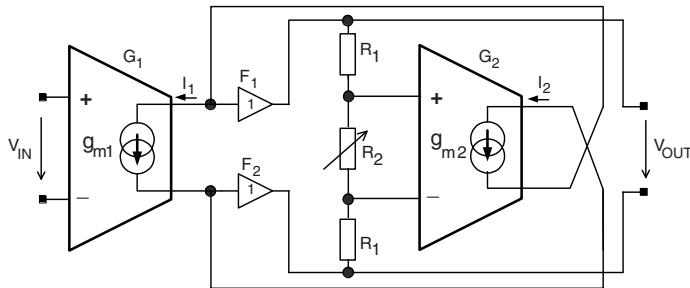


Fig. 19 Improved differential CFA model

The DC gain is

$$A_0 = -\frac{g_{m1}}{g_{m2}} \cdot \frac{1}{A_{Rdiv}} \quad (10)$$

where

$$A_{Rdiv} = \frac{R_2}{2R_1 + R_2} \quad (11)$$

For typical case with $g_{m1} = g_{m2}$ the equation can be even further simplified to

$$A_0 = -\frac{1}{A_{Rdiv}}. \quad (12)$$

Equation (12) for DC gain is exactly what is required for VLSI integration - the gain is defined by the ratio of two easily controllable parameters and is independent of its absolute value.

Practically, the gain accuracy can easily be within the range of $\pm 0.5\%$ without extra trimming circuitry.

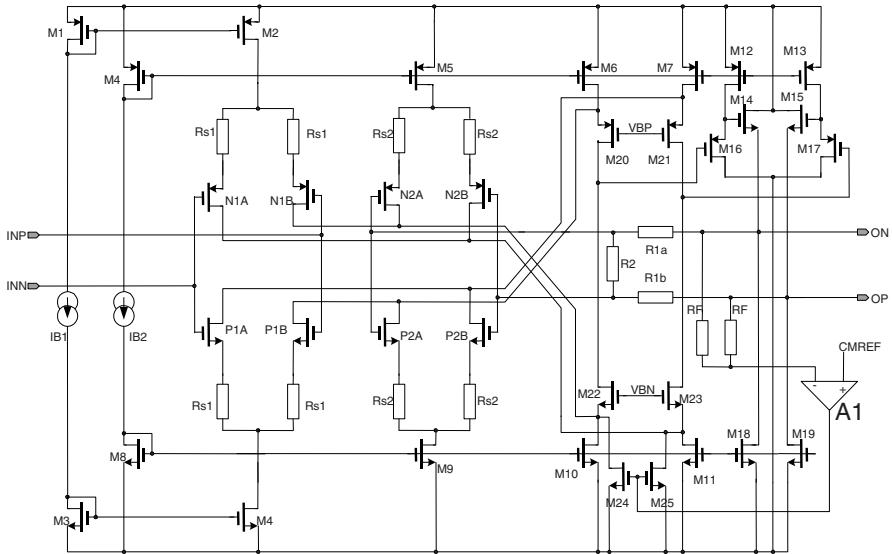


Fig. 20 Differential variable gain CFA

The simplified transistor level schematic is shown in Fig. 20. The structure consists of two complementary, degenerated differential pairs. The input complementary differential pair (N1A, N1B, P1A, P1B) realizes the g_{m1} and differential pair (N2A, N2B, P2A, P2B) realizes the g_{m2} . The folded cascode (M6, M7, M10, M11, M20-23) output impedance corresponds to R_X and C_X of the model (Fig. 18). The simple output followers are realized by M12-19. The gain is adjusted by resistor divider (R1a, R1b, R2). Common mode feedback circuitry (M24-25, RF and A1) is only symbolically included.

The measured DC gain as a function of input signal amplitude for selected gain 1, 2, 4 and temperature -40°C to 160°C is shown in Fig. 21.

4. Sensor accuracy measurement results

For the evaluation of the full system a configuration with analog output driver ratio-metric to the supply was used. The performance evaluation suffers from the fact that it is difficult to distinguish between errors caused by the ASIC and errors caused by the sensor itself. The presented results cover the overall system performance from mechanical sensor to ASIC analog output.

The measurement results of a standard production system, which includes all imperfections of the sensor coil system and the imperfections and temperature drifts of the ASIC actuator/driver is shown in Fig. 22. The angular position error

remains below 0.1% over the whole input range of the sensor. The measured temperature drift of the full system including analog output drivers for the extended temperature range of -40degC to 160degC is also less than 0.1%. The total system has a maximum accuracy error of $\pm 0.1\%$.

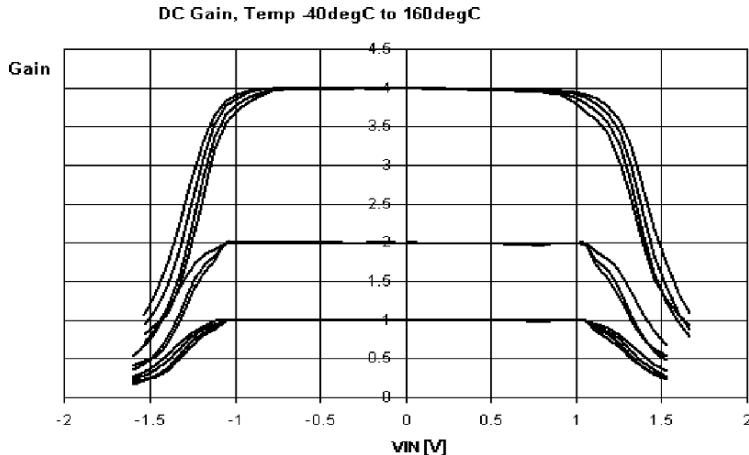


Fig. 21 Measured DC gains adjusted to 1,2 and 4 in temp. range -40 to 160°C

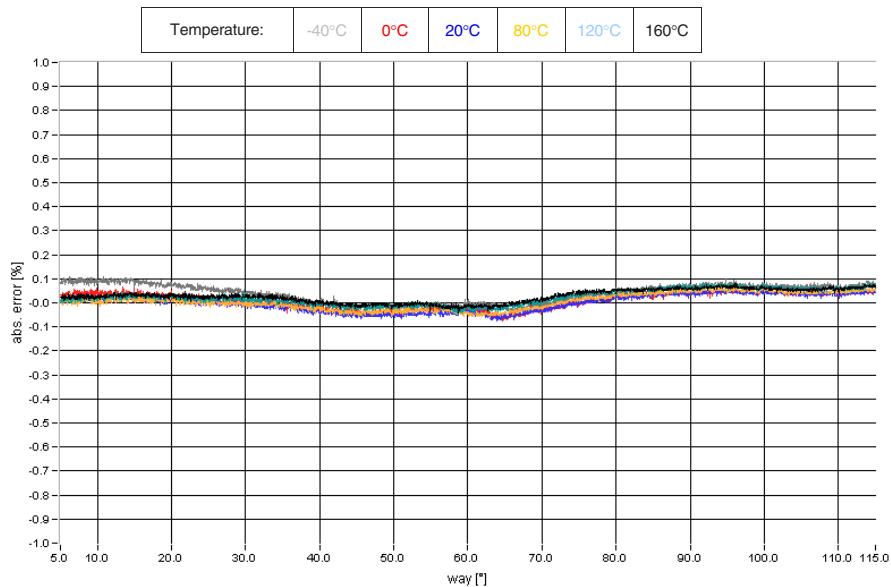


Fig. 22 System accuracy measurement, temperature drift -40°C–160°C, measured with standard production sensor

5. Hella's CIPOS Sensor: Typical Applications

Due to the particularly large flexibility for possible form factors, the CIPOS measurement principle is used in the automotive industry for very different products. In this section some of these are briefly introduced.

Fig. 23 shows two different forms of electronic accelerator pedals equipped with CIPOS sensors. One can see, depending on the structure concept, different shapes of the sensor are used. In the case of the hanging pedal (left photo) that is a segment sensor, in the case of the standing pedal (right photo) a linear sensor.



Fig. 23 Acceleration Pedals with CIPOS-Sensors

The sensors increasingly become also a component of other functional units. A nice example of it are the solenoid operated throttle valves as seen in Fig. 24 left and the headlamp swiveling module seen in Fig. 24 right.

Beside the spatial also the functional integration of sensors plays an important role in many cases. As an example an electro-powered actuator is shown in Fig. 25.

The electronics contains the stator of the sensor for the acquisition of the actual position of the drive lever, an ASIC for the evaluation of the sensor signals, a further ASIC with micro-controller and power-output stage for the motor control as well as the components for a CAN interface. The sensor thereby is not only spatial but also directly functionally integrated into the actuator electronics.



Fig. 24 CIPOS-Sensors in Throttle-Flap and a Headlamp Swivelling Module



Fig. 25 CIPOS-Sensor in an Actuator

6. Conclusions

The presented inductive position sensor provides a powerful solution, compared to other concepts, suitable for the harsh automotive environment. The sensor is very precise, with no contact between the moving part and the stator, it is especially well suited for use in applications with frequent sensor movements even if an exceptional accuracy is required.

The system is based on the sensors planar coils, one ASIC and few external components. The ASIC provides rotor position conversion rate up to 4k samples/sec. The system is fully operational within 5.5ms after supply connection, and consumes typically 8mA. For safety relevant applications, the sensor can be configured as a dual, fully redundant system. Each channel is fully galvanic separated and provides an independent output signal. Both systems share a common rotor structure. The system fulfills all the current automotive EMC requirements.

References

- [1] Hlubeck B.; Hobein D.: "Smart sensor technology - The Basis for Perfect Performance", ATZ 102 (2000) 12.
- [2] Dorissen H.-T.; Hobein D.; Irle H.; Kost N.: "Non-Contacting Linear and Angular Position Sensors - New Technology in Automotive Engineering Applications", ATZ 100 (1998) 10.
- [3] Irle H.; Diekmann J.: "Replacing Throttle Valve Potentiometers with Non-Contacting Inductive Sensors", ECN, April 2002, pp 31.
- [4] P. Horsky, "LC Oscillator Driver for Safety Critical Applications". In Proc. of DATE 05 Conference, Designers Forum, Munich, Germany 2005, pp 34-38.
- [5] I. Koudar, "Variable Gain Differential Current Feedback Amplifier". In Proc. of CICC 04 Conference, Orlando, Florida, 2004, pp 659-662.

CMOS SINGLE-CHIP ELECTRONIC COMPASS WITH MICROCONTROLLER

Christian Schott, Robert Racz, Samuel Huber, Angelo Manco,
Markus Gloor, Nicolas Simonne*

Melexis Technologies, Bevaix (CH) / Tessenderlo (B)*

Abstract

We present a CMOS single-chip electronic compass sensor including the complete digital signal processing for accurate heading calculation, which is the continuation of a work presented at the International Solid State Circuits Conference (ISSCC) 2007 in San Francisco/USA [1]. The device's analog part consists of Hall based three-axis magnetic field transducer with integrated magnetic concentrator that operates as passive magnetic amplifier. The analog amplification chain features a gain of up to 20'000. A true-12-bit extended counting ADC converts the amplified magnetic field signals into the digital domain where a 16-Bit microcontroller calculates the heading information and outputs it via an SPI interface.

The compass sensor is realized in 0.35um low voltage CMOS technology plus a simple batch processing step for deposition of a metal layer. The compact die size of 2.3 mm x 2.8 mm allows for packaging into a standard 4mm x 5mm x 1mm surface mount plastic package.

The heading resolution is better than 0.5 degrees and the accuracy better than +/- 2 degrees.

1. Introduction

CMOS integrated Hall devices are well known as transducer elements for position sensing of a target magnet or for measuring an electrical current. They are fully CMOS compatible, of small size and typically yield output voltages of a few mV when a flux density of a few 10mT is applied. Their inherent bridge offset is also in the few mV range and can be efficiently suppressed by the so-called spinning current method [2], where signal and offset are separated in the frequency domain. For the direct measurement of very low magnetic field densities in the micro-tesla range, only discrete Hall devices in high-mobility compound semiconductors like InSb have been used to date. CMOS Hall

devices feature a relatively low sensitivity of typically 10uV/Gauss, which gives a signal amplitude of merely 2uV in the earth magnetic field of 20uT. In this case, to obtain a reasonable signal-to-noise ratio of 40dB within an integration time of a few milliseconds, the signal must be amplified without adding noise. This can be achieved via a passive magnetic amplifier implemented as thin metal layer on the silicon surface.

We present in this paper a CMOS single-chip electronic compass sensor that includes all the digital signal processing required for an accurate heading calculation. The sensor is realized in a 0.35um low voltage CMOS technology with integrated magnetic concentrator (IMC) post-process [3, 4]. The IMC process has recently been fully automotive qualified for another product, where it is now used for volume production.

2. Compass Architecture

The device's analog part consists of a three-axis magnetic field transducer based on Hall elements and an analog amplification chain. A 13-bit ADC converts the amplified signals into the digital domain where a 16-Bit microcontroller calculates the heading information and outputs it via serial interface. The block diagram in Fig. 1 shows the architectural concept and signal flow inside the circuit.

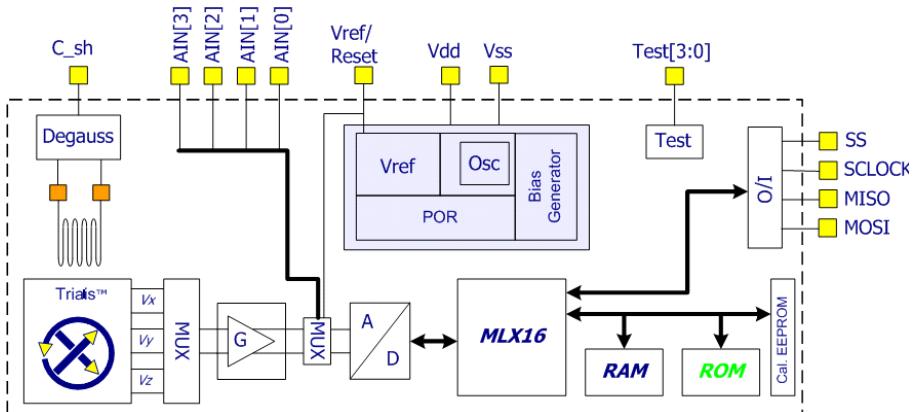


Fig. 1 Block schema of the electronic compass

3. Magnetic Frontend

The three-axis magnetic field transducer is based on a combination of CMOS Hall devices and a structured metal layer on top of the silicon surface. The Hall elements are realized in a pinched n-well with a cross-shaped geometry, and have an active area of about $15 \times 15\mu\text{m}^2$. Without any additional magnetic layer, such Hall elements are only sensitive to the vertical magnetic field component B_z . By integrating a magnetic concentrator (IMC) on top of the silicon, a CMOS Hall sensor can be designed to measure all three magnetic field components B_x , B_y and B_z separately (Fig. 2).

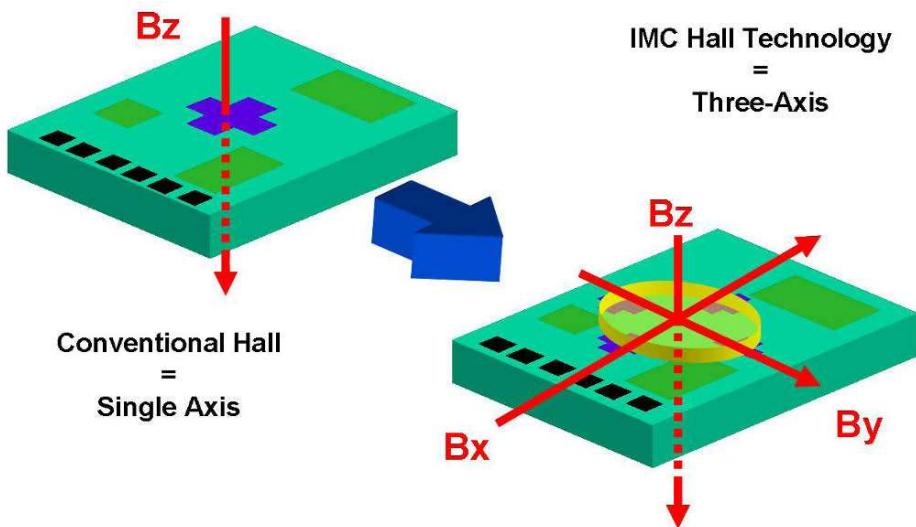


Fig. 2 By integrating a magnetic concentrator (IMC) onto the silicon surface, a conventional singel axis Hall sensor can be enhanced to measure all three magnetic field components B_x , B_y and B_z

Such a 3-axis field transducer generates three Hall voltages V_x , V_y and V_z , corresponding to the three magnetic field components B_x , B_y and B_z . On top of that, the IMC structure can be designed in a way to locally concentrate the flux lines onto the Hall elements, it works as a passive magnetic amplifier with a gain of up to 10.

3.1 IMC Layout

In our electronic compass application we need to achieve maximum magnetic gain and we also need to reduce any deterioration from accidental magnetization of the ferromagnetic layer. We therefore decided to make a layout of five octagonal rings as shown in Fig. 3. Compared to a former layout with circular rings used in [5] the octagonal rings are more robust concerning the post process tolerances. The etching of a constant-width air gap between octagons is easier to control than the etching of a concave gap between rings, resulting in less mismatch of sensitivity between the axes.

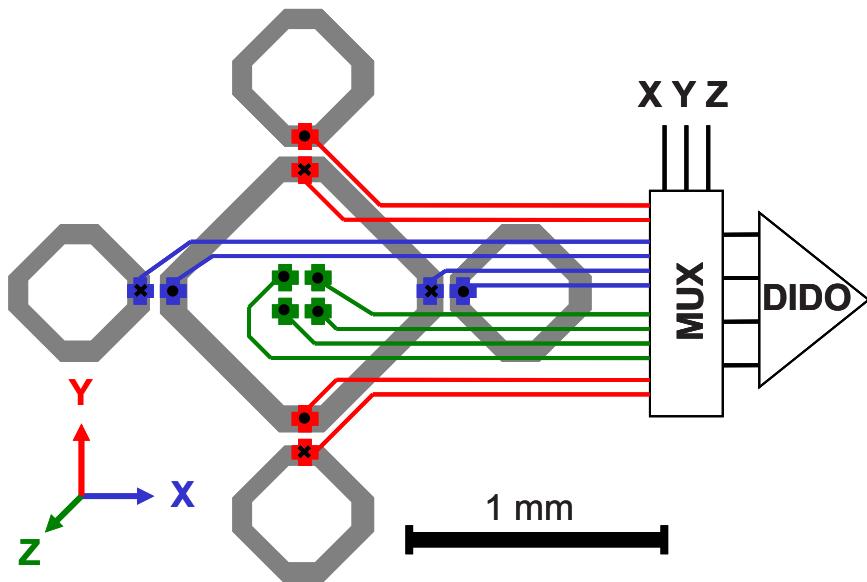


Fig. 3 Entire IMC structure and Hall elements. The dot denotes that the local field comes out of the plane and the cross denotes that the local field goes into the plane for applied external field in X or Y or Z direction

Four Hall elements generate the field proportional signals for each of the three axes X, Y and Z. For X and Y axis two groups of two Hall elements are arranged under the IMC around the airgaps between rings. For Z, a group of four Hall elements is placed in the big ring center. Those work as ordinary Hall elements without taking advantage from the magnetic gain.

3.2 IMC Working Principle

The operation principle for X and Y is illustrated in the top view of Fig. 4 and the corresponding cross-section view of Fig. 5.

With an applied external field (e.g. earth field) from left to right, the magnetic flux is concentrated inside the left octagon, then crosses the gap and continues inside the right octagon. The Hall elements are positioned under the octagons close to the gap.

The interesting part of the operation becomes clear in the cross-section view of Fig. 5. Around the gap, the flux lines are not limited to the metal layer plane, but they are also running in semi-circles above and below the metal. Due to Maxwell's equation of a source-free magnetic flux density $\text{div}B=0$, all flux lines are crossing the border between high permeability metal and low permeability air or silicon at right angle.

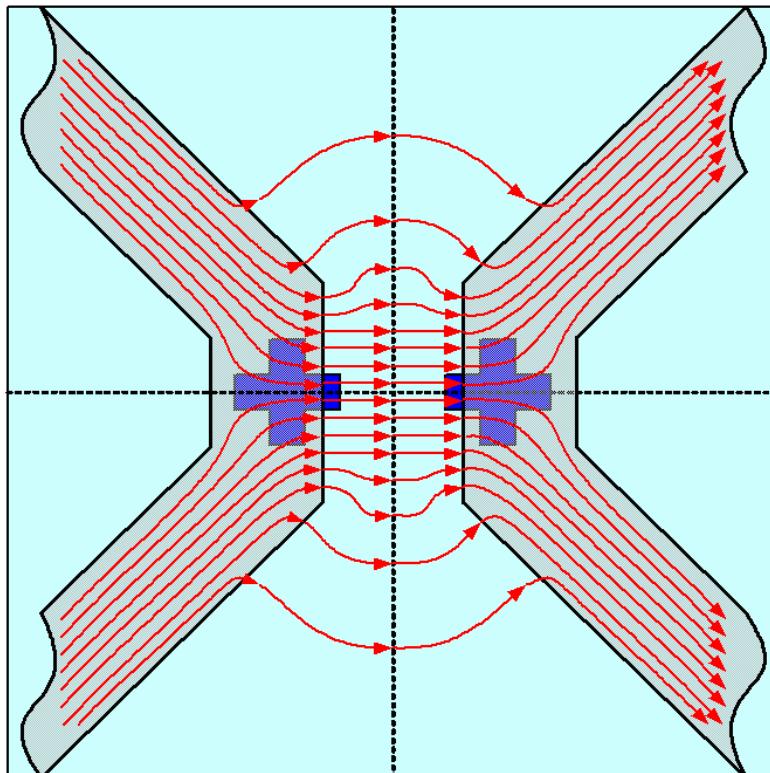


Fig. 4 Magnetic Flux around air gap (top view)

This means that some of the flux lines leaving the left metal piece are going down through the Hall element underneath and on the right up through the other Hall element. In such a way they are close to perpendicular to the silicon surface and can be measured by the Hall elements. Subtracting the output of the two Hall elements yields therefore a voltage proportional to the horizontal flux density. Any external vertical component of the flux density appears as common mode on both Hall outputs and is cancelled out, so that the difference of the two Hall voltages is a pure measure for the horizontal flux density.

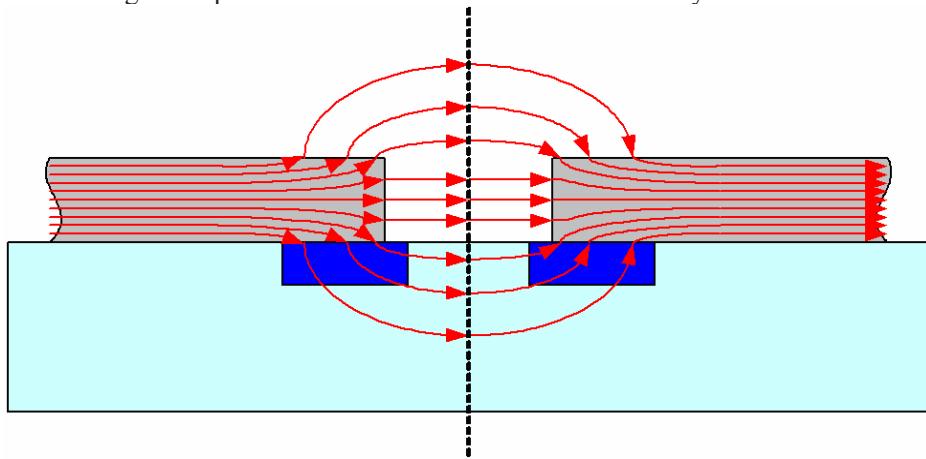


Fig. 5 Magnetic Flux around airgap (cross-section view)

In addition, the flux lines are closer together around the gap area compared to the applied external field, which means that the flux is amplified here.

3.3 IMC Process Technology

The IMC process is a photolithographic batch process which is applied to an 8-inch CMOS wafer at a time. It consists of the following steps (Fig. 6):

1. CMOS wafer with passivation and bond pad openings from the fab
2. a glue layer of a few microns thickness is dispersed on the wafer
3. the 20um thick metal layer is bonded to the wafer and the glue is cured
4. the photoresist is applied on top of the metal layer
5. the illumination mask is aligned above the wafer
6. the photoresist is illuminated by UV light
7. the photoresist is developed
8. the metal layer is etched and remains only underneath the undeveloped photoresist
9. the remaining photoresist is stripped and the wafer is cleaned from remaining glue residues

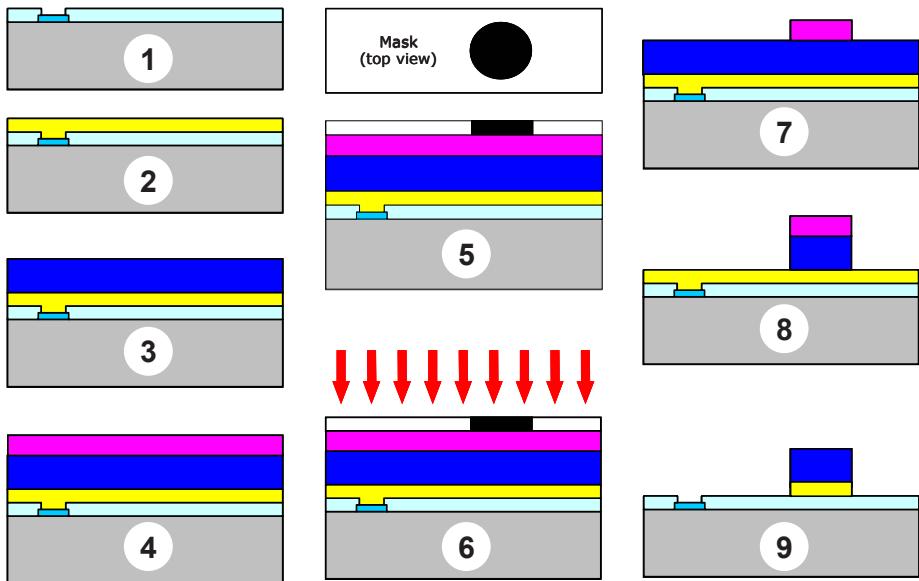


Fig. 6 IMC photolithographic batchprocess

The resulting metal structure after the post process is shown on Fig. 7. The octagons have a ring width of about 60um and the material is about 20um thick. The gap between the rings is about 40um which is the best trade-off between insensitivity to process tolerances and high magnetic gain.

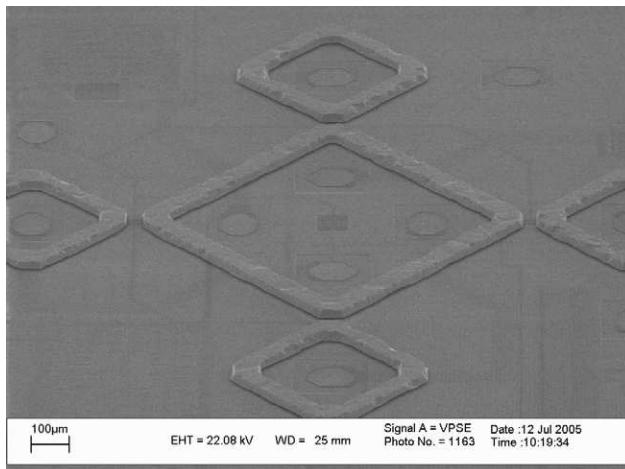


Fig. 7 SEM photograph of the magnetic concentrator

3.4 Degaussing

The sensor additionally contains a circuit block for “low-energy” demagnetization of the ferromagnetic concentrator, to eliminate remnant magnetization caused by strong accidentally-applied external fields. The degaussing coil is a single-turn loop through the IMC ring structures, whose lower part is implemented as integrated metal tracks and whose upper part is finished during packaging by bonding connection. By sending a 10 μ s long current pulse of about 20mA through the center of the octagon rings, all magnetic domains inside the rings can be aligned along the octagon ring Fig. 8.

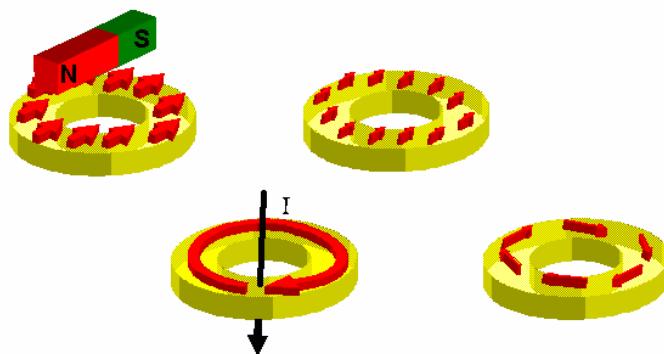


Fig. 8 Ring demagnetization by short current pulse eliminates magnetic offset

This means that the flux lines from material magnetization are closed inside the ring, not going through the Hall elements anymore. Any residual offset from magnetization is cancelled.

4. Analog Frontend

The analog frontend consists of the Hall voltage modulation by spinning current operation, the amplification by up to 20'000, the demodulation and filtering.

4.1 Spinning Current Operation

The principle of spinning current operation for Hall elements is well known to separate the inherent bridge offset from the Hall signal. It consists of cyclic change of supply terminals and output terminals as shown in Fig. 9. In a first phase the current runs from Top to Bottom through the Hall device and the output contacts are left and right and in a second phase the current runs from right to left through the device and the outputs are top and bottom terminals.

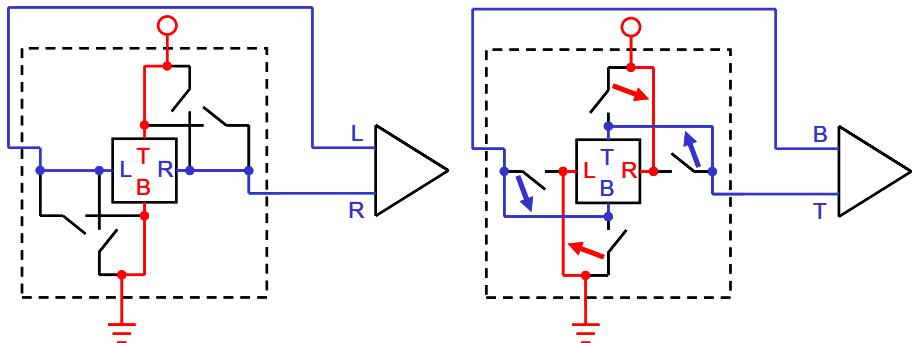


Fig. 9 Spinning Current concept for modulated Hall voltage

The principle as shown here actually rotates the current contacts clockwise and the output contacts counter-clockwise from phase 1 to phase 2. This results in a modulation of the Hall signal with the inherent offset signal being close to DC.

As shown in Fig. 10, the modulated Hall voltage V_1 is amplified and then demodulated. The DC part consisting of Hall offset and amplifier offset is cancelled out by the demodulation (and subsequent filtering).

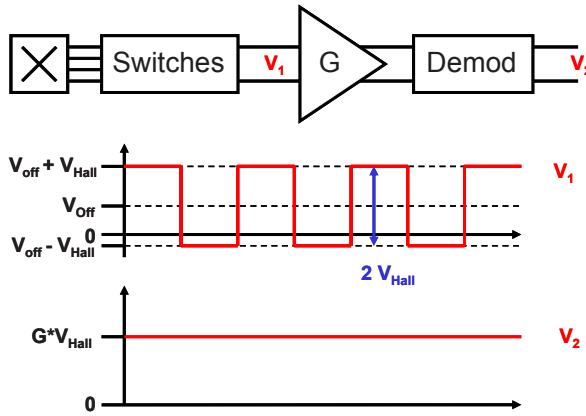


Fig. 10 Spinning Current and Signal Demodulation

4.2 Amplification Chain

As shown in Fig. 11, the Hall elements are current biased and the Hall voltage is modulated to 25kHz by using the spinning current technique to prevent the offset voltages of the Hall elements from saturating the amplifiers.

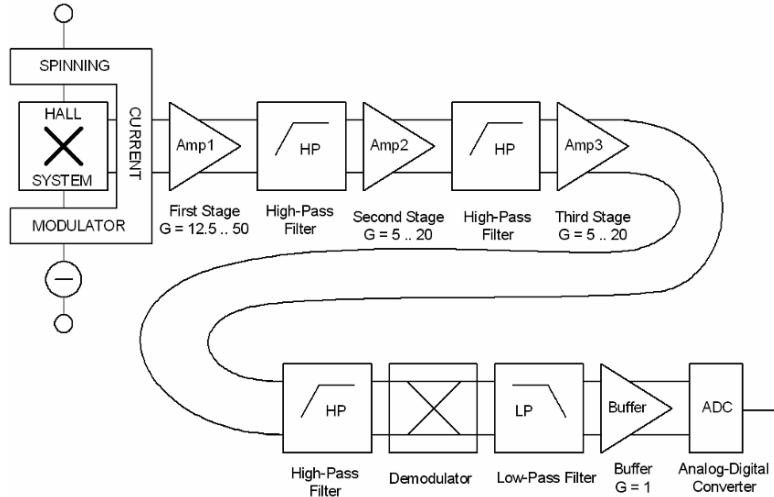


Fig. 11 Hall Frontend and analog amplification chain

The fully differential amplification chain consists of three stages, and has a programmable analog gain, which can be varied between 312.5 and 20'000. The stages are connected via high-pass filters for further elimination of any DC offset voltages. After demodulation, the signal is 8kHz low-pass filtered.

The X and Y axis sensitivity is programmable between 0.9 and 58 LSB/ μ T, while the Z axis sensitivity is about 8 times smaller due to the lack of IMC gain.

4.3 A/D Converter

The ADC uses an extended counting technique [6], which is a blend of $\Sigma\Delta$ modulation with its high resolution but relatively low speed and algorithmic conversion with its higher speed but lower accuracy.

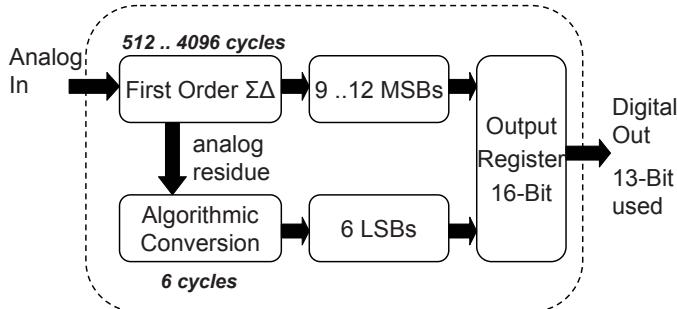


Fig. 12 Extended Counting ADC architecture

The converter successively operates first as a first-order $\Sigma\Delta$ modulator (counting phase) to convert the MSB's, and then the same hardware is used as an algorithmic converter (extended phase) to convert the remaining LSB's. The ADC operates at low current (typ 250 μ A) and low sampling frequency (400kHz) and it is programmable between 15 and 18 bits of which only the 12 MSBs are used. During the counting phase between 512 and 4096 samples are integrated.

4.4 On-chip Microcontroller

The digitized data is transmitted to an on-chip microcontroller, where it is first passed through a 100Hz digital low-pass filter for further noise reduction. After temperature correction of offset and sensitivity, the angle of the applied magnetic field (earth's field) is computed with the help of a firmware algorithm stored in the ROM (Fig. 13). This heading information is then output to an external host microcontroller via the digital interface. By using the on-board EEPROM, parameters like offset fields, axis selection, analog gain, zero setting and interface protocol (I^2C or SPI) can be defined. Those parameters allow the end-user to entirely calibrate the single-chip compass even in the presence of magnetic field distortions caused by ferromagnetic objects, e.g. batteries, sensor housing, car body etc. The sensor can also digitize four analog inputs from external sensors and transmit the results to a host system via the on-board digital interface.

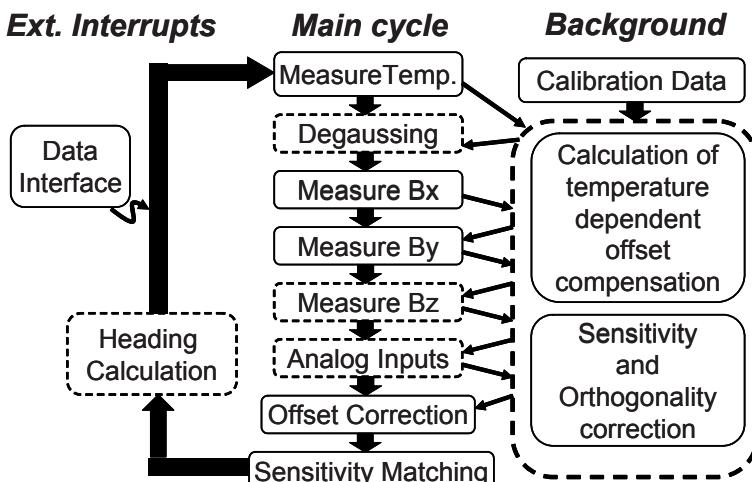


Fig. 13 Microcontroller digital signal processing

5. Realization in 0.35 μ m CMOS

The sensor chip has been realized in low-voltage 0.35 μ m mixed-signal CMOS technology (Fig. 14). After completion of the CMOS process, the IMC layer is applied and photolithographically structured. The on-chip degaussing coil is then completed during the wire bonding to the package leadframe.

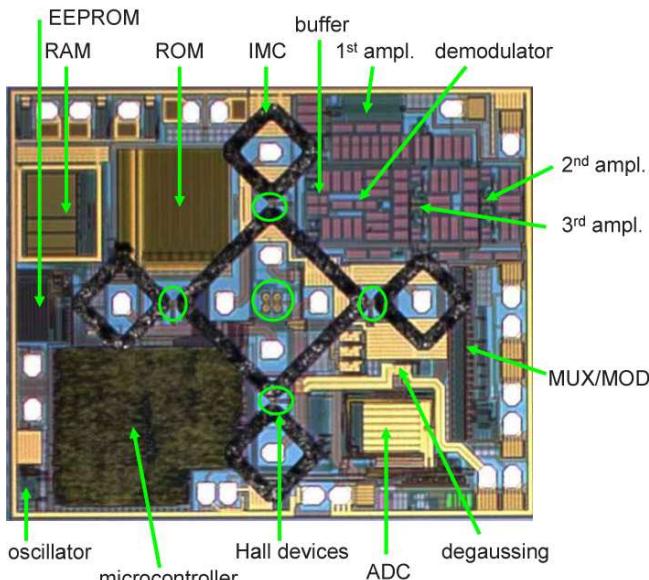


Fig. 14 Photograph of silicon die with IMC

The compact die size of 2.3 x 2.8 mm² allows for packaging into a standard 20-pin plastic package, which measures only 5mm x 4mm x 1mm (Fig. 15). Due to rigorous low-voltage design of all circuit parts, the compass sensor works within a supply range of 2.2 to 3.6 Volts and features a current consumption of 5mA in normal power mode and less than 2mA in low-power mode.

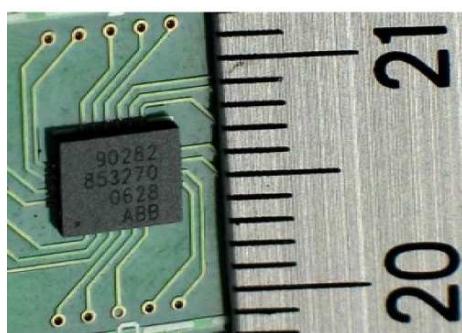


Fig. 15 Photograph of 4 x 5 x 1 mm³ plastic package

6. Measurements

The output characteristics of the packaged sensor were measured by using a 3 dimensional Helmholtz coil. The angular error of max 1.5° was measured by rotating the sensor within the earth magnetic field at constant temperature (Fig. 16). The heading resolution is better than 0.5° and the accuracy better than $\pm 2^\circ$.

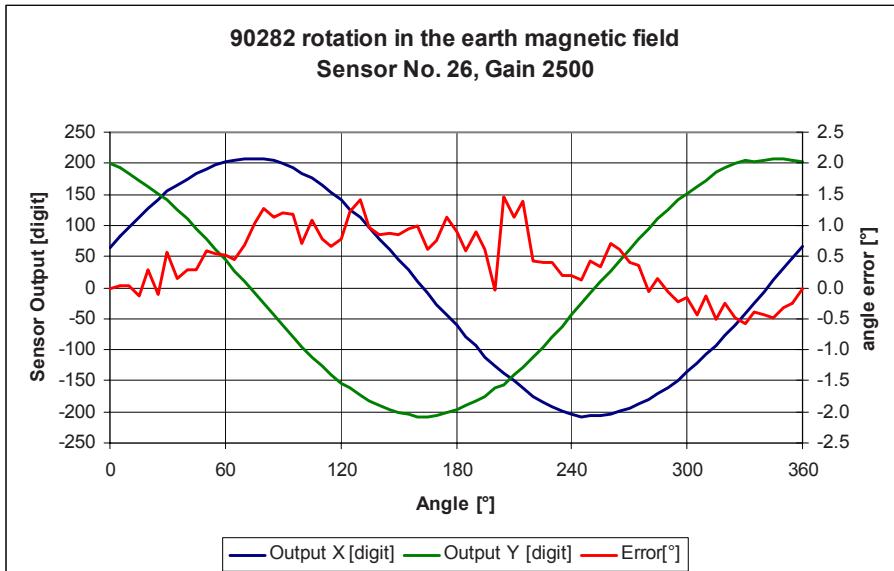


Fig. 16 Output angle error over 360° rotation in the earth magnetic field

Temperature stability is a very important issue, since the compass sensor shall be used in automotive environment within the temperature range from -40°C to $+85^\circ\text{C}$.

First measurements on offset drift over the temperature range from -25 to $+100^\circ\text{C}$ at a sensitivity set to $7.25\text{LSB}/\mu\text{T}$ show a value of 150 LSB on X and Y (Fig. 17). This drift is relatively high compared to other recent work [7] and it has its origin in the variation of local stress with temperature under the edge of the IMC where the Hall elements are placed.

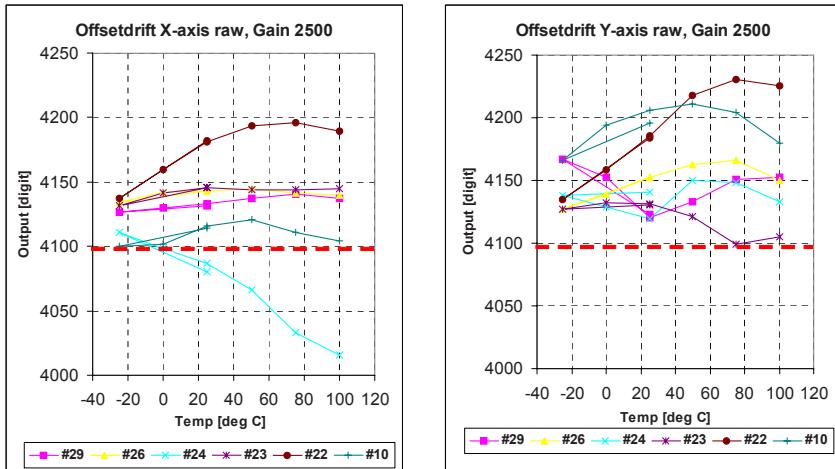


Fig. 17 Offset-drift versus temperature for X (left) and Y (right)

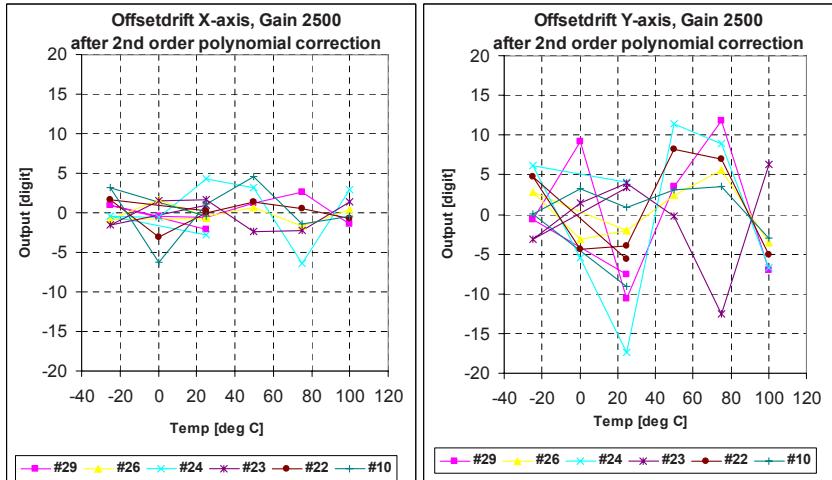


Fig. 18 Residual offset drift after second-order polynomial compensation

However, thanks to the microcontroller, second order polynomial compensation reduces the drift to about 5 LSB for X, but only to 15 LSB for Y (Fig. 18). An improved frontend architecture with more efficient offset compensation and a modified assembly process will be implemented in the next version to reduce this drift.

The sensor's nonlinearity was measured by a magnetic field sweep of $\text{FS} = \pm 500\mu\text{T}$ to below 1%FS. The onset of magnetic saturation of the IMC is observed above 1mT.

7. Conclusions

Even though silicon Hall devices do not at all seem to be a good choice for earth magnetic field measurement, they turn out to be well performing when combined with an integrated magnetic concentrator structure and adapted frontend electronics. In this challenging task, the on-chip microcontroller is essential for the implementation of digital compensation schemes for temperature drift and matching errors.

Possible host systems for such a miniature complete electronic compass system range from watches to mobile phones to automobiles. Due to its three-axis feature, the compass can either be used as conventional heading sensor, or as a recalibration reference for gyroscopes in more complex systems, or even as a three-axis low-field transducer in mapping applications

References

- [1] C. Schott, R. Racz, S. Huber, A. Manco, M. Gloor, "A CMOS Single-Chip Electronic Compass with Microcontroller", Proc. of Internat. Solid State Circuit Conference 2007, pp. 382-383
- [2] A. Bilotto, G. Monreal, R. Vig, "Monolithic magnetic Hall sensor using dynamic quadrature offset cancellation", IEEE Journal of Solid-State Circuits, Volume 32, Issue 6, June 1997, pp. 829-836
- [3] Patent application EP772046W
- [4] R. S. Popovic, R. Racz, C. Schott, "A new CMOS Hall angular Position sensor", tm – Technisches Messen, 68, June 2001, pp. 286-291
- [5] R. Racz, C. Schott, S. Huber, "Electronic Compass Sensor" Proceedings of IEEE Sensors 2004, 24-27 Oct. 2004, vol. 3, pp. 1446-1449
- [6] Pieter Rombouts and Ludo Weyten, "A versatile Nyquist-rate A/D converter with 16-18 bit performance for sensor readout applications", ISSN 0167-9260- 2005, vol. 39, no1, pp. 48-61
- [7] J. van der Meer et al., "A Fully-Integrated CMOS Hall Sensor with a $4.5\mu\text{T}, 3\sigma$ Offset Spread for Compass Applications", Proc. of ISSCC 2005, Feb. 6-10, San Francisco, USA, pp. 246-7

Protection and Diagnosis of Smart Power High-Side Switches in Automotive Applications

Andreas Kucher
Infineon Technologies Austria AG
A-9500 Villach – Austria
Andreas.Kucher@infineon.com

Abstract

In automotive power applications protection and diagnosis is getting more and more important. Main reasons for this trend are safety requirements, high reliability and complex power management of modules in a car. This paper describes aspects and challenges of designing smart power high-side switches embedded in automotive systems to drive inductive, capacitive and resistive loads.

1. Introduction

Within the last few years the number of power modules in the automotive environment has rapidly increased. Smart power systems had been developed to solve essential tasks. To be able to manage this in a proper way a lot of diagnostic feedback is required. Smart power high side drivers are not simple switches any longer, but they also have to provide information about power and fault conditions. Beside accurate information the system has to withstand fault conditions like short circuit, reverse battery, inverse operation or overvoltage. The design of circuits within a tough automotive environment is challenging design technique and technology as well. The on-resistance of switches is typically in the range of 6 to 200 mOhm. Most of the applications require PWM operation with frequencies between 50 and 200 Hz typically to serve the needs of loads like bulbs, small motors or heating systems.

2. Application and Block diagram

A typical application is shown in figure 1. On the left hand side you can find the vehicle's battery including the resistor for the inner resistance of the battery plus the line resistance from battery to the application module. A line inductance which is typically $5\mu\text{H}$ represents the wiring harness of the module's supply line.

The ECU (electronic control unit) itself contains a microcontroller, with the appropriate voltage regulator and the smart power high side switch. The green box in figure 1 shows the basic circuit blocks of an intelligent switch which includes interface logic, diagnosis and protection circuitry and last but not least the power switch itself. In most cases the switch is an n-type MOS transistor, due to the better $\text{Ron} \times \text{A}$ performance compared to p-type devices.

Additional components like capacitors, diodes and resistors are present for protection reasons. Some of these components can be left out, if the protection functions of the switch are able to handle high voltages and/or currents.

On right hand side of figure you can see the load including the line impedance.

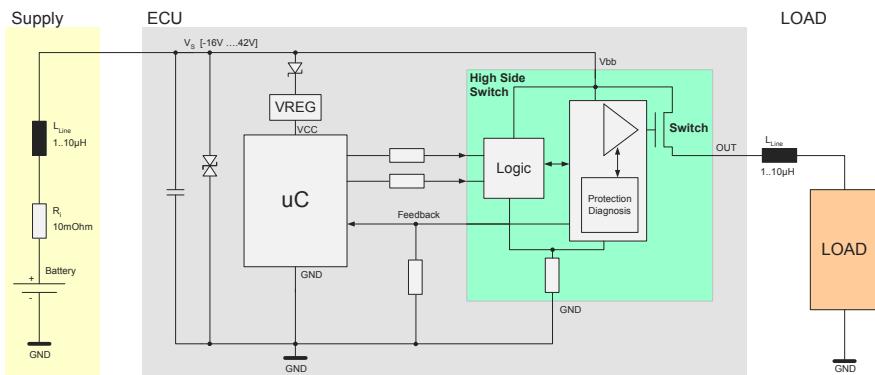


Fig. 1: System overview for smart power high side switch application

Figure 2 illustrates typical ground related loads for smart high side driver devices.

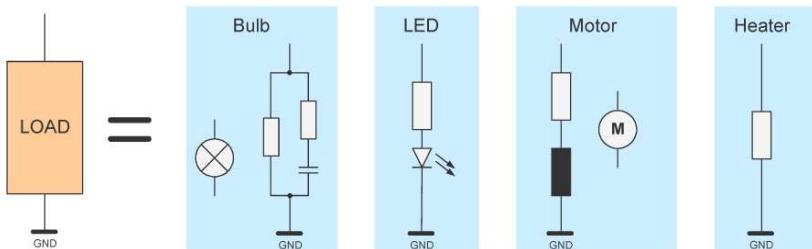


Fig. 2: Various loads for smart power high side switches

Real applications need to handle a comprehensive variety of different loads. Figure 3 shows an example of a system switching more than 500 Watts of output load.

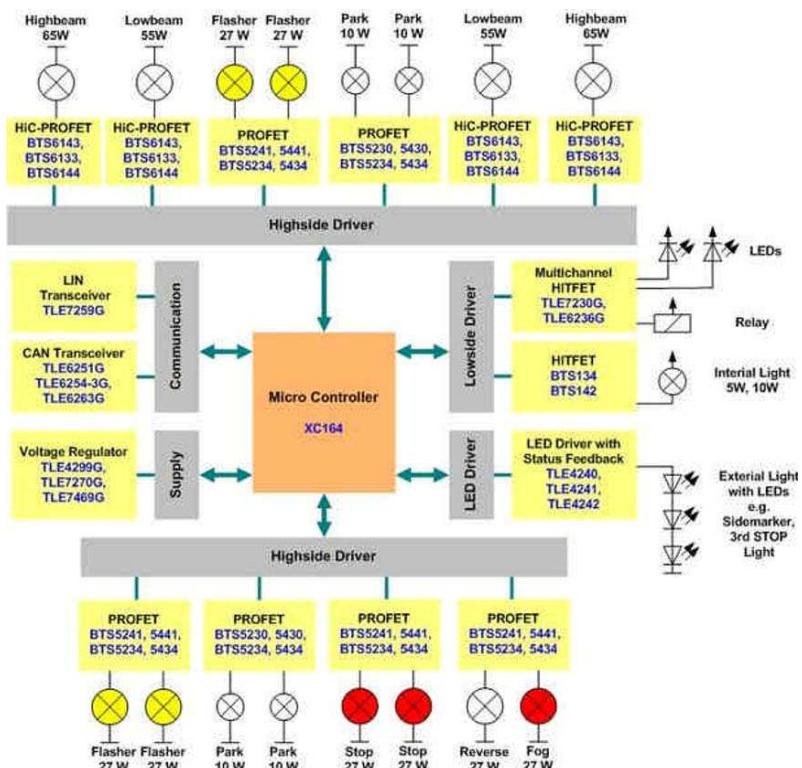


Fig. 3: Block diagram of light application

In the next step we are going to discuss the partitioning of functions for a smart power high side switch device on the example of a five channel high side switch.

The basic architecture of the device contains the main blocks chargepump, DMOS driver stage, logic interface, diagnosis and protecting functions against overvoltage, overcurrent and overtemperature.

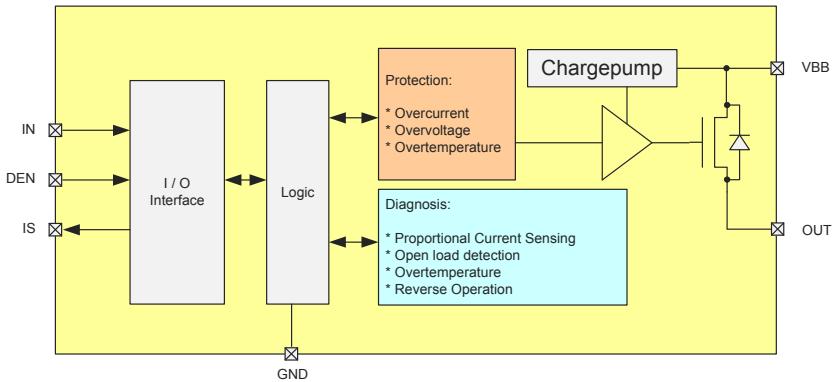


Fig. 4: Block diagram of high side switch

Later in this chapter some of the most interesting blocks are described in more detail.

3. Technology

High side switches are basically done monolithically or in a system on chip (SoC) way (i.e. chip-by-chip (CbC), chip-on-chip (CoC)). Today junction-isolation BCD processes with p- or n-substrate are commonly used in automotive applications whereas more integrated wafer technologies are based on p-substrate.

The power transistors are implemented as lateral or vertical DMOS transistor. The product of R_{ON} times Area is determined by the DMOS concept. The “up-drain” concept allows drivers with high R_{ON} only. Better in $R_{on} \times \text{Area}$ are trench-DMOS concepts.

Figure 5 shows the basic devices of a BCD [1] technology for a monolithic integration of a complex driver system. The implementation of a low ohmic device in BCD technology would require a higher effort in area and costs. Another field of application for BCD processes is the realization of chips for CbC or CoC solutions.

Since the substrate is connected to ground level, the circuits must be designed very carefully, if signals might go below ground. Any n-epi node in the technology creates a diode, which injects substrate current due to a negative n-epi node voltage. This event may lead to a functional disturbance due to the activation of parasitic substrate NPN structures.

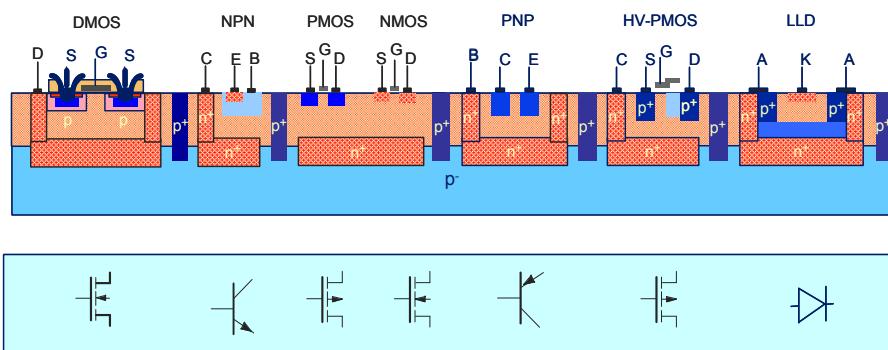


Fig. 5: Cross section of BCD devices with p-substrate

Low resistive power switches are implemented as discrete devices in DMOS trench technologies. In this case the substrate is n⁺ doped and the backside of the

die is connected to the drain of the power DMOS. This kind of technology can be utilized for monolithic power devices and control chips as well.

Figure 6 shows the cross section of devices with n^+ substrate.

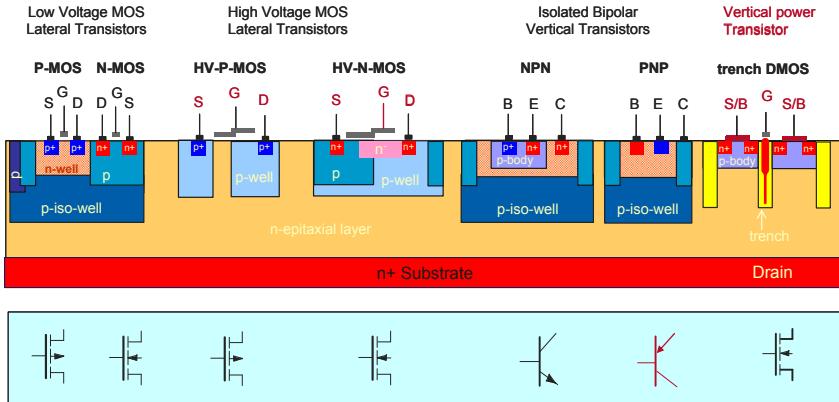


Fig. 6: Cross section of HV_Cmos devices based on n^+ substrate

The decision to use CoC, CbC vs. a monolithic approach is often answered by the total costs of a product. However further considerations like the required complexity or a limited leadframe area must be taken into account.

Typically the monolithic IC design [2] of smart power high side switches is accomplished in wafer technologies based on n^+ substrate. Thus power and logic is implemented within one technology, however the design complexity is limited due to the feature size of such a technology.

The limit of monolithic design is in most cases driven by cost and the possibility to fit into a small package. Here the CoC/CbC approach has the advantage to be able to use the best DMOS properties combined with a high voltage BCD or CMOS process.

4. Diagnosis requirements

The most important information for power switches is to know how much power is being consumed at a certain time. Therefore the designers responsibility is the correct supervision of voltage, current, time and/or the resulting silicon temperature.

Which basic diagnostic feedback signals are reported to the micro controller?

- 4.1. Current sensing – proportional current including overcurrent detection
- 4.2. Overtemperature sensing

In this chapter current sensing and overttemperature sensing is considered.

4.1 Current sensing

The proportional current sensing in protected high side smart power switches is done according the block diagram in figure 7. The concept is based on “active current mirror”.

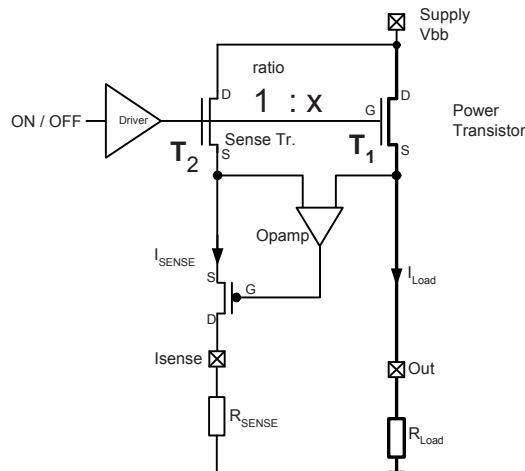


Fig. 7: Active current mirror concept

On the right hand side there is the power transistor T_1 with a given gate source voltage $V_{GS}(T_1)$. Transistor T_2 is the sense transistor. The geometric dimension of T_1 is x times larger than the one of T_2 . The Opamp is responsible for regulating the source of T_2 to the same voltage than the source of T_1 . This means the gate source voltage is the same as well. Under that condition the drain current of T_2 is proportional to the drain current of T_1 divided by the factor “ x ”.

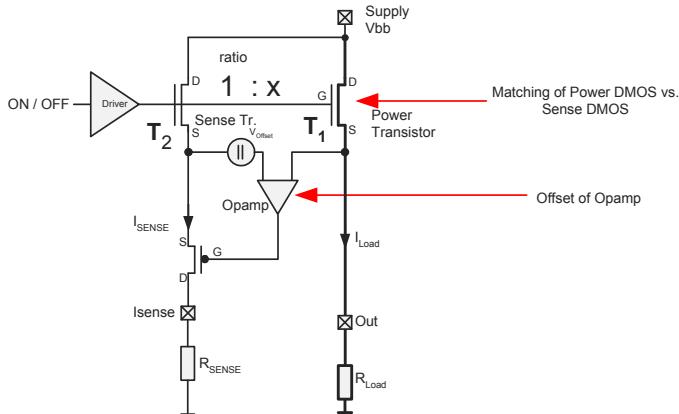


Fig. 8: Block diagram of current sensing

To do the calculation the resistance of T₁ and T₂ has to be calculated.

$$R_{\text{Sense}} = V_{DS(T2)} * I_{\text{SENSE}} \quad (1)$$

$$R_{\text{Power}} = V_{DS(T1)} * I_{\text{Load}} \quad (2)$$

Assuming that the matching of the power transistor to the sense transistor is perfect we can set up the following equation:

$$I_{\text{Sense}} * R_{\text{Sense}} = I_{\text{Load}} * R_{\text{Power}} \quad (3)$$

The ratio is also called kilis. The kilis is the factor of load current I_{Load} divided by I_{Sense}.

Ideal equation to calculate the k_{ilis0}:

$$k_{\text{ilis}(0)} = \frac{R_{\text{Sense}}}{R_{\text{Power}}} = \frac{I_{\text{Load}}}{I_{\text{Sense}}} \quad (4)$$

In reality there are two major influences.

- a. the mismatch of the power transistor to the sense transistor
- b. the offset voltage of the operational amplifier

The mismatch of the power transistor to the sense transistor is given by technology and the operating point of the DMOS. From the view of design nothing can be done against the mismatch of geometry of the two DMOS Transistors T₁ and T₂. When considering the operating point of the DMOS the gate source voltage and the drain current are influencing the matching accuracy. Once the gate source voltage (V_{GS}) is high enough the mismatch is determined by the geometry. In case of lower V_{GS} the matching accuracy is getting worse.

The second important parameter is the offset of the operational amplifier. This amplifier has to regulate the drain to source voltage of T₂ to the same value as the drain to source voltage of T₁. The drain source voltage of T₂ is given by equation 5:

$$V_{DS(T2)} = R_{Sense} * I_{SENSE} \quad (5)$$

This means at low load currents (I_{Load}) the V_{DS} of T₂ is very low.

Figure 9 shows the influence of offset voltage and DMOS mismatch at high V_{GS} voltage.

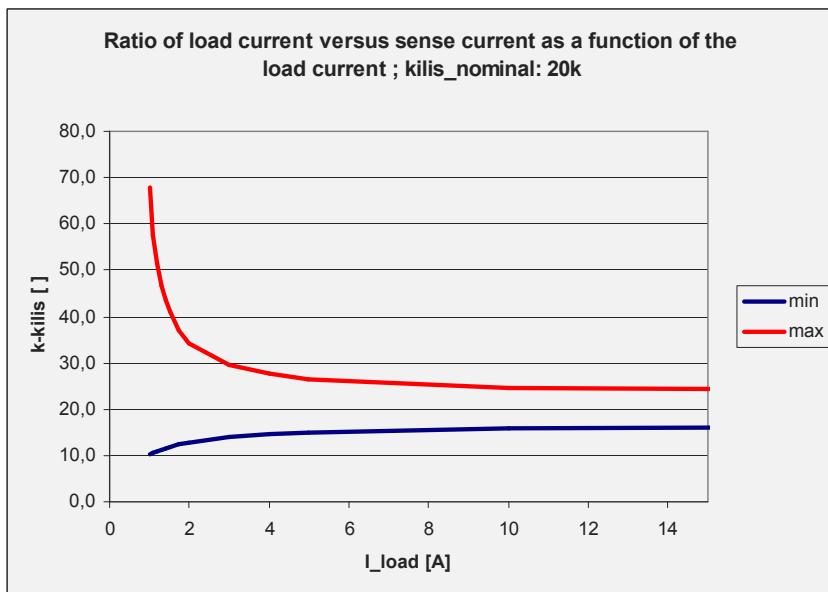


Fig. 9: Error of kilis caused by opamp and DMOS mismatch

At low load currents the V_{DS} of the sense transistor T_2 is very low. The drain current of T_2 is in the hundred of μA region. The V_{DS} of T_2 is getting close to the offset voltage of the opamp.

Improvement to increase the sense accuracy at low load currents:

To avoid an operation with very low drain to source voltage during the load current measurement, actions have to be taken to keep the V_{DS} of the power DMOS (T_1) in a range where the offset of the current sensing opamp can be neglected. An additional operational amplifier (Opamp 2) can be implemented for improvement.

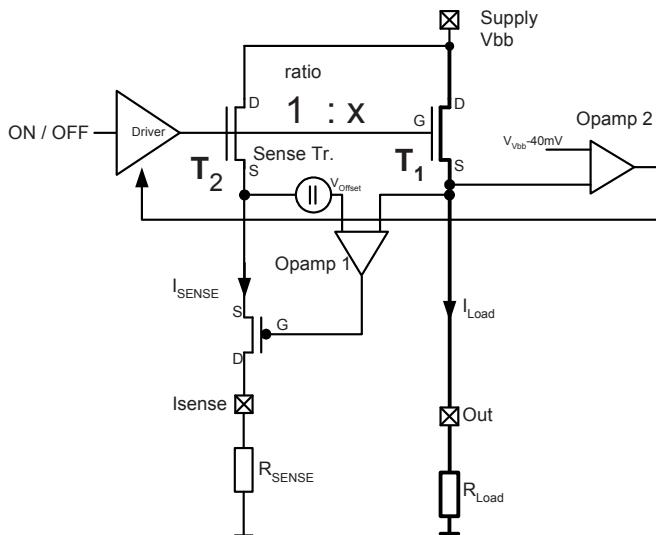


Fig. 10: Block diagram of improved current sensing

In order to avoid a very low V_{DS} the source voltage of the power transistor is observed. Once this V_{DS} is below a preset value (i.e. 40mV) the gate source voltage of T_1 and T_2 is reduced by the gate driver stage. A reduced V_{GS} leads to a higher on-resistance of T_1 and T_2 and the offset voltage of opamp 1 is less critical. On the other hand at reduced V_{GS} the matching of T_1 to T_2 is worse. At the end it is a trade off because the offset mismatch is dominant compared to the change of the transistor mismatch at low V_{GS} . This method is called gate back regulation (GBR) because the gate voltage of T_1 and T_2 is reduced.

Figure 11 shows the improvement of the current sensing circuit.

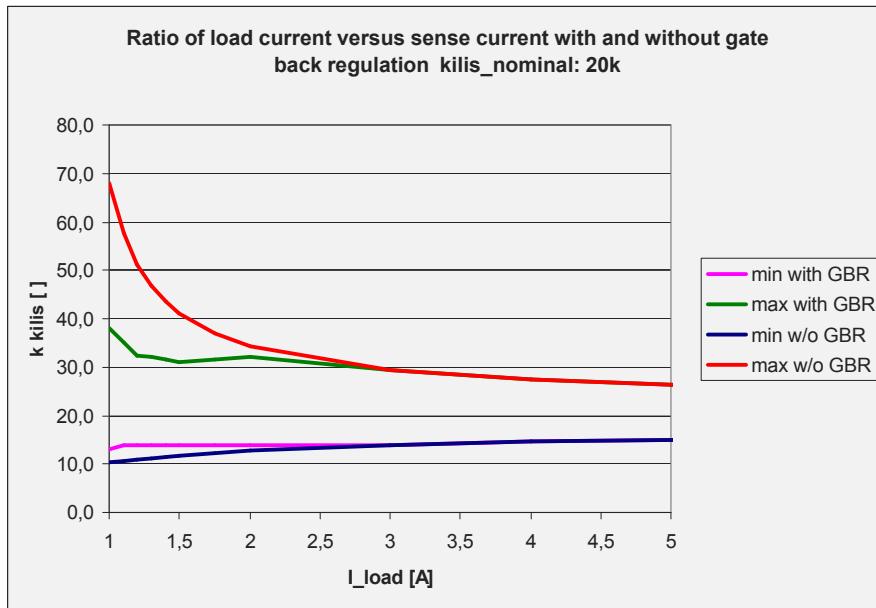


Fig. 11: Comparison of current measurement accuracy with and without gate back regulation circuit (GBR)

5. Protection

In this chapter the protection of the device against overvoltage, overcurrent and overtemperature is described.

5.1 Overtemperature sensing

Overtemperature protection is implemented not to exceed the maximum temperature of a device. Once the temperature has reached the overtemperature limit the circuit has to be shut down. After turning off there is no power dissipated and the device is cooling down. If the temperature falls below an uncritical value it is possible to reactivate the switch again.

Overtemperature is caused by an overload condition of the device, especially during current limitation with high power dissipation. The power dissipation is given by.

$$P_{\text{loss}} = I_{\text{current_limit}} * V_{\text{DS}} \quad (6)$$

Once the overtemperature sensor has reached the overtemperature threshold the device turns off and cools down. The temperature sensing circuit has a built in hysteresis. When the junction temperature is below the threshold minus a hysteresis value a restart of the power stage is possible.

A situation could occur that keeps the device in OFF state unless a device reset is sent. This latching overtemperature shutdown (see figure 12) causes low temperature stress to the device, but for some loads this behaviour might be a disadvantage. In such a case the load could be a bulb with higher inrush current than expected leading to premature shut down of the power switch and avoiding a bulb illumination.

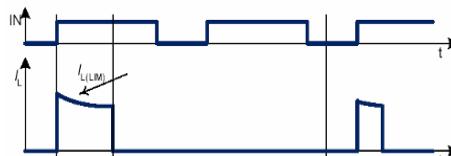


Fig. 12: Latching over temperature shutdown

A further concept for a toggling overtemperature shut down with additional reduction of the power consumption after restart is demonstrated in figure 13. This keeps the temperature stress to a minimum whilst shutting down the device. In the proposed solution the overcurrent threshold is set to a lower value after the first overtemperature event to avoid hazardous overheating of the die.

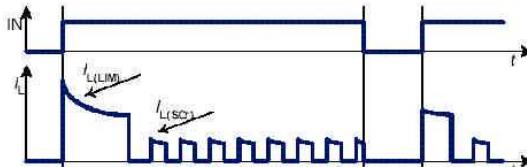


Fig. 13: Toggling over temperature shutdown with reduced power

For accurate temperature detection, the position of the temperature sensor is an important issue. Therefore transient thermal simulation helps to determine the ideal position of the temperature sensor. Figure 14 shows the result of a thermal simulation in which the temperature sensor is placed close to the hot spot of the silicon. The optimal position for placement of the temperature sensor in figure 14 is in the lower middle position of the picture.

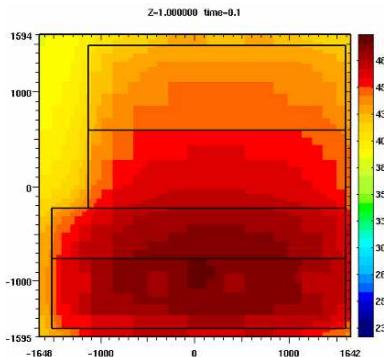


Fig. 14: Transient thermo simulation of a power DMOS

5.1.1 Temperature sensors

In most cases temperature sensors are based on p-n junctions or resistors. A p-n junction can be biased in forward or reverse direction. The voltage drop at constant current flow of a forward biased junction is used to measure the

temperature. Another method utilizing a reverse biased junction relies on the principle of increasing leakage currents at high temperature to realize a very simple temperature sensor (see figure 15).

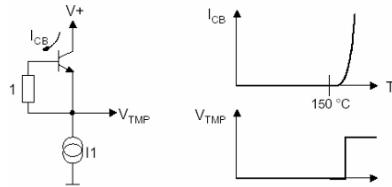


Fig. 15: Leakage current temperature sensor

The leakage current depends on doping while the temperature threshold is more or less fixed to a certain value for a given p-n junction.

$$I_s = qA \left[\frac{D_p}{L_p} \frac{1}{N_D} + \frac{D_n}{L_n} \frac{1}{N_A} \right] n_j^2 + qA \frac{w}{\tau_0} n_j^2 \quad (7)$$

More accurate results can be expected from a p-n junction which is forward biased however the need of a voltage reference has to be taken into consideration.

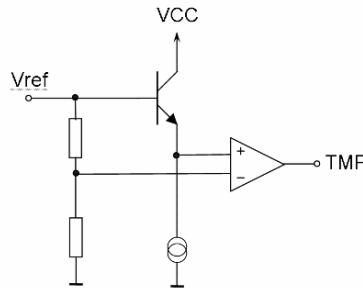


Fig. 16: PN junction forward biased temperature sensing

The positive or negative temperature coefficient of integrated resistors is the core of resistor based temperature sensors. The accuracy depends on production tolerances and is worse than the accuracy of the bipolar transistor.

5.2 Overvoltage requirements

In automotive environment ESD and ISO pulse robustness is required to fulfill the car manufacturer's specifications. For pins having a connection to wires outside the ECU a HBM robustness (human body model) of 4kV is demanded.

The definition of the ISO pulse shapes can be found in the ISO 7637 Standard. The pulses include voltages up to +/- 200V at supply pin directly at the device. Because the voltage class of typical automotive technologies is below 100V the device could be damaged and therefore it has to be protected. To avoid this kind of injury a decoupling network to ground has to be arranged.

Figure 17 shows implemented clamping structures to avoid damage to the device during an ISO pulse stress applied on VBB pin. Once the voltage from VBB to GND exceeds the maximum ratings of the technology the Z-Diode clamping circuitry is getting active. The ground inside the device and the ground pin are connected via a resistor which drops overvoltage during ISO stress to achieve the required robustness.

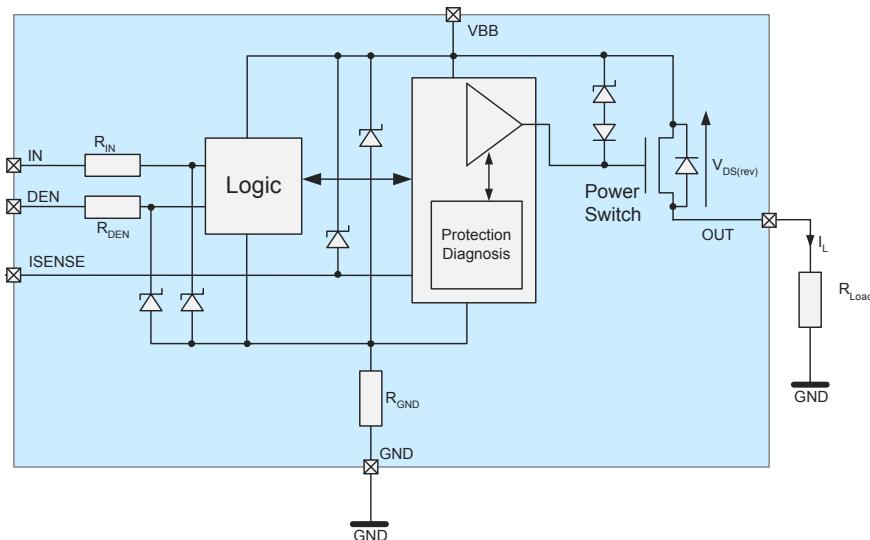


Fig. 17: Block diagram of device showing Z-Diode clamping

Car manufacturers require reverse battery robustness up to two minutes without any damage of the devices therefore reverse polarity is tested and must be withstood.

The behaviour of the available technology has a big influence on the device. During battery reversal the reverse diode of the power switch gets active and dissipates power given by the current in reverse operation multiplied by the forward voltage of the reverse diode of the power transistor.

The dissipation during this event is:

$$P_{\text{diss}(\text{rev})} = \sum (V_{\text{DS}(\text{rev})} \cdot I_L) + \frac{V_{\text{bb}}^2}{R_{\text{GND}}} \quad (8)$$

Depending on package and power dissipation a certain temperature rise will occur which needs to be reviewed in terms of the maximum allowed temperature. If the temperature is below 150°C no additional effort has to be done but if the temperature is too high the power transistor has to be switched on during reverse battery operation. With a lower V_{DS} of the power transistor also the power dissipation is decreased.

5.2.1 Loss of ground and inductive clamping

When switching inductive loads the source output of the DMOS is forced to negative voltage. The limit is the maximum permitted drain to source voltage of the DMOS or the voltage capability of the driver stage.

The substrate potential is connected to the positive supply in case of a monolithic die based on n⁺ substrate. The voltage between the supply pin and the source output of the DMOS has to be below the maximum ratings of the technology.

For chip-by/on-chip systems also p-substrates are used for driver ICs. The design of such a driver stage needs to be done very carefully because the p⁻ substrate is connected to ground. To prevent breakthrough of devices additional clamping structures from IC ground to DMOS output are included.

Why is there a need of a negative clamping voltage? One of the main reasons is the loss of the module ground. In this case on one hand the IC ground is pulled up to positive supply voltage and on the other hand the output of the DMOS is pulled down to ground level by the low resistive load.

5.3 Overcurrent

One of the most challenging failure modes of high current devices is over current. Here we have to distinguish between overcurrent limitation and overcurrent shutdown. Both methods have pro's and con's for certain applications.

Beside the fault condition (i.e. short circuit) the behavior of the load is also significant. Switching on a bulb means a high inrush current. To avoid overcurrent detection the limit has to be set to a value which is about 10 times higher than the nominal current. Figure 18 shows an example of a 25W bulb current during inrush.

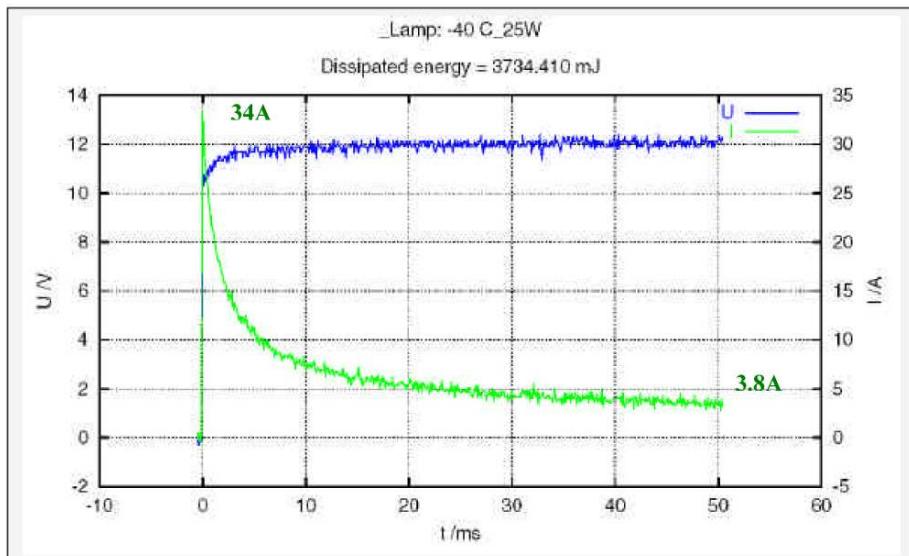


Fig. 18: Inrush current for a 25W bulb

For bulb applications the current limitation principle is more common. During an overcurrent event the maximum power dissipated inside the switching device depends on the voltage and the overcurrent limitation threshold. For a fixed overcurrent threshold the power dissipation increases linear with V_{DS} .

An improvement can be achieved by a multi-stage current limitation, where the current limitation value depends on the drain to source voltage of the power switch. With this concept the peak power dissipation can be reduced. Furthermore the device can be sure not to report an overcurrent condition to the microcontroller during turning ON.

Figure 19 shows the current limitation threshold. At low V_{DS} the threshold is higher than the inrush current of a bulb. To avoid high power dissipation inside the DMOS the overcurrent threshold is reduced at high V_{DS} .

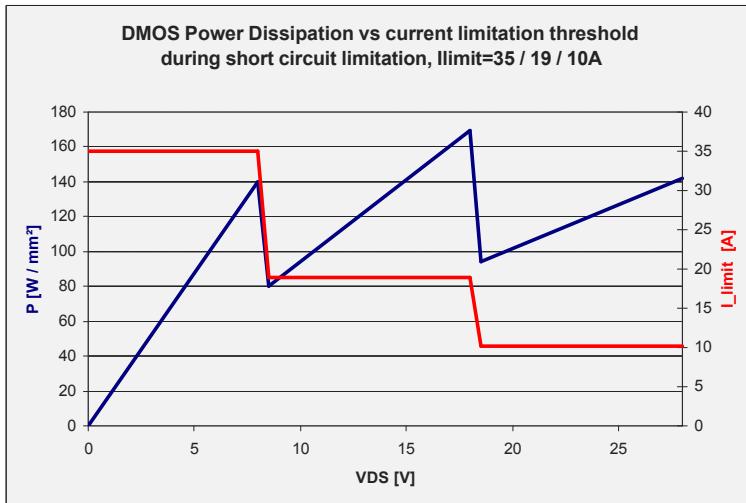


Fig. 19: current limitation depending on V_{DS}

For very low ohmic high side switches with very high overcurrent thresholds a overcurrent shutdown is needed because the dissipated power is too high to handle the event by limiting the load current. Figure 20 shows the setup of such a short circuit test to test. The purpose of this device is to show the short circuit robustness of the device.

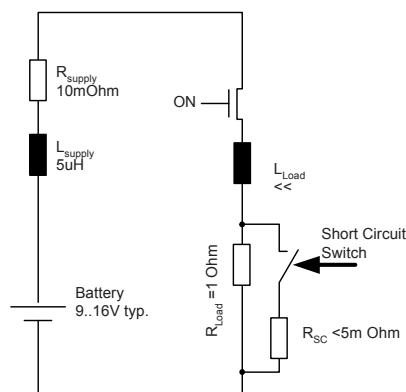


Fig. 20: Short circuit test setup

Figure 21 shows the result of the test described in the setup of figure 20. The device under test was a 6 mOhm high side switch.

The current rises up to about 200A in less than 10 μ s. To avoid an inadvertent shutdown some glitch filters have to be implemented. It can be seen that a delay of less than 5 μ s can cause a change of the DMOS load current of about 150A.

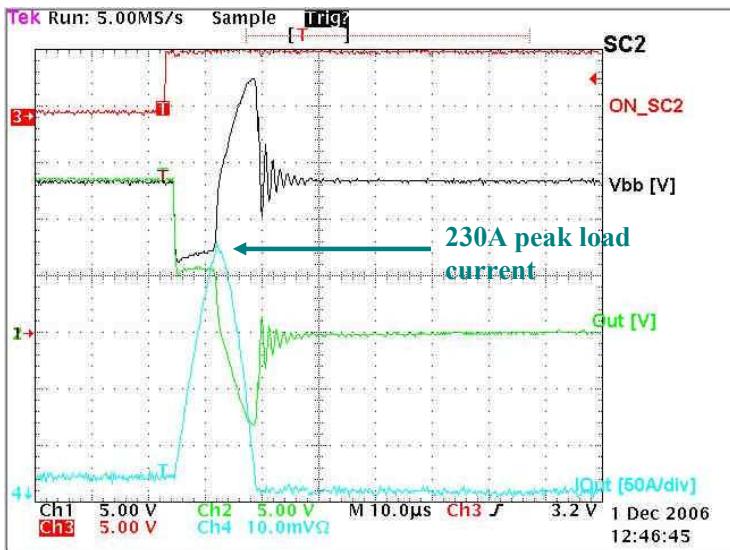


Fig. 21: Voltages and currents during short circuit type 2

Figure 21 shows the fast transient behaviour during an overcurrent event. The combination of very high current and fast transient makes the design of such circuitry difficult. This gives room for additional improvement of protection concepts in future designs.

6. Conclusions

Protecting a device in automotive environment is difficult. High voltages and currents cause a lot of challenges. The diagnosis for a power switch is quite important for near future. Here the requirements are well known. Improvement has to be done in terms of accuracy for an optimized power management. For protection purposes the effort for robust and reliable design is increasing significantly. Nevertheless the consecutive technology shrink and new assembly processes of devices lead to a smaller power transistor area and increased power density. To improve the robustness and diagnosis capability of smart high side transistors additional effort in design is needed.

References

- [1] M. Stecher et al., "Key Features of a Smart Power Technology for Automotive Applications", Conference on Integrated Power Systems CIPS 2002, Bremen, Germany
- [2] B. Murari, F. Bertotto, G.A. Vignola (eds.), "Smart Power ICs: Technologies and Applications", Springer 1995
- [3] Infineon, "Bridging Theory Into Practice", Fundamentals of Power Semiconductors; ISBN-0-9789866-0-1

Part II: Integrated PA's: from Wireline to RF

One of the bottle necks in fully integrated circuit systems is getting the signals out. As a result output drivers are becoming the key building block of integrated systems. In this part, different output drivers covering kHz up to 60 GHz are discussed. The first three papers deal with RF Power Amplifier output structure. The last three papers deal with wireline drivers. Each of them have their own specific requirements and design challenges.

The first RF PA discusses the trends towards linear Broadband amplifiers. Different architectures are presented. Special care in the power distribution network becomes an important role towards stability, power supply rejection and common mode gain. The hype to more Moore would suggest to go to very high frequencies (60GHz and higher), but may be no more more Moore techniques, such as MIMO UWB chip solutions, can bring nice solutions as well.

The second RF PA paper goes indeed towards 60GHz and beyond. Since modern systems require more power and high efficient linear PA's, linearization techniques are analyzed. Several techniques such as LC-type combining and planar transformer techniques are described. Several fully integrated examples up to 60 GHz in standard CMOS are presented.

The third RF PA work evaluates efficient integrated power amplifiers based on burst mode techniques for the linearization. By using these techniques efficient PA's can be achieved but due to the burst approach care in the spurious have to be taken. By using Self Oscillating Power Amplifier (SOPA) techniques in combination with the filtering properties of the SAW antenna filter, it is demonstrated that telecom performances are within reach.

The fourth paper describes high speed serial interface solutions. Modern systems, such as GSM, laptops, etc require a lot of communication between the different subsystems. In order to reduce cost, pin count and improved speed performances, instead of the classical bus communications, the trends is towards high-speed serial interfaces. This of course requires drivers which are efficient both when low date rate as high date rate is transmitted. For that line drivers which can handle both 'classical' digital CMOS levels (low data-rate) as drivers for terminated differential transmission lines (high data-rate) are developed and discussed.

The fifth paper discusses xDSL drivers in nanometer technologies. Those line drivers require high output voltages to reduce the transformer ratio. On top high efficiency is required. In this work the Self Oscillating Power Amplifier in combination with dynamic-biased high output voltage stages in nanometer technologies are described. It is shown that high voltages (up to 7 V) can be achieved in standard 1.2 V nanometer CMOS technologies without violating the 1.2V technology breakdown requirements. As such efficient xDSL line drivers in nanometer technologies have been achieved.

The last paper deals with SLIC (Subscriber Line Integrated Circuits) drivers. Although the modern communication systems, still the Plain Old Telephone Systems (POTS) are in many places in use. Also more integration and better efficiency in the drivers are required. The high voltage requirements, ESD protections in combination with the ugly line characteristics are major challenges. Due to the combination of CMOS and DMOS devices single chip SLIC-CODEC combination can be achieved.

Michiel Steyaert

Integrated CMOS Power Amplifiers for Highly Linear Broadband Communication

K. Mertens¹, M. Unterweissacher², M. Tiebout¹, C. Sandner¹

¹Infineon Technologies GmbH Development Center Villach, Austria

²Department of Electronics, TU Graz, Inffeldgasse 12, 8010 Graz, Austria

Abstract

In this paper we will explain the trends in short-range broadband communication and investigate the technology requirements necessary for implementing the PA building block. Further we will discuss the importance of the power grid on the stability of the PA. Finally a single-ended integrated PA, part of a WiMedia/MBOA compliant UWB transceiver, is presented.

1. Introduction

In the recent years there is a trend to go to high data rate transceiver chips for short-range communication. All these transceivers have in common that they increase their bandwidth while keeping the radiated output power modest. Hence from Shannon, we know that this is a way to maximize the digital data throughput of the communication channel.

$$C = B \cdot \log_2 \left(1 + \frac{P_o}{kT} \right) \quad (1)$$

Seen the huge bandwidths, multi-carrier modulation schemes, like orthogonal frequency-division multiplexing (OFDM) or discrete multi tone (DMT), are being used. Both techniques split a highly selective wide band channel into a large number of non-selective narrow-band channels. Hence, the effect of multi-path propagation –fading– can be easier tackled, since the symbol duration is extended by the factor N equal to the number of sub-carriers used. The longer symbol duration relaxes the equalization implementation in the receiver and makes therefore the transmitter robust against fading. From the above-mentioned two multi-carrier modulation schemes OFDM is by far the most popular one. The reason for this is that it uses the available bandwidth in the most spectral efficient way. The disadvantage of an OFDM scheme is the high peak to average output power, which emanates from the addition at the modulator of independent sub-carriers all having a random amplitude and phase. Therefore high linear power amplifiers, such as class A or AB, are required. A good example, fitting above description, is the emerging UWB specification. This standard is intended as a cable replacement technology and is therefore also

referred to as wireless USB. The IEEE.8.15 UWB standard aims to cover a communication distance up to 10m, which is suited for most indoor applications. The UWB standard allocates a frequency spectrum ranging from 3GHz up to 10GHz. This frequency range is split into five band groups, which consist of three bands each, see figure 1. Each band has a bandwidth of 528MHz, which is modulated with an OFDM modulation with 128 sub-carriers. The supported data rates are: 55, 80, 110, 160, 200, 320 and 480 Mbps. The maximum allowed output power seen at the antenna is $-41\text{dBm}/\text{MHz}$. The 1dB compression point of the UWB PA is therefore situated around 3dBm, when we take the required back-off and the external losses into account.

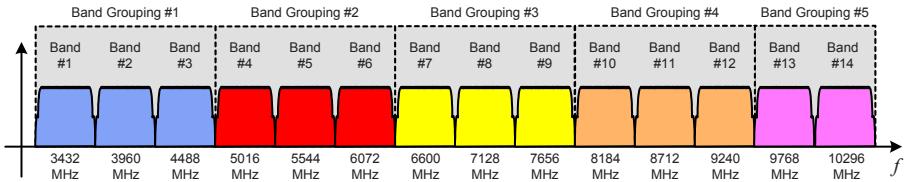


Figure 1: UWB Frequency allocation

The applications targeted with this standard are mainly intended for the consumer market and therefore a low price mainstream technology, such as CMOS or BiCMOS, is required. Possible implementation architectures for the transceiver RF front-end are zero or low IF. A crucial component in the TX chain is the modulator. The modulator is typically implemented as a Gilbert cell and has a negative conversion gain of -10dB . This means that typical gains between 20dB or 30dB are required in the PA driving chain to reach an output power of 3dBm . Note that in contrast to UWB, constant envelope systems such as Bluetooth and GSM use nowadays the frequency synthesizer (in most cases a fractional PLL loop) to drive the PA. The advantage is that less driving stages are needed for obtaining a certain output power level. For implementing the transceiver an appropriate CMOS or BiCMOS technology node must be selected. Hence, we have to examine the transistor specifications that are relevant for the PA, since it turns out that this building block is the most difficult one to realize.

2. Technology Selection

Choosing a technology able to do the job is more than half of the work in power amplifier design. Technology parameters as F_{MAX} and the Johnson product of a transistor are here the relevant specifications to consider. In figure 2 the measurements of the unilateral power gain (U) and the current gain ($H21$) on an active device are shown. The frequencies where both curves are equal to 0dB , are the cut-off frequency F_T respectively the maximum oscillation frequency F_{MAX} . Both measurements shown in figure 1 can be used for constructing an

equivalent RF small signal model. Remark that this model is the same for most existing RF devices, such as BJTs and HBTs. Hence, comparing different

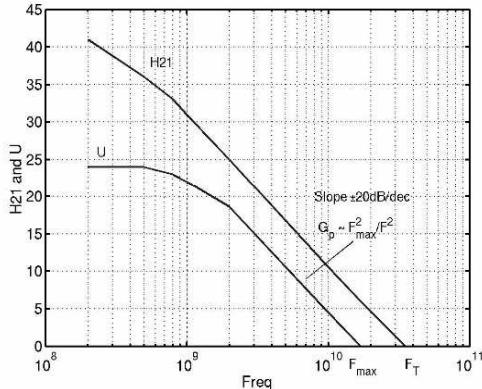


Figure 2: Measurements of the current gain $H21$ and the power gain U on a HBT

transistor technologies among one another is straightforward. Out of this small signal model the cut-off frequency F_T and the maximum oscillation frequency F_{MAX} can be calculated as:

$$F_T = \frac{gm}{C_{gs}} \quad (2)$$

$$F_{MAX} = \sqrt{\frac{F_T}{8\pi R_g C_{gd}}} \quad (3)$$

From equation 3 it can be seen that F_{MAX} depends on the time constant formed by the Miller capacitance (C_{gd}) and the gate resistance. Hence, depending on the time constant, the value of F_{MAX} can be smaller, equal or larger than F_T . The latest nm CMOS technologies can sufficiently lower the gate resistance, needed in digital circuits to improve the digital gate delay, so that F_{MAX} performance can track the F_T performance. Knowing that F_{MAX} rolls-off with a -20dB slope, let us estimate the maximal stable power gain (MSG) for a sub-micron CMOS transistors as

$$MSG \cong \frac{F_{MAX}^2}{F^2}. \quad (4)$$

The so obtained MSG will fix the minimum amount of required amplifier stages needed in the TX chain. For even a more accurate guess, the quality factors of the used inductors can be taken into account. The inductor used at the gate to tune-out the capacitance of the transistors, can be incorporated into the gate resistance, to redefine a new transistor F_{MAX} . The gate resistance and the parasitic resistance of the coil are standing in series with their reactive element and form thus a LRCR network. Transforming this LRCR network at a certain frequency into a LCR network, gives the equivalent parallel resistance of the LC tank. Remapping this parallel resistance to a resistance in series with the gate

capacitance gives the new gate resistance. The newly found gate resistance value will exceed the initial gate resistance of the device and therefore

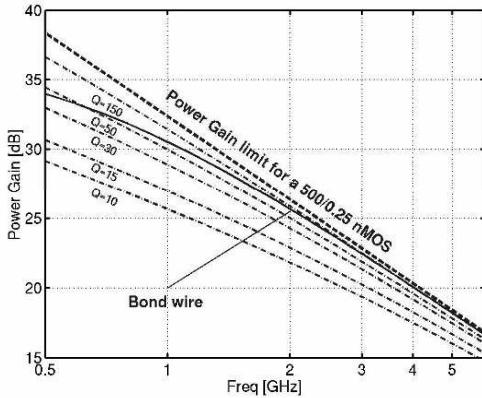


Figure 3: Power Gain in function of inductors Q factor

a reduction in maximum oscillation frequency has to occur. Hence, the quality of the used inductor will lower the power gain of the power amplifier stage. Figure 3 shows the available gain for different inductor qualities. Under the assumption that the inductors can be layouted as straight lines (slabs), we can easily predict the maximum quality that can be achieved in a CMOS process. Moreover the slabs are situated above a silicon substrate, which lowers the quality factor. The quality of the slabs can be expressed by following product [1].

$$Q_{\text{slab}} = Q_{Rs} \cdot \text{Substrate_loss_factor} \quad (5)$$

The last term in above expression represent the degradation in quality due to the substrate and has a value between zero and one. On the other hand the first term represents the quality of an inductor with series resistance R_s . The skin effect increases the series resistance of the inductor at higher frequencies, due to the non-uniform current distribution. A formula often used for estimating the frequency dependency for this series resistance is

$$R_s = \frac{L}{W \cdot \delta \cdot \sigma \cdot (1 - e^{-(T/\delta)})}, \quad (6)$$

where the skin depth δ is defined as

$$\delta = \sqrt{\frac{2}{\omega_0 \cdot \mu_0 \cdot \sigma}}. \quad (7)$$

Slabs realized in a CMOS process, by stacking metal layers, have a typical thickness T of around $7\mu\text{m}$. Hence, the skin effect is stronger for slabs than for bonding wires with a diameter of $25\mu\text{m}$. Increasing the width of the slab reduces

the DC-resistance and thus elevates the obtainable quality factor Q_{RS} , as can be seen from underneath expression.

$$Q_{RS} = \frac{\omega_0 \cdot l_s}{R_s} \quad (8)$$

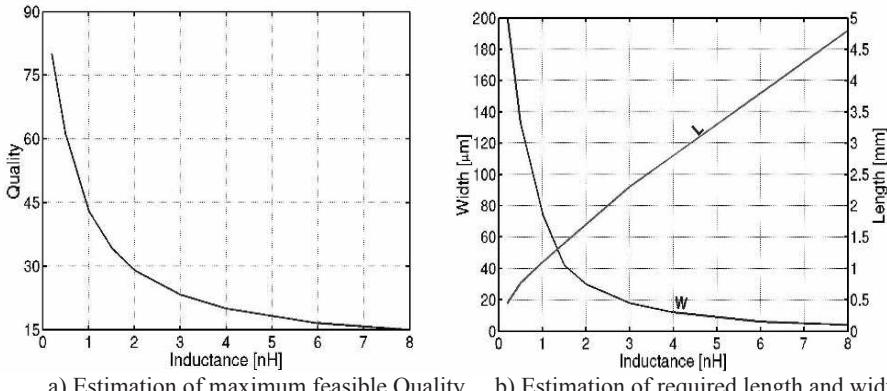
However, enlarging the slab width will cause more harmful substrate effects and as a consequence a reduction in the substrate loss factor will be observed. The above mentioned substrate loss factor is given by

$$\text{Substrate_loss_factor} = \frac{R_p}{R_p + [Q_{RS}^2 + 1] \cdot R_s}, \quad (9)$$

and is depending on the equivalent substrate resistance R_p . Assuming that the underside of the chip is conductive glued on to the substrate, makes that R_p can be calculated as

$$R_p = \frac{T_{ox}^2}{\epsilon_{ox}^2 \cdot \omega_0 \cdot W \cdot L \cdot \rho_{sub} \cdot T_{sub}} + \frac{\rho_{sub} \cdot T_{sub}}{W \cdot L}. \quad (10)$$

Large slabs (inductance values) can only be realized area efficient when they are winded up to a planar coil geometry. When we leave out second order effects, such as mutual coupling between the windings and eddy currents, than the above formulas let us easily calculate the maximum quality achievable for a specific inductance value. The result of this calculation is shown in figure 4 for a CMOS process with a substrate resistivity of $20 \Omega\text{cm}$.



a) Estimation of maximum feasible Quality b) Estimation of required length and width
Figure 4: Obtainable Quality and required dimensions for Integrated inductors on a $20 \Omega\text{cm}$ substrate

Increasing the substrate resistivity to values exceeding $20 \Omega\text{cm}$, will certainly help in achieving higher qualities for the integrated inductors. However, increasing the substrate resistivity will lead to latch-up. For standalone power amplifiers, made out of purely nMOS transistors, a higher substrate resistivity is not harmful, but for integrated PAs this would certainly result in an unacceptable yield due to the digital logic (pMOS transistors).

Consequently, the main advantage for RF CMOS would than be violated, since the goal is to combine the analog RF circuits together with digital logic. Hence further improvements in inductor quality by increasing the substrate resistance in a standard digital process are out of the question. However, for a CMOS application that uses a broad bandwidth, the quality factor could be kept relative low. The needed inductor quality Q for implementing a UWB band group bandwidth of around 1.5GHz is around three ($Q=f/BW$). The conclusion is that for broadband communication systems the transistor dominates the gain of an amplifier stage. So, wideband amplifier systems will benefit from CMOS generations to come. This is in contrast to narrowband systems working at a lower carrier frequency. These systems are limited by the quality factor of the integrated inductors and can't exploit the full transistor performance.

For the output stage not only F_{MAX} is of relevance, but also the supply voltage. Therefore the well-known Johnson limit [2], will give an indication of the performance that could be achieved in the last amplifier stage. The Johnson product doesn't improve when scaling down the transistor devices and stays around a value of 200GHzV. Hence, only transmitter PAs with relatively moderate output power specifications are suited for CMOS integration. On the other hand F_{MAX} gets better for every new CMOS technology node, therefore VCO's, LNA's and up-conversion mixers can be made at higher operating frequencies or with lesser power consumption. Without doubt the receiver section must therefore be implemented in a RF CMOS process, since RF CMOS offers the lowest cost price.

The above elaboration reveals that fully integrated solutions are not always desirable or possible. For the 802.11a/g WLAN specification, which also uses an OFDM multi-carrier scheme, the PAs 1dB compression point is fixed between 5 and 10dBm. Hence, an external PA is needed to boost the output power to the maximum allowed equivalent isotropic radiation power (EIRP) of 20dBm. Nevertheless, WLAN applications targeting a transmission distance up to 10m can directly connect an antenna to their CMOS transceiver chipset. Thus the output power of this short-range WLAN solution comes close to the one defined for UWB. Apparently it seems that WLAN and UWB are competing technologies. However, UWB has a much lower current consumption per transmitted bit and should therefore be used in battery driven applications. In table 1 a comparison is made between some existing wireless standards. UWB clearly demonstrates itself as the application with the highest data rate, but also as the one who is the most power efficient. The table also shows the Bluetooth standard, which is clearly not intended for broadband communication. The reason why it is included in the table, is to show that the PAE of a PA tells only half of the story. Bluetooth uses a constant GFSK constant envelope modulation scheme, as such are very efficient switched or saturated power amplifiers applicable in the TX path. Hence you would expect that this modulation scheme

	Bluetooth [3]	WLAN 802.11b [4]	WLAN 802.11a/g [4]	WLAN 802.11n [5]	UWB [6]
Data rate	2 Mbps	11 Mbps	54 Mbps	270Mbps	480Mbps
TX current	35mA	200mA	107mA	260mA	289mA*
RX current	30mA	65mA	75mA	251mA	258mA*
TX μ A/bit	0.17	0.18	0.019	0.0096	0.0060
RX μ A/bit	0.15	0.059	0.0138	0.0093	0.0054

Table 1: TX and RX Current per Bit

would outperform the others, but table 1 reveals that this is certainly not the case. The conclusion is that PAE of a power amplifier is a good comparison for PAs targeting the same application, but significantly loses its importance when wireless standards are considered. It is clear that UWB, which has only one output power class, is really optimized for short haul communication. Standards as WLAN respectively Bluetooth have more output power classes and therefore they are able to transmit outdoors. The regulation restricts therefore the allowed frequency bandwidth and that's why they have a disadvantage in power consumption. The transceiver chips of table 1 show that the transceivers TX and RX chain consume a similar power. Hence, the PA will not put a special constraint on the used package concerning heat transfer. That's way the main selection criterion for the package is the number of pins for accommodating the digital I/O-puts. Low cost very thin profile quad flatpack no-lead (VQFN) and ball grid array (BGA) packages are for this reason very popular.

The PAs used in the UWB and the WLAN transmission chains need to have a certain OIP3 for accommodating the OFDM signals. Papers [7] and [8] shown that the linearity can be written in function of the transistors F_T as

$$OIP3_{F_T=0} \approx \sqrt{\frac{8 \cdot F_T}{F_T''}}. \quad (11)$$

Above formula indicates that the maximum OIP3 of a CMOS transistor is reached for the biasing point that provides the maximum F_T frequency. Hence, maximizing linearity is equal to minimize the number of cascaded CMOS transistor stages. As a consequence the area occupied by the PA building block is as small as possible. For further technology nodes a strong bending in the F_T and the F_{MAX} curves can be expected. This effect will cause that the linearity degrades for diminishing transistor size. However the gain surplus, offered by the scaling, can be used to correct the linearity by applying local resistive feedback. Furthermore less PA stages are required and this has also a favorable influence on the linearity. Scaling the CMOS transistor size will thus not only reduce area and power consumption in the digital part of the chip.

* Plus 100mA assumed for the base-band

3. Stability of PAs

Guaranteeing that the PA will not oscillate is an important task in linear PA design. This in contrast to PAs using a constant envelope scheme, which even use mode locking or injection locking schemes to improve the efficiency of the PA [9]. The main mechanism responsible for oscillations is the power supply. Hence, the power supply should be an integral part of the design. Two possible approaches can be followed. The first approach is to use a PA design where the supply routing is incorporated into the RF design strategy. The traveling wave-guide amplifiers (TWAs), often used in very high frequency designs, are here the most well known implementation form. The second approach is to simulate the normal cascaded PA stages together with a model of the power grid.

3.1. TWA amplifiers

In TWA the parasitic capacitances of the transistor are absorbed into the transmission lines. A schematic of a TWA structure is given in figure 5. From this figure it can be seen that there are two transmission lines in the design. The first transmission line connects the input of the PA to the gates of the transistors, where the second transmission line connects the drains of the transistors to the

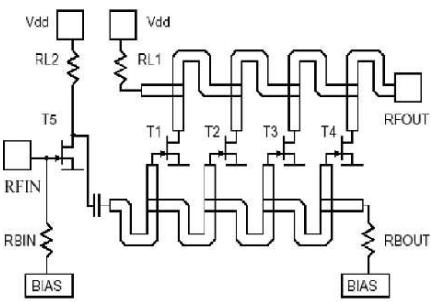


Figure 5: schematic of a TWA

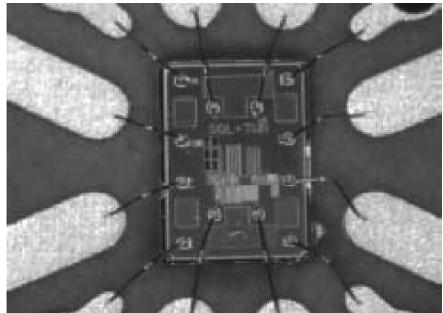


Figure 6: photograph of a UWB TWA [10]

RF output. In general one side of the transmission line is terminated to the characteristic line impedance Z_0 . For the transmission line going to the output of the amplifier the termination resistance is connected to the power supply. Therefore the power supply becomes part of the design strategy. The power distribution in a TWA is thus automatically designed, shaped and layouted. Instabilities will be detected automatically while performing the required design simulations. Accurate modeling the power supply at very high frequencies is difficult and therefore the TWA amplifiers are loved in the RF community. Picture 6 shows a TWA amplifier implementation for UWB [10]. In this design special measures are being taken to overcome some drawbacks associated with

the TWA structure. One of the stringent design limitations is the close relationship between gain and bandwidth. For achieving a certain bandwidth the transistor size should be limited. Limiting the transistor size restricts the gm – the gain – that a device can deliver. The gain in a TWA structure is additive and not multiplicative. A high gain for the total PA structure is therefore not to be expected. In the implementation of [10] the gain issue is tackled by using strip-lines in combination of a non-uniform transmission line approach. Hence, a gain of 17dB and a bandwidth of 8GHz could be achieved. The bandwidth was highly influenced by the discontinuity, which arise of going off-chip. When this discontinuity can be avoided then very high 3dB corner frequencies can be achieved as papers [11], [12] and [13] demonstrate. For example the probe measurements on the CMOS 90nm TWA structures of [13] demonstrate that a pass-band frequency of 80GHz can be achieved in CMOS. However, the power gain of 7,4dB does not fit to existing TX chain specifications. Previous discussion indicates that the used package will have a big impact on the overall performance of the integrated PA. The area consumption of a UWB TWA is around 1mm², when micro-strip structures or CPW are being used. It is quite difficult to utilize the benefits coming from newer CMOS processes to scale down the die size of a TWA structure. The consequence is that TWA amplifiers are being used in the phase where the transistor process allows new communication products to emerge. Recently many UWB CMOS TWAs are being presented at conferences. In the future these designs, inherited from the RF community, will most probably be replaced by analog design approaches.

3.2. Stability of Cascaded PA Stages

3.2.1. Unintended Feedback Loop on Power Grid

Analog design normally cascade amplifier stages and use feedback to improve linearity and stabilize gain. The gain of current deep submicron CMOS devices is not sufficient to apply local feedback in a really effective way. Hence, most cascaded amplifiers are using an open loop structure, where inductors are used to tune every gain stage. The area consumption for todays implemented cascaded UWB power amplifiers is therefore in the same range as for UWB TWAs. Unfortunately the common power grid of the amplifier stages can close unwanted feedback loops and render the circuit unstable. Figure 7 illustrates a possible feedback loop closed by the VDD grid for a two-stage differential PA. A supply ripple at VDD2 is distributed via the common power grid lines to the first stage. Although damped by the power supply rejection ratio (PSRR) of the first stage this signal appears at the signal lines between the stages that close this loop. Hence, the differential power amplifier can exhibit a common mode oscillation. The impedance of a low dropout regulator (LDO) or a bond wire is much higher at high frequencies compared to the illustrated signal path,

therefore the feedback loop only consists of on-chip elements. More stages even increase the risk of oscillations, because additional loops exist and every amplifier stage inside the loop amplifies the unwanted ripple. Moreover the simple circuit topologies used in low voltage RF PAs do not offer good power supply rejection ratio (PSRR) especially at higher frequencies. The circuit analyzed later on in this paper is a four stage PA with CMOS differential amplifiers similar to figure 7. The circuit, realized in a deep submicron CMOS process with a bandwidth from 3 to 5 GHz, has a high risk to suffer from unwanted oscillations.

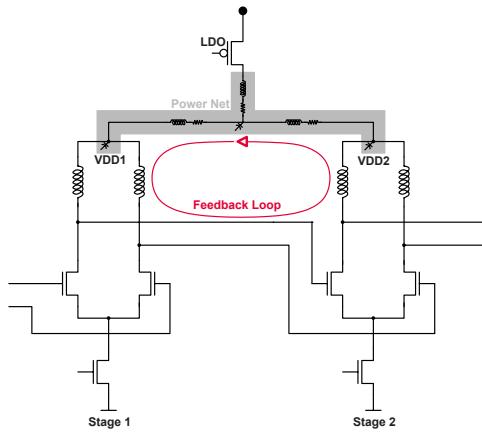


Figure 7: Illustration of feedback via power net

3.2.2. Supply Current in Differential Stages

In addition to doubled signal swing and lower distortion, differential circuits have the advantage unlike single ended designs, that the total current consumption of an amplifier stage is almost DC. (ig in figure 7) Due to the nonlinear behavior of differential amplifiers especially at large signal swings, a supply current ripple at twice the signal frequency cannot be avoided. The current consumption of a differential amplifier using NMOS transistors in strong inversion is

$$i_g = i_+ + i_- = \kappa_n \left[\frac{(V_B + v_{sig} - V_t)^2 + (V_B - v_{sig} - V_t)^2}{2} \right] \quad (12)$$

$$i_g = \frac{\kappa_n}{2} \cdot \left[(V_B - V_t)^2 + V_p^2 \cdot \frac{1}{2} \cdot (1 + \cos(2\omega t)) \right] \quad (13)$$

V_B is the bias voltage and V_t the threshold voltage of the NMOS. In most cases the input signal amplitude V_p is at least an order of magnitude smaller than V_B -

V_t so according to equation 13 only a small fraction of the supply current is at a high frequency. This simplifies the specifications of the power supply, since the inductance of metal lines and bond-wires does not degrade the performance of the circuit.

3.2.3. Power Distribution Network

Besides the power supply rejection ratio and the common mode gain of the amplifier stages, the feedback loop mainly depends on the impedance of the power grid hence a power grid model is mandatory for the stability analysis of the PA. The high operating frequency of the circuit requires the inclusion of inductances in the model. The dashed graph in figure 10 shows the result of a loop gain simulation using a simple RC network for the power grid. The simulation results above 1 GHz are far too optimistic therefore the inductive behavior of the network must not be discarded. Typical RLC models from lumped-model parasitic extraction or from 3D field-solver algorithms are very accurate, but too large for fast simulations. Model order reduction makes these models usable for simulations, enabling accurate post layout stability analysis. An overview of the inductive behavior of interconnects may be found in [17], [18]. However such models are not scalable easily and layout data for extraction is not available during the design phase. A pre-layout power grid model was developed, since it is reasonable to incorporate estimated power grid parasitics early during the design to avoid tedious layout changes later.

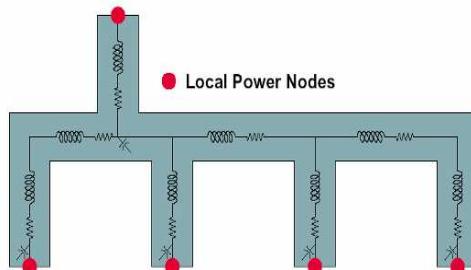


Figure 8: A sketch of a simple power network connecting 4 subcircuit blocks. The corresponding RLC network is also illustrated

Figure 8 shows a typical bus structure and the corresponding RLC network, whose values are defined by a set of interconnect wire lengths and widths. Initial wire dimensions and the connectivity are stored in a configuration file where bus or star-connections may be combined hierarchically to define more complex power grids. To improve the user-friendliness, netlist based tools were developed to add the power grid models to the original circuit and to automatically reconnect all affected power nets. The advantage offered by such tools is that no modifications of the original circuit schematics are required. All

required actions to use the power grid model, except defining the physical dimensions of the metal lines, are automated. To keep the resulting RLC network compact several simplifications were carried out. Blocking capacitors dominate the capacitance of the grid so it is possible to neglect the metal-substrate and metal-metal capacitance of the wires. Inductive coupling is also neglected to keep the amount of circuit elements low. The inductance of each wire segment is calculated with (14) published by Grover [14].

$$L = 2 \cdot 10^{-7} \cdot l \cdot \left[\ln \frac{2l}{w+t} + \frac{1}{2} - 0.2235 \cdot \frac{w+t}{l} \right] \quad (14)$$

l , w and t are the physical dimensions of a wire segment and a 3rd order Taylor series (15,16) is used for approximating the logarithm.

$$\ln\left(\frac{2l}{w+t}\right) = 2 \cdot y + \frac{2}{3} \cdot y^3 + \frac{2}{5} \cdot y^5 \quad (15)$$

$$y = \frac{2l - w - t}{2l + w + t} \quad (16)$$

This formula is used because it is possible to calculate the inductances directly in Spice netlists, which enables a convenient way to sweep wire dimensions. Several commonly used inductance formulas are compared in [16] showing sufficient accuracy of the Grover formula for typical dimensions of on-chip metal lines. Note that these pre-layout estimations are used to compare and scale different power grid topologies and to identify sensitive blocks and nets. The model is not designed to replace a post layout extraction and simulation.

3.2.4. Loop Gain Analysis

Since the power grid closes several feedback loops where each may have more than one return path, a standard open loop AC simulation cannot identify such an unstable circuit easily. Transient simulations show these oscillations but it is time consuming to check stability for a wide frequency range. The most effective way to check the stability of the PA is a small signal loop gain analysis. The circuit is stable if the loop gain amplitude of all loops in the PA is smaller than 1, regardless of the phase. We used a method that does not require opening the feedback loop, only the point where the loop gain is calculated has to be defined. In normal small signal analysis power and ground networks are assumed ideal thus not available as signals in the simulation. When the power grid RLC models are added to the simulation, local VDD and VSS connections become “normal” Spice nodes enabling a meaningful loop gain analysis. Injection of voltage or current test signals to a closed feedback loop is a

common method to measure the loop gain described in detail in [15]. We utilized the combined current and voltage injection. Compared to voltage- or current-only injection, this method is not dependent on the impedance at the injection node. Just the direction of the feedback loop has to be defined. In two consecutive runs, the simulator injects an AC test voltage and a test current to

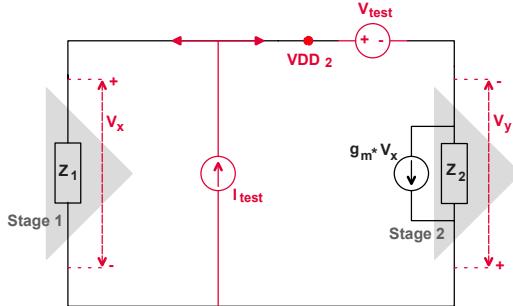


Figure 9: Norton equivalent of feedback loop plus current and voltage injection to measure the loop gain. VDD_2 is shown in Fig. 7

the point where the loop gain is measured. Figure 9 shows in bold the test sources and the Norton equivalent of a feedback loop according to figure 1.

The loop gain T of this circuit is

$$T = g_m \cdot \frac{Z_1 \cdot Z_2}{Z_1 + Z_2}. \quad (17)$$

It is not possible to determine the values of Z_1 , Z_2 and g_m directly, therefore T is calculated differently. A test voltage V_{test} is applied to the circuit and the ratio of the voltages V_x and V_y yields to T_v , the voltage gain of the loop

$$T_v = \frac{V_y}{V_x} = g_m \cdot Z_2 + \frac{Z_2}{Z_1}. \quad (18)$$

In an analogous manner the test current yields to T_i , the current gain of the loop

$$T_i = \frac{i_y}{i_x} = g_m \cdot Z_1 + \frac{Z_1}{Z_2}. \quad (19)$$

The following term allows getting a formula for the true loop gain T

$$\frac{1}{T+1} = \frac{1}{T_i+1} + \frac{1}{T_v+1} \Rightarrow T = \frac{T_v \cdot T_i - 1}{T_v + T_i + 2}. \quad (20)$$

Figure 10 shows the loop gain magnitude of the PA measured at the local power node of the 4th stage. The last stage is the most critical one since the loop gain includes the common mode amplification of all previous stages. Due to this amplification, the loop gain has an in-band peak in the frequency band of the amplifier.

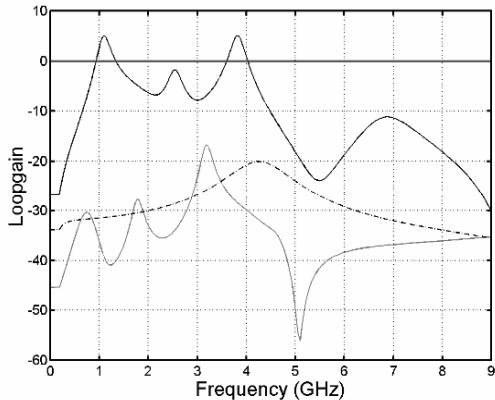


Figure 10: Loop gain magnitude for different power grid models.
black: unstable RLC model; grey stable RLC model, dashed: RC only model

3.2.5. Simulation Results

The main purpose of the simulations presented in this chapter was to identify the critical parts of the power distribution network. Several different power grid configurations were analyzed and power grid dimensions sweeps were made. It was possible to run a very large batch of simulations within a short time, since the AC loop gain analysis together with our pre-layout model is very fast. Statistics of the maximum loop gain were made to identify the critical nets. Figure 11 shows the proportion of simulation results with loop gain below 0dB, referring to the complete batch of simulations, while ordered by grid parameters. Using this data it was possible to find a floorplan that ensures stable circuit operation without sacrificing the amplifier performance.

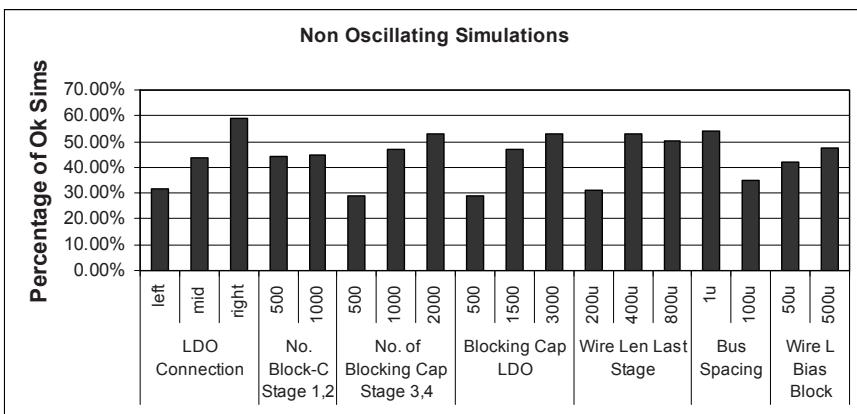


Figure 11: Percentage of simulations with loop gain below 0 dB

A test chip layout that is not optimized for stability was used as starting point for our analysis. As can be seen in figure 11, that one of the most critical parameters for stability are the length of power wire to the 4th stage and the size of the blocking capacitors. Long separate power wires to all amplifier stages and a star connection improve the stability of the PA compared to a more compact power grid (Fig. 12). The LDO has only a small effect on the stability since the regulation loop does not work at RF frequencies. As illustrated in figure 12, a large amount of blocking capacitors is added at the point where all power wires are connected together to filter RF signals at the LDO. Large blocking capacitors at the amplifier stages improve the stability as long as no LC resonant circuits in the frequency band of the amplifier are created. The overall circuit dimensions remain unchanged, only the separated supply lines require slightly more space.

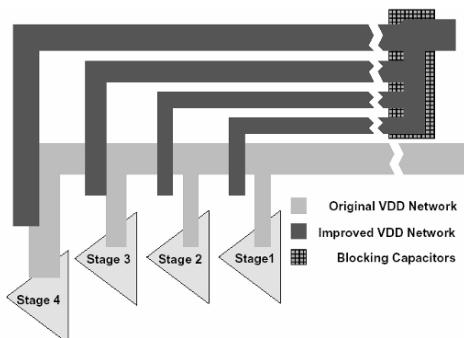
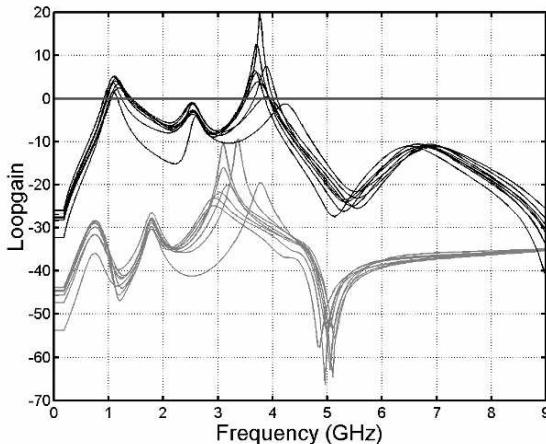


Figure 12: A sketch of original and optimized power grids

In figure 13 loop gain process and temperature corner simulations of the original and the optimized power grid are shown. While the original power grid resulted in an unstable circuit the improved grid allows stable operation. Especially in the critical in-band frequency range from 3 to 5 GHz the improved network shows a strongly reduced loop gain. Only the most critical loop of the last amplifier stage is presented in this paper, but to ensure a stable circuit all possible feedback loops have to be analyzed in an analogous way. Estimated pre-layout lumped-RLC power grid models are sufficiently accurate to identify the critical wire dimensions, and enabled us to analyze and compare different power grid configurations, without having to draw layouts. Differential circuits are essential for this kind of power distribution network, since longer supply wires with higher inductance can be used without affecting the circuit performance. As indicated in this paper future power amplifier will work at higher operating frequencies while using the lower supply voltages of future deep submicron technologies. These circuits will be even more sensitive to this kind of oscillations, so power grid concepts and stability analysis will become a mandatory part during circuit design.



*Figure 13: Process and temperature corner simulation of the loop gain.
Black: original power grid; Grey: improved power grid*

3.3. A WiMedia/MBOA-compliant BG1 CMOS RF Transceiver for UWB

A differential output for the PA is not always desired. For single-ended implementations it is recommendable to use separated power supplies, since achieving a stable loop gain is difficult to accomplish. The nicest implementation way is to connect a low drop out regulator with an internal capacitance at every PA stage. Figure 14 shows the photograph of the chip, which is packaged in a low-cost VQFN plastic package with 48 pins. It is fabricated on Infineon's 0.13um standard digital CMOS technology with 1-poly 6-layer copper metal stack and MiM-Capacitors used as the only RF add-on feature. On TX side the baseband I/Q analog input signal is converted to a current signal by a highly linear voltage-to-current converter and fed into Gilbert-type folded up-converting mixers. To reduce the LO leakage caused by DC-offset within the mixer stage, a compensation DAC is added and controlled by a serial interface bus. The differential output signal of the mixer is converted to single ended, followed by a programmable gain stage and an integrated three-stage power amplifier. A similar structure is used for the three stages. In figure 15 we can clearly recognize that each PA stage is foreseen of a DC regulator, which ensures the stability of the PA. This regulator is constructed by means of a pMOS, which is steered by an OPAMP. The OPAMP senses the voltage at the drain of the nMOS and ensures therefore that the nMOS amplifier transistor is biased for optimum gain and linearity. Using the correct metal widths for the used inductors solves electro-migration issues. The high DC current, of the class A power amplifier, leads in most cases to an inductor quality which is too high for realizing a 1.5GHz bandwidth.

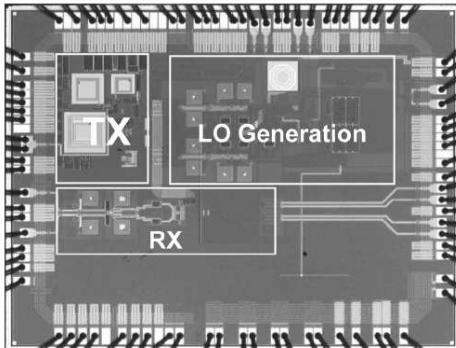


Figure 14: A UWB transceiver [6]

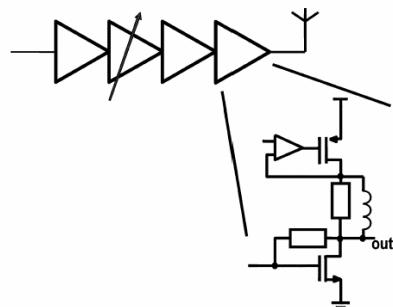


Figure 15: Schematic of a PA stage

Placing a poly resistance in parallel with the inductor, which lowers the quality of the resonance tank, solves the problem. Maximizing the gain in the stage requires that the parallel resistance is as high as possible. To guarantee still a low Q-factor we need to use very high inductance values, as can be seen from the equation below.

$$Q = \frac{R_p}{L \cdot \omega} \quad (21)$$

The coils in the PA are realized as stacked inductors to save area. Stacking two similar coils above each other increases the inductance value by a factor of four compared to a single inductor coil [19]. The parasitic capacitance of the stacked inductor can be easily incorporated into the design. Gain-switching is implemented by a capacitive divider with switchable divider ratio, yielding a variable gain range of 30dB with a resolution of 1 dB for high gain settings. Taking into account some back-off due to external losses, antenna and impedance mismatch enables cost-efficient control of the actual output power to fulfill TX emission mask requirements without need for additional external components. The transmitter is tested with an OFDM WiMedia/MBOA compliant signal [20]. On the left side of figure 16 the constellation diagram at the power amplifier output for band 1 is shown at an output power of -7dBm. The corresponding measured EVM is -28dB. For the band 2 and band 3 the measured EVM is -27.5 and -27dB, respectively. On the right side of figure 16 the EVM in band 1 is drawn while varying the output power. For large output power levels when the PA is approaching its compression region the EVM degrades. At a power close to 0dBm, the transmitter still meets the required EVM of -20dB [20].

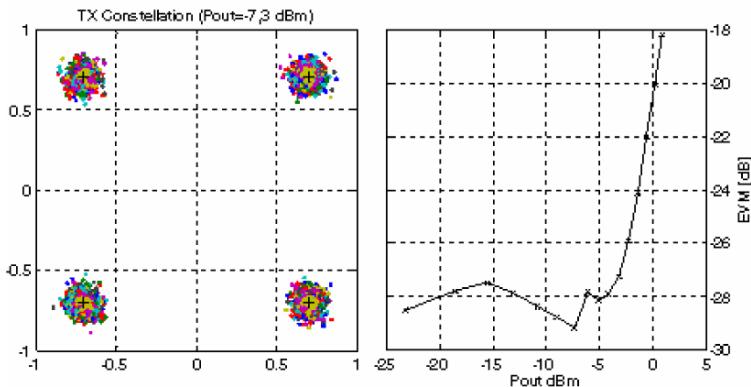


Figure 16: Measured TX constellation (left) and EVM versus Pout (right)

Figure 17 represents the spectrum at the PA output when operating in time frequency interleaved (TFI) mode [20], hopping between all 3 bands. The resolution bandwidth set at the spectrum analyzer is 1MHz as recommended in [21]. The two graphs show the spectrum with power level set close to -41.3dBm/MHz, and with and without external band-pass filter (TDK DEA453960BT). Using the filter all spurious are well below -30dBc.

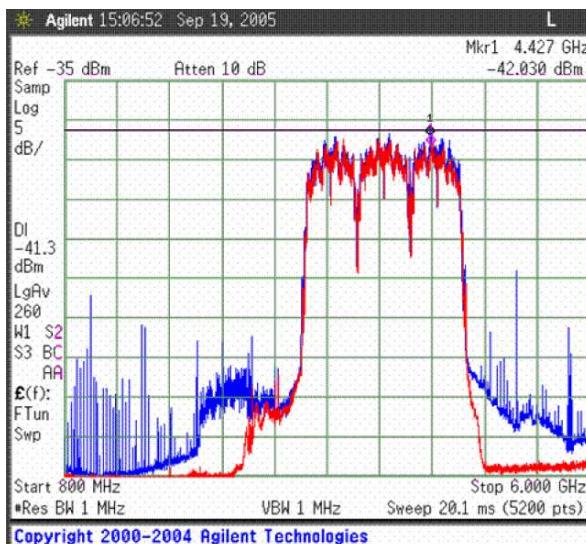


Figure 17: Measured TX spectrum in TFI mode

4. The near Future in Broadband Communication

For boosting the bit rate of current applications as WLAN and UWB two alternatives arise on the horizon. The first possibility is to use multiple input multiple output (MIMO) systems. The second option is to move to the unlicensed frequency band around 60GHz. A MIMO system uses multi-path propagation –fading– to increase the capacity of a transmit channel. The increase in channel capacity is proportional to the used antennas and the number of independent transceivers. A result of a MIMO system is that the data rate improves, but not the current required per transmitted or received bit. For a good operation the required antennas should be spatial separated by 0.4 times the wavelength [22]. This distance is important for guaranteeing that the transceivers are uncorrelated. MIMO systems may not be mixed-up with antenna arrays (AAS or beam forming systems) or selection diversity combining (SDC). The mentioned systems like AAS and SDC increase link budget (thus data rate and range), but this without using unique wireless signals in the RX and TX paths. In SDC and ASS systems the digital signal processor (DSP) is connected to only one AD/DA converter, where for MIMO several ADs/DAs are used. Notice, that for existing standards the modulation scheme is fixed and for these applications AAS and SDC systems will only extend transmission range and NOT throughput. This is an important distinction compared to MIMO systems. Existing transceivers can be combined to a MIMO system with minimal overhead and therefore the approach will fit into a no more Moore scenario. The approach of using the unlicensed band around 60GHz follows of course the opposite direction and depends therefore strongly on further improvements in transistor scaling. The appealing features of the 60GHz band are the possible high bandwidth and the short size required for realizing quarter wavelength structures. The last point will cause that RF design techniques will become more and more prevalent. Going of chip at such high frequencies is difficult and therefore on-chip antennas are potential candidates to rectify the problem. A dipole antenna at 60GHz integrated in a standard CMOS process is 1mm long and has a gain of -6dB [23]. Observe, that the applications at 60GHz are six times higher in frequency than the current developed UWB transceivers. Every CMOS technology node improves transistor speed by a factor of 1.4. Hence, it will take four CMOS generations (a 32nm CMOS technology) before such transceivers could be implemented using standard design techniques. It is clear that without new design and transceiver architectures the arrival of such systems will take still a long time. Normally every 2 years a new technology node is introduced and therefore a period of 8 years is a realistic assumption. It seems that the industry will therefore go on the short term to MIMO UWB systems to implement Gigabit communication systems. On the long term these systems should be replaced by applications at

60GHz. However, this is only possible when the current consumption per transmitted bit is sufficient lower than for a MIMO UWB system.

5. Conclusion

We have seen that the current CMOS technologies are able to realize a UWB transceiver system. The output power required for these systems can easily be achieved without external PAs, as the implementation of [6] proves. UWB systems have the lowest current per bit ratio and should therefore be used in portable applications to extend the battery operation time. The future technology roadmap will allow that the PA can be implemented by analog amplifier structures. For assuring the stability of the PA, it is required to incorporate the power grid into the simulations. A method how this can be achieved is presented in the paper. The most probably scenario for achieving Gigabit communication on the short term is to realize a MIMO UWB transceiver chipset.

References

- [1] C. Yue and S. Wong, "On chip Spiral Inductors with Patterned Ground Shields for Si-Based RF IC's", IEEE Journal of Solid-State Circuits, vol. 33, no. 5, pp. 743-751, May 1998.
- [2] E.O. Johnson, "Physical Limitations on Frequency and Power Parameters of transistors", RCA review, vol. 25, June 1965.
- [3] P. van Zeijl, et al., "A Bluetooth Radio in 0.18 μ m CMOS", IEEE Journal of Solid-State Circuit, vol. 37, no. 12, December 2002.
- [4] S-C Yen, et al., "A Low-power Full-band 802.11abg CMOS Transceiver with On-chip PA", IEEE RFIC symposium, pp. 103-106, June 2006.
- [5] G. Chien, et al., "A Fully-Integrated Dual-Band MIMO Transceiver IC", IEEE RFIC symposium, pp. 99-102, June 2006.
- [6] C. Sandner, "A WiMedia/MBOA -compliant CMOS RF Transceiver for UWB", IEEE Journal of Solid-State Circuits, vol. 41, no. 12, December 2006.
- [7] M. Vaidyanathan, et al., A Theory of High-Frequency Distortion in Bipolar transistors, IEEE transactions on microwave theory and techniques, vol. 51, no. 2, February 2003.
- [8] L. E. Larson, "Advances in Silicon Semiconductor Device Technology for Radio and Wireless Applications", IEEE Radio and Wireless Conference RAWCON, pp. 353-355, Augustus 2003.
- [9] K. Tsai and P. Gray, "A 1.9GHz CMOS class E power Amplifier for wireless communications", IEEE Journal of Solid-State Circuits, vol. 34, no. 7, pp. 962-970, July 1999.
- [10] C. Grewing, et al., "Fully integrated power amplifier in CMOS technology, optimized for UWB transmitters", IEEE RFIC Symposium, pp. 87-90, June 2004.

- [11] M. D. Tsai, "A Miniature 25Ghz 9dB CMOS Cascaded Single-Stage Distributed Amplifier", IEEE Microwave and Wireless Components Letters, vol. 14, no. 12, December 2004.
- [12] L-H Lu, T-Y Chen and Y-J lin, "A 32GHz Non-Uniform Distributed Amplifier in 0.18 μ m CMOS", IEEE Microwave and Wireless Components Letters, vol. 15, no. 11, November 2005.
- [13] R.-c Liu, et al., "An 80 GHz travelling-wave amplifier in a90nm CMOS technology", ISCC conference, pp. 145-155, 2005.
- [14] F. W. Grover, "Inductance Calculations: Working Formulas and Tables," Dover Publications, New York, 1962.
- [15] R. D. Middlebrook "Measurement of loop gain in feedback systems," IEEE International Journal of Electronics, vol. 68, no. 4, pp. 485-512, 1975.
- [16] K. Hyungsuk C. Chung-Ping Chen, "Be Careful of Self and Mutual Inductance Formulae," Technical Report University of Wisconsin-Madison, 2001.
- [17] M. W. Beattie L. T. Pileggi, "Inductance 101: Modeling and Extraction," DAC Design Automation Conference, pp. 323-328, 2001.
- [18] K. Gala, et al., "Inductance 101: Analysis and Design Issues," DAC Design Automation Conference, pp. 329-334, 2001.
- [19] A. Zolfaghazi, A. Chan, B. Razavi, "Stacked Inductors and Transformers in CMOS Technology", IEEE journal of Solid-State Circuits, vol. 36, no. 4, April 2001.
- [20] Multiband OFDM physical layer specification, Release 1.0, Jan 2005, <http://www.multibandofdm.org>
- [21] Federal Communications Commission, Revision of Part 15 of the commission's rules regarding Ultra-Wideband Transmission Systems, Feb 2002, http://www.fcc.gov/Bureaus/Engineering_Technology/Orders/2002/fcc02048.pdf
- [22] B. Bisla, et al., "RF System and Circuit Challenges for WiMAX", Intel Technology Journal, vol. 8, issue 3, August 20, 2004, ISSN 1535-864X, <http://developere.intel.com/technology/itj/index.htm>
- [23] B. Razavi, "CMOS Transceivers for the 60GHz Band" IEEE RFIC symposium, pp. 231-238, June 2006.

POWER COMBINING TECHNIQUES FOR RF AND MM-WAVE CMOS POWER AMPLIFIERS

Patrick Reynaert

M. Bohsali, D. Chowdhury and A. M. Niknejad

University of California at Berkeley

Department of Electrical Engineering and Computer Science

Berkeley Wireless Research Center

2108 Allston Way, Suite 200

Berkeley, CA, 94704

USA

reynaert@eecs.berkeley.edu

niknejad@eecs.berkeley.edu

Abstract

This paper gives an overview of several design issues for CMOS Power Amplifiers (PA) for wireless and mobile communications. The challenges, faced by the RF PA designer, and the different trade-offs, encountered during the design process, are clearly indicated. The idea of power combining is introduced and it is clarified how this technique can alleviate some of the problems related to the aggressive CMOS scaling. The theory is clarified by several design examples that cover a frequency range from the lower GHz as high as 60 GHz.

1 Challenges and trends in CMOS RF PA design

Generating power at a high frequency has always been a challenge in circuit design, both for the old *cm-scale* vacuum-tubes as well as for today's *nm-scale* transistors. However, it is without doubt that the aggressive scaling of CMOS has seriously impeded the design of CMOS PAs. The low supply voltage of nanometer CMOS technologies directly, and quadratically, affects the output power of a power amplifier. With supply voltages below 1 Volt, achieving sufficient output power at RF frequencies is not an easy task. And while the supply voltage of CMOS has scaled

down, the output power requirements of the newest wireless standards clearly has not.

A second challenge in CMOS PA design is the high peak-to-average power ratio of many wireless standards [1]. The first successful CMOS integrated transceivers [2] were targeted for GSM and Bluetooth, both constant envelope systems that only employ phase or frequency modulation. For these systems, the PA thus only needs to have phase linearity. More recent systems like WLAN, WiMAX, Bluetooth-II, W-CDMA, CDMA2000, etc. all employ a non-constant envelope modulation scheme, resulting in amplitude modulation of the output carrier. This in turn requires an amplifier with both amplitude and phase linearity, which often results in a lower efficiency.

A third important challenge, related to the previous one, is the importance of both the long term and the short term average power efficiency of the power amplifier. Two mechanisms will have an impact on the efficiency numbers: *amplitude modulation* and *power control*. After all, an RF PA for mobile communication is seldom operated at the peak output power. Most of the time, less output power is transmitted to save the battery lifetime. It is therefore of importance that the PA achieves a good efficiency for the average output power, rather than a maximum efficiency at the peak output power.

These three major challenges are clarified in more detail in the subsequent sections.

1.1 Output power and power matching

1.1.1 Impedance matching versus power matching

To achieve sufficient output power in a low-voltage CMOS technology, the 50-Ohm load impedance (which in most cases is the input impedance of the antenna or the antenna filter) needs to be converted or transformed into a lower value, and this is typically done by an impedance matching network. This *impedance transformation to achieve more output power* or in short *power matching* is substantially different from the common *impedance matching*.

Impedance matching would strive to convert the 50-Ohm load impedance into the complex conjugate of the source impedance in order to achieve a maximum power transfer between the transistor and the load. Doing so indeed guarantees a maximum power transfer between source and load, but half the signal power is lost in the source impedance, and it ignores the maximum output power that actually can be achieved for a given DC power consumption and input swing. Indeed, a transistor is not an ideal current or voltage source, but has a limited output voltage and output current swing. Furthermore, in a power amplifier it is of more interest to

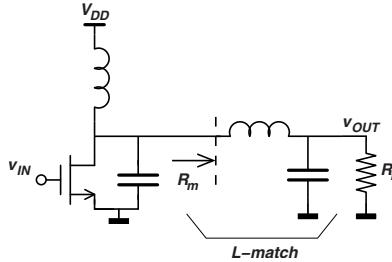


Figure 1: Basic amplifier with an L-match network.

achieve a high output power for a given power consumption, rather than maximizing the power transfer from a given source.

The current and voltage limitations of the transistor therefore give rise to an optimal load in terms of maximum (large signal) output power. This optimal load is substantially different from the optimal load in terms of maximum (small signal) power gain.

1.1.2 L-match network

Several impedance transformation networks exist in literature, but only a few of them are suitable for CMOS integration. The popular L-match, shown in figure 1 together with the skeleton of a basic RF amplifier, is a nice example and only requires two components. The impedance transformation ratio r is defined as

$$r = \frac{R_L}{R_m} \quad (1)$$

Of course, since this network is placed at the output of the PA, it is of utmost importance that it has a low power loss. After all, any power lost in the impedance transformation network will directly affect the power efficiency of the amplifier.

It can be shown [3] that the loss of an L-match network is proportional to the impedance transformation ratio of that network. Furthermore, if the L-match network has power loss, less power than intended flows towards the output and the impedance transformation ratio needs to be made even larger. This will then further decrease the efficiency of the transformation network alone. Therefore, the *power enhancement ratio*, E , defined as the product of the impedance transformation ratio and the efficiency of the transformation network, is a much better number to evaluate the performance of the impedance transformation network [3, 4]. The power enhancement ratio is thus a direct indication of how much more power flows to the

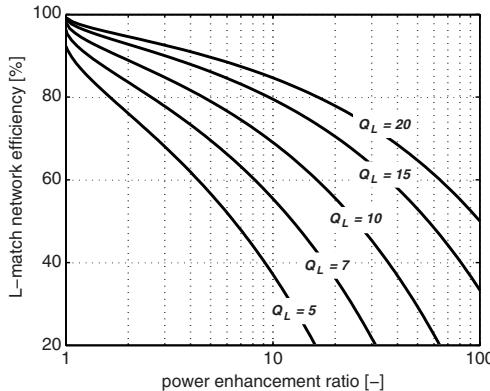


Figure 2: Efficiency versus Power Enhancement Ratio for different inductive quality factors.

load, compared to the case where the 50-Ohm load is directly connected with the output of the PA.

Figure 2 shows the efficiency of an L-match network versus the power enhancement ratio for different values of (unloaded) inductor quality factor Q_L . Let us take a numerical example to evaluate this graph. An amplifier that can deliver a sinewave with an amplitude of 0.5V, would deliver an output power as low as 2.5mW in a 50-Ohm load! If an output power of 50mW is wanted, the required power enhancement ratio needs to be $E = 20$. If one can have inductors with a quality factor of 10, one can see from figure 2 that the L-match network alone will have an efficiency of about 55%. The impedance transformation ratio is equal to 36, much higher than the value of E , and the L-match converts the 50-Ohm load into an impedance as low as 1.4-Ohm. It will be shown later that a power combining architecture is capable of achieving a higher efficiency for a given power enhancement ratio.

1.2 Average efficiency

Recent years have seen a shift from constant envelope modulation schemes (e.g. GMSK) to modulation schemes that use both amplitude and phase modulation, as shown in figure 3. This requires a power amplifier with both amplitude and phase linearity.

On the other hand, and in most cases, peak efficiency is only achieved at maximum output power. When the output power reduces, either because of amplitude modulation or power control, the efficiency will also reduce. Simply put, at lower output power the power amplifier is over-sized, and no longer optimized for maximum efficiency.

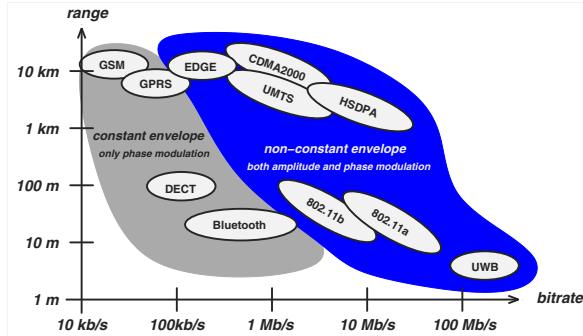


Figure 3: The shift from constant envelope systems to non-constant envelope systems.

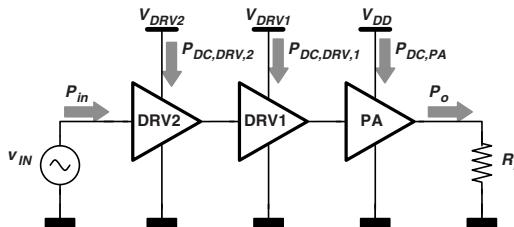


Figure 4: DC and RF power flows in a power amplifier.

The *efficiency versus output power*-curve is thus a crucial graph for modern communication standards. However, one has to be cautious on how *efficiency* and *output power* are defined. Only looking to the output or the last stage of a PA, one can define the drain efficiency as (see figure 4)

$$\eta_d(P_o) = \frac{P_o}{P_{DC,PA}} \quad (2)$$

The output power P_o is defined as the power at the frequency of interest, when a CW signal is applied at the RF input. The power at the harmonics should not be taken into account. To emphasize that the efficiency is dependent on the actual output power, it is written here as $\eta_d(P_o)$.

Linear power amplifiers in general are characterized by a more-or-less linear relationship between the RF input power and the RF output power. Typical examples are Class A, AB and B. When the RF input power is reduced, the drain efficiency of these amplifiers drops. This is most pronounced in a Class A, where the DC power

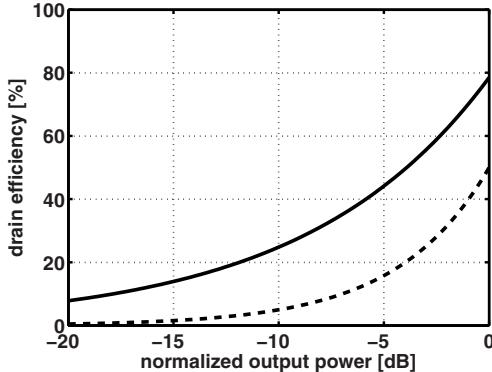


Figure 5: Drain efficiency of the Class B (solid line) and Class A (dashed line) amplifier versus normalized output power.

consumption is independent of the RF input power, resulting in a quadratic relationship between drain-efficiency and RF output amplitude or a linear relationship between drain-efficiency and RF output power. In a Class B, the DC power consumption changes with the RF input power, resulting in a more favorable η -vs- P_o curve. In a class B, the drain-efficiency drops with the square-root of the output power. Both curves are shown in figure 5.

There are two mechanisms that will cause a *movement* or a *trajectory* on the η -vs- P_o curve: *amplitude modulation* and *power control*. Amplitude modulation typically occurs *fast* and requires a high linearity. Due to the amplitude modulation, the RF signal will have an average amplitude and a peak amplitude. The ratio of the two is denoted as the crest-factor or peak-to-average power ratio (PAPR). Power control on the other hand is a relative slow process. The goal of power control is to increase the battery lifetime or, in CDMA systems, to ensure that all signals arrive at the base station with an equal field strength. Power control does not require a continuous amplitude control range but rather uses discrete power steps. There is clearly a big difference between the two mechanisms, and this reflects itself in the different architectures that are used to increase the efficiency.

As the instantaneous output power changes, either by amplitude modulation or by power control, the instantaneous efficiency also changes. One could think of an *average efficiency* defined by the long-term average output power divided by the long-term average DC consumption. The average efficiency is then defined as

$$\langle \eta_d \rangle = \frac{\langle P_o \rangle}{\langle P_{DC,PA} \rangle} \quad (3)$$

This average efficiency can also be obtained by looking at the statistical distribution of the output power, defined as the Probability Density Function (PDF) of the output power, and written as $p(P_o)$ [5]. If one assumes that the amplifier has no memory effect, the above calculation can also be done by using the η -vs- P_o curve, which is actually a steady-state curve. In that case, the average efficiency can be written as:

$$\langle \eta_d \rangle = \frac{\int_{P_{o,min}}^{P_{o,max}} P_o \cdot p(P_o) \cdot dP_o}{\int_{P_{o,min}}^{P_{o,max}} \frac{P_o}{\eta_d(P_o)} \cdot p(P_o) \cdot dP_o} \quad (4)$$

As clarified before, one can define two PDF functions of the output power: one related to the amplitude modulation of the carrier, and one due to the power control.

1.3 Efficiency Enhancement versus Linearization

It was shown in the previous paragraph that the average efficiency is much more important than peak efficiency. After all, amplitude modulation and power control will cause the amplifier to operate at an average output power which can be substantially lower than the peak output power. Therefore, several techniques exist to achieve a high efficiency over a wide power range. These techniques try to achieve a high average efficiency.

Efficiency enhancement strive to *push* the η -vs- P_o curve upwards, at lower power levels. Two major techniques exist in literature: *Doherty* [6, 7] and *Envelope tracking* [8].

Linearization is another approach to achieve a high average efficiency. It starts from a non-linear amplifier with a high peak efficiency. A non-linear amplifier can only transmit a single output power (Class E) or has a poor linearity (Class C). To expand the linear range of such a non-linear amplifier, the supply voltage can be modulated according to the amplitude signal [9]. Figure 6 shows the conceptual difference between the two approaches.

2 Power Combining Techniques

Power combining techniques, in general, combine the output of several power amplifiers into one single output. This approach has an important advantage (1) to achieve sufficient output power, (2) to improve the average efficiency and (3) has potential for amplifier linearization. In what follows, three different power combining architectures are discussed and a CMOS implementation of each architecture is given.

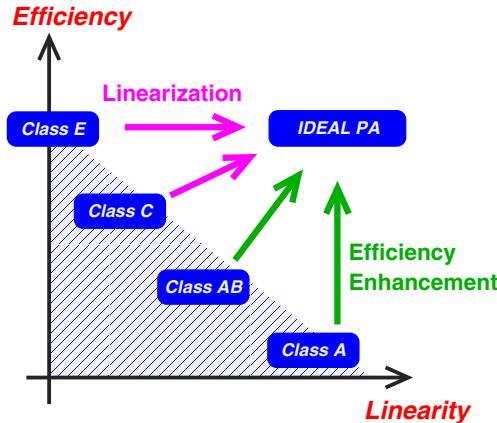


Figure 6: Difference between Efficiency Enhancement and Linearization.

2.1 The multi-section lattice-type LC balun

To combine the outputs of two differential power amplifiers into a single-ended output, and to perform the impedance transformation simultaneously, a lumped element LC balun is often used [10, 11]. This network was originally used in 1932 as an antenna balun to convert a balanced antenna, like a dipole or loop antenna, to an unbalanced transmission line, such as a coaxial cable [12]. In the approach discussed here, the LC balun will be used to combine the outputs of several power amplifiers to one single-ended output.

2.1.1 Basic Equations

Figure 7 depicts the proposed topology with multiple LC baluns placed in parallel [13]. The transformed load impedance, seen by each power amplifier, equals

$$Z_{in,C} = Z_{in,L} = R_m = \frac{R_L}{2} \cdot \frac{1}{N} \cdot \left(\frac{B}{R_L} \right)^2 \quad (5)$$

with N the number of parallel section and B equal to

$$B = \omega L_m = \frac{1}{\omega C_m}. \quad (6)$$

This topology thus consists of $2 \cdot N$ amplifiers, and two differential amplifiers are needed for one section. As with the L-match network, decreasing the ratio (B/R_L) will reduce the impedance seen by each of the power amplifiers. Yet another means

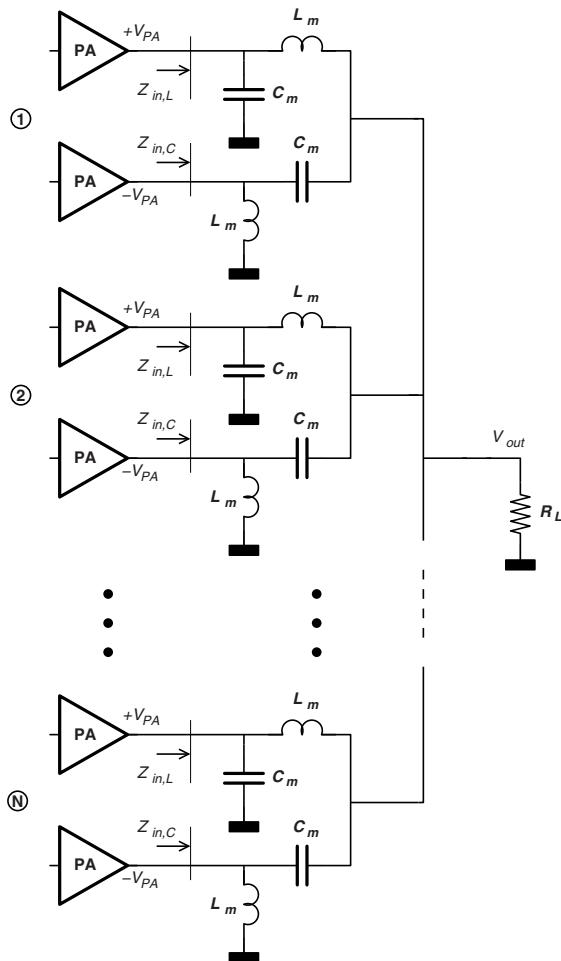


Figure 7: Parallel amplification topology, consisting of N sections.

to decrease the transformed impedance, as clearly visible from equation 5, is by increasing the number of parallel sections N .

The total output power, i.e. the power dissipated in R_L , is equal to

$$P_{out} = 4 \cdot N^2 \cdot \frac{V_{PA}^2}{R_L} \cdot \left(\frac{R_L}{B} \right)^2 \quad (7)$$

in which V_{PA} represents an RMS voltage. The impedance transformation ratio for each individual power amplifier becomes

$$r = \frac{R_L}{R_m} = 2 \cdot N \cdot \left(\frac{R_L}{B} \right)^2 \quad (8)$$

The efficiency of the proposed topology will depend on the impedance transformation ratio of each section. In this regard, one should realize that the ratio R_L/B has a quadratic effect on both the output power and the impedance transformation ratio r . On the other hand, the number of parallel sections N has the same quadratic effect on the output power, but it only has a linear effect on the impedance transformation ratio. As such, increasing the number of sections will quadratically increase the output power, and it will only linearly increase the impedance transformation ratio r of each section. Since the efficiency of each section depends on the impedance transformation ratio, increasing the number of parallel sections will allow to achieve a high output power with a relative low impedance transformation ratio for each section, and thus a high efficiency.

2.1.2 Power control

The topology of figure 7 also allows a discrete form of efficient power control. If the two differential power amplifiers of a specific section are not operational, and the output of both PAs behave as an AC ground, the LC balun of that section becomes a high-impedance LC tank which is connected in parallel with the 50-Ohm load. This idea is depicted in figure 8. The inactive section will not contribute to the total output power, N is reduced to $N - 1$ and because of the high-impedance nature of the parallel LC tank, the corresponding output network will, in the ideal case, not influence the operation of the other amplifiers.

2.1.3 Average efficiency improvement

The power control ability will have a positive influence on the drain efficiency of the switching amplifiers. After all, the drain efficiency of a switching amplifier strongly depends on the ratio of the on-resistance of the switch r_{on} to the load resistance of

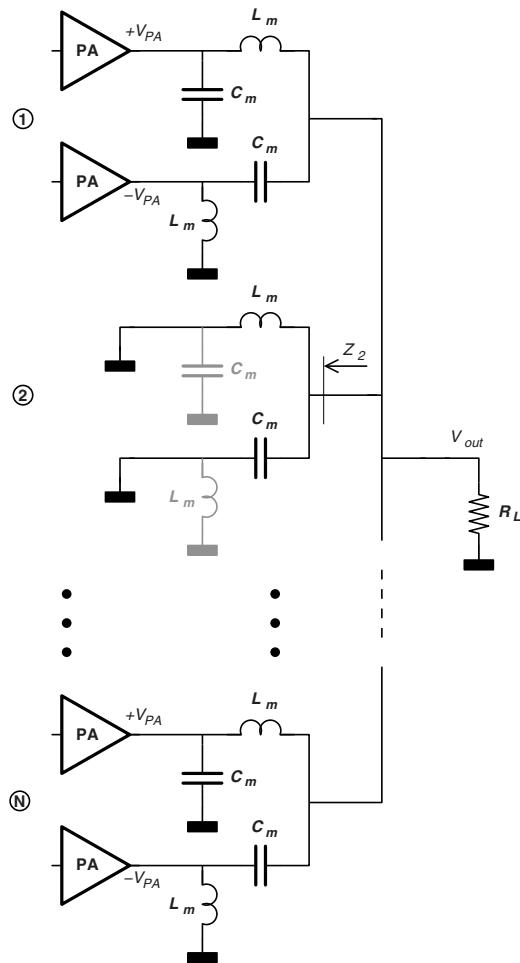


Figure 8: Parallel amplification topology, consisting of N sections, with an inactive section 2.

the power amplifier R_m . An approximated formula is given by [14] and [15].

$$\eta_d \approx \frac{1}{1 + 1.4 \cdot \frac{R_m}{R_m}} \quad (9)$$

By de-activating one or more differential power amplifiers, thus decreasing N in equation 5, the load of the remaining active power amplifiers, R_m , will increase. This in turn increases the drain efficiency of the remaining power amplifiers, according to equation 9. This effect is actually a discrete form of load pull, and is also related to the efficiency improvement that occurs in the Doherty amplifier [6] [7].

Apart from the positive influence on the drain efficiency, the discrete form of power control will have a more pronounced positive effect on the global efficiency, which can be understood as follows.

A power amplifier never stands alone, but one or more driver stages need to be added to increase the power gain of the transmitter. CMOS devices have a relative large gate capacitance and as a result, the power consumption of the driver stages becomes relatively large compared to other RF technologies. The global efficiency of the amplifier takes this power consumption into account and is defined as

$$\eta_g = \frac{P_{out}}{P_{DC,PA} + P_{DC,DRV}} \quad (10)$$

with $P_{DC,PA}$ the DC power consumption of the power amplifier stage and $P_{DC,DRV}$ the DC power consumption of the driver stages. Another commonly used definition is the power added efficiency, defined as

$$PAE = \frac{P_{out} - P_{in}}{P_{DC,PA} + P_{DC,DRV}} \quad (11)$$

where P_{in} is the RF input power. If the total power gain is large, the power added efficiency and global efficiency are almost equal.

The transmitted output power of a switching or saturated amplifier is commonly regulated by changing the supply voltage. Reducing the supply voltage will quadratically reduce the output power while the switching amplifier still maintains its high drain efficiency. On the other hand, the supply voltage of the driver stages can not be reduced as otherwise the power amplifier will no longer operate as a switching amplifier and the drain efficiency would drastically reduce. As such, the power dissipation of the driver stages will remain constant when the supply voltage is reduced and from equation 10, this means that the global efficiency will reduce at lower output power.

Instead of reducing the power supply of the last stage, it is also possible to turn off one or more amplifiers in order to reduce the output power. In that case, the driver stages of the corresponding sections can be shut down as well. The total

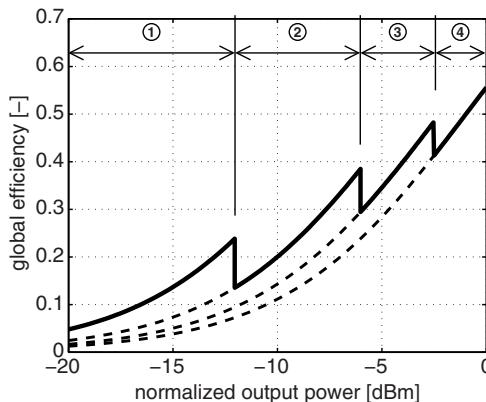


Figure 9: Global efficiency versus output power for 1,2,3 and 4 sections in parallel.

power consumed by the driver stages will then reduce and the global efficiency will increase at a lower output power. Figure 9 shows an example of this effect. In this figure, eight amplifiers, i.e. four section, are placed in parallel. The DC power consumption of the driver stages is assumed to be 20% of the peak power dissipation of each individual amplifier, and the drain efficiency is assumed to be 100%. The peak output power is achieved when the four sections are all active. When the supply voltage of all four amplifiers is reduced, the global efficiency decreases as indicated by the bottom dashed line in figure 9. However, at lower output power levels one or more sections can be turned off and this will result in a higher global efficiency for the same output power. As such, this topology allows to select the optimum number of parallel sections that gives the best global efficiency for the required output power. This optimal curve is shown by the solid black line in figure 9.

Example: A 2.45 GHz Class BE PA in 130nm CMOS

Circuit Implementation

The presented parallel amplification structure was used to implement a Bluetooth PA in a 0.13 μm CMOS technology with an f_T of 70 GHz. The goal was to demonstrate that with native 0.13 μm transistors and the presented power combining network, sufficient output power can be achieved to meet the Class 1 Bluetooth requirements. No thick gate oxide transistors or device stacking techniques were used. The circuit of this amplifier is shown in figure 10.

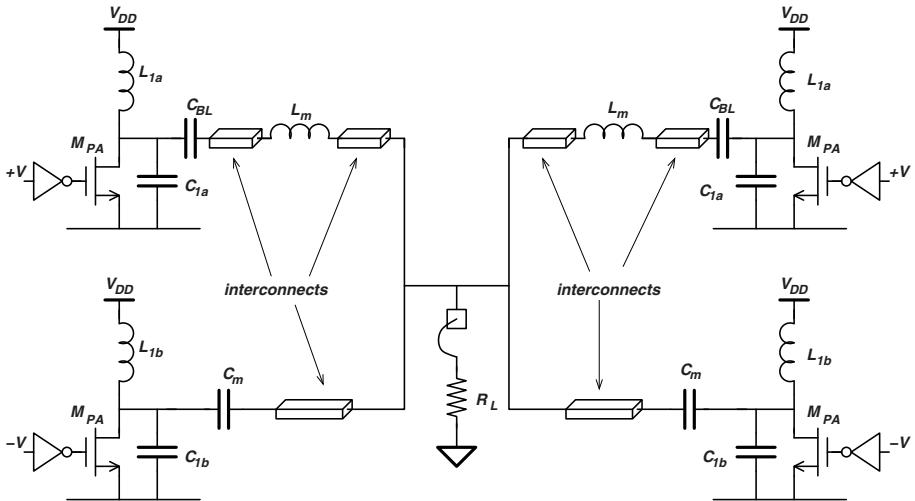


Figure 10: Circuit implementation of the Bluetooth PA.

To achieve the 20dBm output power specification at 2.45GHz, four amplifiers are placed in parallel and their outputs are connected by means of two LC baluns. To provide some margin, each amplifier is designed to deliver 60mW of output power, resulting in a theoretical maximum output power of 240mW or 23.8dBm. For the design of the output stage, a switching topology was chosen that combines the high efficiency of Class E with the high output power capability of Class B. Class E suffers from the high peak drain voltage, especially in deep sub-micron CMOS where the breakdown voltage of the MOS devices is the main limiting factor for PA design. However, compared to Class E, the peak drain voltage of this amplifier is reduced to about 2.5 times V_{DD} which is more favorable for CMOS integration. On the other hand, the conduction angle of this operating class remains 50% to ensure the maximum output power of Class B. Hence, we have denoted this operating regime as Class BE [16]. At zero drain current, the nMOS transistor of this technology can safely withstand a drain voltage of 4V. Assuming a Class BE waveform with a peak drain voltage of 3.5V, the equivalent load impedance required to have an output power of 60 mW is $R_m = 9\Omega$. From the equations derived above, it can be found that $B = 44\Omega$ and thus at 2.45GHz, $L_m = 2.86\text{nH}$ and $C_m = 1.48\text{pF}$.

The RF driver stages are digital inverters, operating at the nominal supply voltage of 1.2V. Digital inverters consume less silicon area, but the drawback is the increased DC power consumption since the gate capacitance of the output transistor is not tuned by an inductor. The input power is only -6 dBm which enables a direct connection of the upconversion mixers to the first RF driver stage. All the inductors and capacitors of figure 10 are integrated on-chip and no external matching or tuning is necessary.

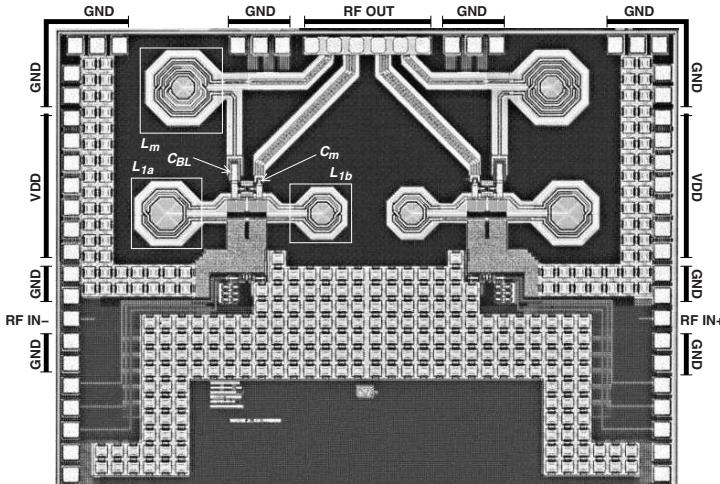


Figure 11: Micrograph of the fully integrated CMOS Bluetooth PA.

Layout

Figure 11 shows the photograph of the fully integrated RF amplifier. The size of the chip is 2.74 mm by 2.00 mm. A reduction of the consumed silicon area could be achieved by merging L1a and L1b together in one single inductor. However, since $L_{1a} > L_{1b}$, this would require an inductor with a non-centered common tap. To connect the four outputs, long interconnections are necessary. Underneath the interconnect lines lies a patterned metal ground plane to accurately model the parasitic capacitance, to avoid capacitive signal injection into the substrate and to short-circuit the substrate losses. Finally, a total value of two times 206 pF is implemented to bypass the supply voltage of the Class BE stage.

Measurements

With all four amplifiers operating in parallel, a maximum output power of 200 mW or 23 dBm can be achieved at a global efficiency of 28%. The corresponding drain efficiency of the amplifier is 34% and the driver stages consume 118 mW. The latter number could be reduced if tuned driver stages were used, but it would drastically increase the chip area. When two amplifiers are on, the peak output power is 60mW or 17.8dBm, the maximum global efficiency is 21% and the corresponding drain efficiency is 27%. The power dissipation in the driver stage is divided by two, resulting in a consumption of 59mW.

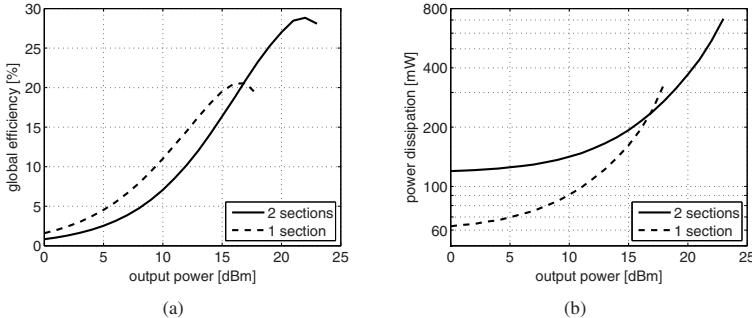


Figure 12: Efficiency improvement: (a) global efficiency versus output power and (b) power dissipation versus output power.

Table 1: Measured constant envelope performance summary.

parameter	measured performance	
	2 sections - 4 PAs	1 section - 2 PAs
maximum output power	200 mW - 23 dBm	60 mW - 18 dBm
maximum drain efficiency	42 %	32 %
maximum global efficiency	29 %	21 %
input power	-6 dBm	-6 dBm
driver stage power consumption	118 mW	59 mW

When only one section is used, the dissipation of the driver stage is divided by two, which results in an increase of the global efficiency at lower power levels. This mechanism is clearly indicated on figure 12(a) in which the output power is changed by reducing the supply voltage of the last stage only. At a power level of 17 dBm and below, it is beneficial to use only one section or two power amplifiers. To further demonstrate the benefit to switch off one section at lower power levels, figure 12(b) shows the measured power dissipation of the entire amplifier, including the driver stages. Table 1 summarizes the power and efficiency measurements. This amplifier also meets the Bluetooth spectral mask specifications, and details can be found in [13].

This work clearly demonstrates how a power combining topology is capable of achieving sufficient output power at low supply voltage, how it can efficiently implement power control and how the efficiency at lower output power level can be improved.

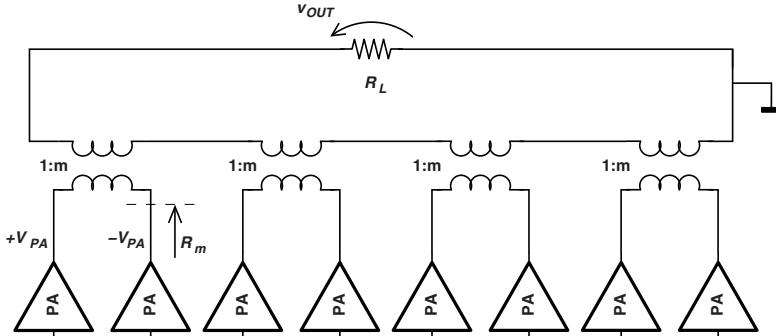


Figure 13: Power amplifier topology using transformers for power combining.

2.2 Transformer-based Power Combining

At GHz-frequencies, it becomes feasible to design and implement low-loss transformers in a mainstream CMOS technology. This allows to design novel transformer-like power amplifier topologies that combine the output of several power amplifiers, resulting in all the benefits of the previously discussed LC-balun power combining technique.

2.2.1 Basic Equations

A straightforward way to implement a power combining architecture by using transformers, is depicted in figure 13. In contrast to the distributed active transformer approach of [3], the topology in figure 13 also allows power control, since each stage works independently and can be turned-off.

As before, the impedance that each individual amplifier sees is determined by two factors, in this case the impedance transformation ratio, i.e. the turn ratio of each transformer, as well as by the number of parallel stages. Let N be the number of transformers and m be the turn ratio. In that case, the impedance that each single-ended PA sees, becomes:

$$R_m = \frac{R_L}{2 \cdot N \cdot m^2} \quad (12)$$

and the total output power equals

$$P_o = \frac{4 \cdot N^2 \cdot m^2 \cdot V_{PA}^2}{R_L} \quad (13)$$

The output power can be increased either by increasing m or N . The turn ratio m can be increased, maybe to a value 2 or 3, but in most cases it is much more

convenient to use a 1:1 transformer, i.e. $m = 1$. Transformers with a 1:1 ratio are much easier to layout and tend to have a lower loss.

2.2.2 Power Control and Power Back-off

Let us examine the output power control capability of the power amplifier topology of figure 13. Assume that four 1:1 transformers are used together with four differential Class B power amplifiers, characterized by a large signal G_M .

At peak output power, each amplifier is on. The single-ended load seen by each amplifier is

$$R_m = \frac{1}{8} \cdot R_L \quad (14)$$

The voltage gain of each amplifier equals

$$G_A = G_M \cdot \frac{1}{8} \cdot R_L \quad (15)$$

and the total voltage gain equals

$$G_T = G_M \cdot R_L \quad (16)$$

The total output power is equal to

$$P_o = 4 \cdot 16 \cdot \frac{V_{PA}^2}{R_L} = P_{o,max} \quad (17)$$

Sections can be turned off by making their G_M equal to zero. However, the secondary inductance of the transformer of the in-active sections is then still placed in series with the load. Therefore, one has to re-tune the primary side of the in-active transformers, in order to reduce the remaining secondary series inductance.

When one section is turned off, and assuming that the secondary inductance can be made small enough, the load seen by each amplifier drops down to

$$R_m = \frac{1}{6} \cdot R_L \quad (18)$$

and the voltage gain of each stage increases to

$$G_A = G_M \cdot \frac{1}{6} \cdot R_L \quad (19)$$

To ensure that the remaining active amplifiers still operate near saturation, the input voltage needs to be reduced by 3/4 or 75%. The output power then becomes

$$P_o = 4 \cdot 9 \cdot \frac{V_{PA}^2}{R_L} = \frac{9}{16} \cdot P_{o,max} \quad (20)$$

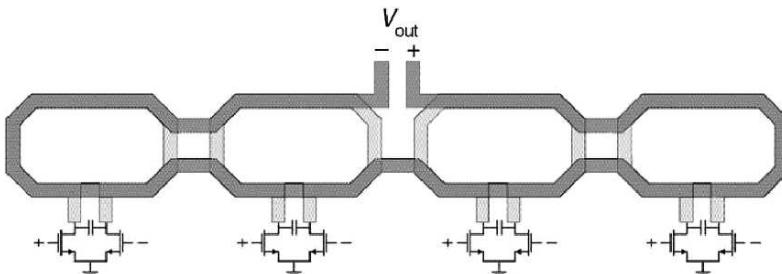


Figure 14: Simplified schematic of the 130nm 2.5GHz power amplifier.

which is 2.5dB below the maximum output power. The output power is thus reduced and each amplifier still operates at maximum drain efficiency. Clearly, the topology allows to improve the efficiency at power back-off.

The same analysis can be repeated when only two and only one amplifier is active. This results in a power back-off of 6dB and 12dB respectively.

Example: a 2.5 GHz Class AB PA in 130nm CMOS

The described topology was implemented in a 130nm CMOS technology with two thick metal layers [17]. Figure 14 shows the simplified topology. A Class AB biasing was chosen to achieve sufficient linearity.

Two thick metal layers were available for the transformer, so a stacked layout was chosen to minimize the losses. The magnetic coupling factor of the transformer is about $k = 0.7$ and the efficiency of the transformer, simulated with ADS Momentum, is approximately 80%.

The amplifier core makes use of a shared-junction thin-oxide cascode structure (not shown in the figure). The cascode structure improves the reliability, and the shared junction area reduces the parasitic capacitance on the cascode node. A tuning capacitor is connected to the primary side of the transformer to re-tune the sections that are turned off. In doing so, the remaining inductance of the secondary side of the transformer, which is in series with the load, can be minimized. The entire amplifier works from a 1.2-V power supply and requires no external tuning elements to achieve optimal performance. Figure 15 show a micrograph of the chip. The total area is 2 mm by 1.2 mm, including the bondpads.

Since this is a one-stage design, the gain of the amplifier is only about 10dB. Figure 16(a) shows the measured output power and efficiency versus input power. The PA transmits up to 24dBm of linear output power with 25% drain efficiency.

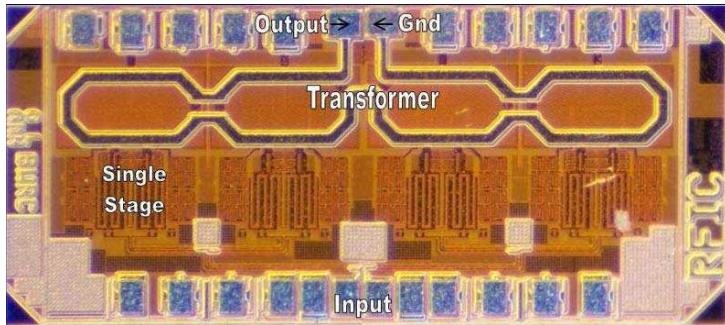


Figure 15: Micrograph of the 5.8 GHz PA.

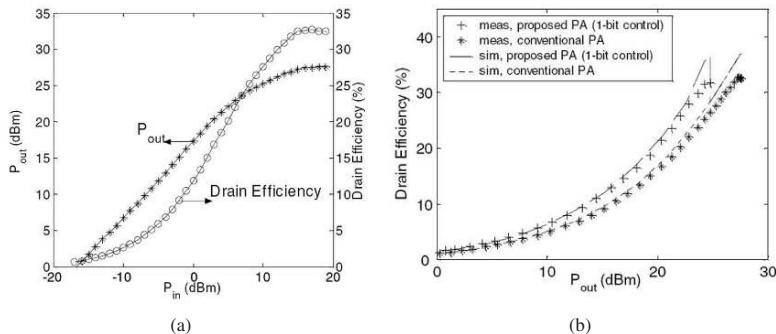


Figure 16: Measured output power and drain efficiency versus input power (a) and efficiency versus output power at full output power and at 2.5dB power back-off (b).

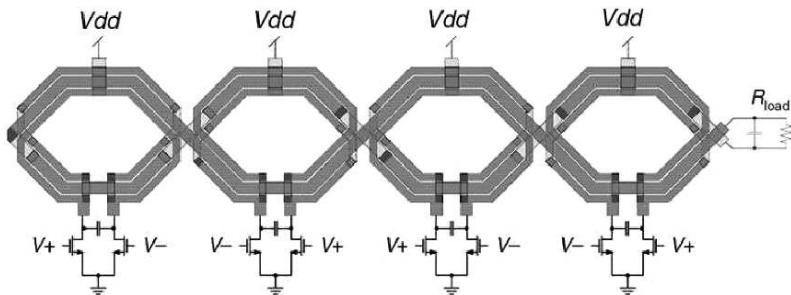


Figure 17: Planar transformer layout using four 1:1 transformers.

When driven into saturation, it delivers a power as high as 27dBm with a drain efficiency of 32%. The efficiency improvement at power back-off is clearly shown in figure 16(b) for a 2.5dB power back-off, and matches very good with the simulation results¹.

Example: a 5.8 GHz planar transformer in 90nm CMOS

When only one thick metal layer is available, a planar transformer layout is mandatory in which both the primary and the secondary turns make use of the same metal layer. Figure 17 shows the layout of power amplifier topology that makes use of four planar 1:1 transformers in CMOS [18]. The primary side of the transformer is split-up in two parallel windings that are twisted around the secondary coil. This will reduce the proximity effect and current crowding, and will thus reduce the series loss resistance. Another drawback of the layout in figure 14 is the partial flux cancellation that occurs since adjacent primary windings have currents in opposite directions. By reversing the current directions of the adjacent transformers in figure 17, this flux cancellation can be minimized. Also note the figure-eight shape of the primary which minimizes the coupling to common-mode magnetic fields.

The simulation of this structure in a 90nm digital CMOS process with only one thick top metal layer gives a power efficiency of 0.752 at 5.8 GHz. The complete structure only occupies 0.65mmx0.15mm.

2.3 Microwave Power Combiners

The recent allocation of 7GHz of spectrum around 60GHz for data communication has triggered the research for mm-wave CMOS design [19, 20]. Transistors in 90nm

¹It was not possible to measure higher back-off levels due to an error in the measurement board.

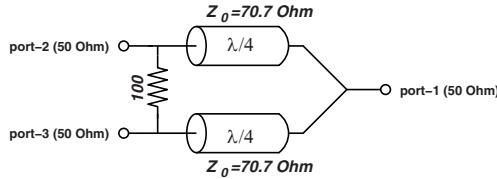


Figure 18: The Wilkinson power combiner.

and 65nm CMOS indeed achieve an f_T and f_{max} higher than 100GHz, meaning that fully integrated mm-wave CMOS transceivers become feasible.

2.3.1 mm-wave CMOS design

The performance of CMOS transistors at mm-wave is strongly dependent on the parasitic resistances. Especially the gate resistance is of primordial concern. To reduce the gate resistance, MOS devices with finger widths of 1 μm or less are required at mm-wave. Power amplifiers, however, typically require ‘large’ transistors, with a total gate width of several millimeters. This means that a huge amount of small fingers need to be placed in parallel. The parasitic resistance and inductance of the interconnections, needed to combine the gate, drain, source and bulk of all these small transistors, will come into play at mm-wave frequencies and will in most cases deteriorate the mm-wave performance.

Therefore, power combining is needed to achieve sufficient output power from small transistors with low output power capability.

2.3.2 A Wilkinson Power Combiner in CMOS

The Wilkinson power splitter, invented around 1960 by Ernest Wilkinson, splits an input signal into two equal phase output signals, or combines two equal-phase signal into one in the opposite direction. The Wilkinson combiner or splitter uses quarter-wave transformers to match the split ports to the common port, as shown in figure 18. For power combining purposes, port 2 and port 3 would be connected to two amplifiers, and port 1 would be the output port. Of course, a lossless reciprocal three-port network cannot have all ports simultaneously matched. Therefore, a resistor is added to the combiner which allows all three ports to be matched, and it also fully isolates port 2 from port 3 at the center frequency. Amazingly, the resistor adds no resistive loss to the power split, so an ideal Wilkinson splitter is 100% power efficient.

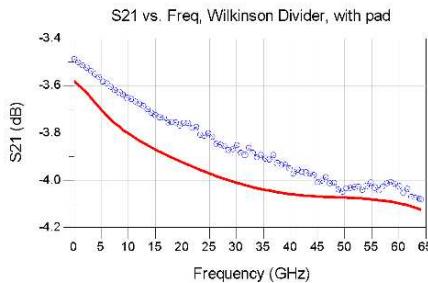


Figure 19: Measurement (circles) and model (solid line) of a 2-to-1 Wilkinson power combiner in 90nm CMOS.

Figure 19 shows the measurement of a 2-to-1 Wilkinson power combiner implemented in a 90nm CMOS technology, using coplanar transmission lines. The insertion loss at 60GHz is about 1-dB and matches well with the Ansoft HFSS simulation.

A 4-to-1 combiner can be build using three 2-to-1 combiners. This would result in an insertion loss of 2.2dB, giving a power efficiency of about 60%. Another approach is to use one single 4-to-1 combiner, which would require transmission lines with a Z_0 of 100Ω , as shown in figure 20. The simulation result of this structure in Ansoft HFSS show an insertion loss of 1.4-dB, resulting in an efficiency of 72%. On the other hand, the 4-to-1 combiner needs transmission lines with a characteristic impedance of 100Ω , which in turn requires a coplanar structure with a large gap spacing of about 20 μm . This poses a lot of problems in nm-CMOS technologies which typically requires *dummy filling* or *tiling* to meet minimum metal density rules. Tiling has little or no effect on inductors since the magnetic field is not disturbed by the small metal tiles. The electric field, obviously present in a coplanar transmission line, are heavily influenced by the presence of the tiles. Slow wave structures [21] are a possible road to solve these issues.

Example: a 60GHz PA in 90nm CMOS

Using the 2-to-1 Wilkinson power combiner/splitter, a 60GHz PA was designed in a 90nm CMOS technology. Figure 21 shows the schematic of the PA, using a 2-to-1 Wilkinson power splitter at the input, and a combiner at the output. The two nMOS transistors in this schematic consist of 80 fingers with a width of 1 μm and minimal gate length. Such a transistor alone achieves a measured MSG of 7.6dB at 60GHz and an f_{max} of 176GHz for a DC bias current of 25mA from a 1-V supply.

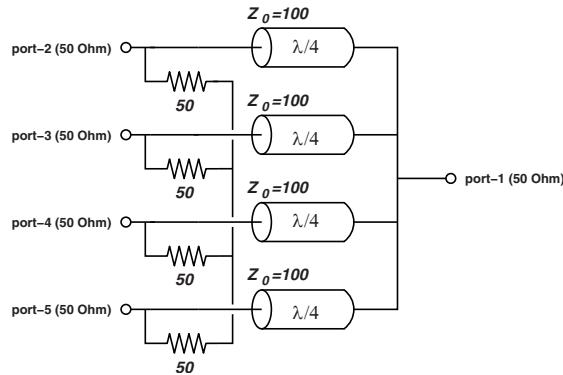


Figure 20: A 4-to-1 Wilkinson power combiner.

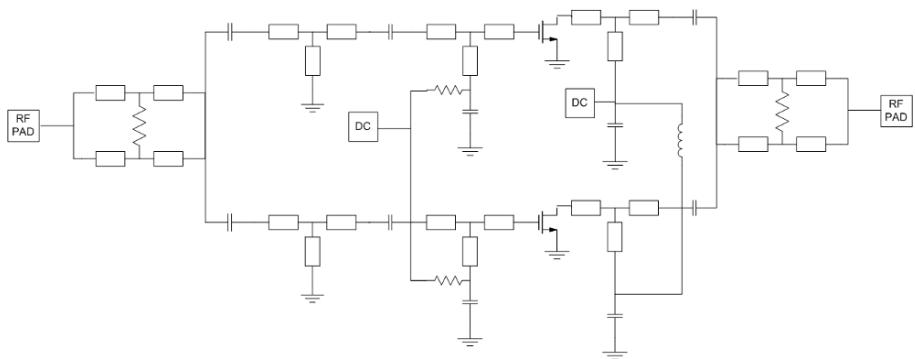


Figure 21: Schematic of the 90nm amplifier with a Wilkinson power splitter and combiner.

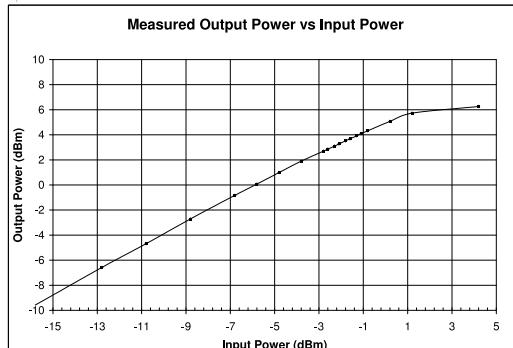


Figure 22: *Measured output power versus input power of the 90nm 60GHz amplifier.*

Figure 22 shows the measured performance of the amplifier. At 60GHz, a saturated output power of +6dBm and a 1-dB compressed output power of +4dBm (2.5mW) was achieved. This results in a power efficiency of 5% at the 1-dB compression point. The power gain of the amplifier is 6.2dB. The simulated 1dB compressed output power of this amplifier was 6.75dBm, which is almost twice the measured value.

Figure 23 shows the measured and simulated S-parameters of the amplifier. Clearly, a mismatch has occurred between the simulation and the measurements, mainly at the input port (S_{11}). Also the voltage gain (S_{21}) is about 3.5-dB higher than simulated. It is believed that this variation is caused by a too aggressive design methodology that placed the design close to the input stability circle. An increase of C_{gs} by only 6fF (10% of its nominal value), caused e.g. by process variations, pushes the design beyond available power gain circle, towards the input stability circle.

3 Conclusions

CMOS RF Power Amplifier design is heavily impeded by the low supply voltage of nm CMOS technologies. Furthermore, the high—and increasing—dynamic range requirement, originating from both amplitude modulation and power control, increases the need to achieve good long-term average efficiency. CMOS can still be the technology of choice for RF PA design if circuit innovation and a flexible PA topology are combined with digital PA control and predistortion. In fact, since digital CMOS transistors are virtually for free, such a topology is likely to become the solution of choice for CMOS power amplifiers in the near future.

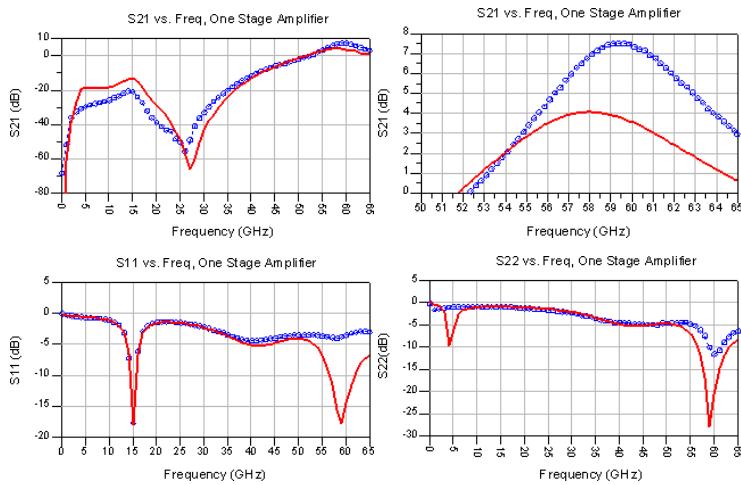


Figure 23: S-parameter measurement (circles) and simulations (solid line) of the 90nm 60GHz amplifier.

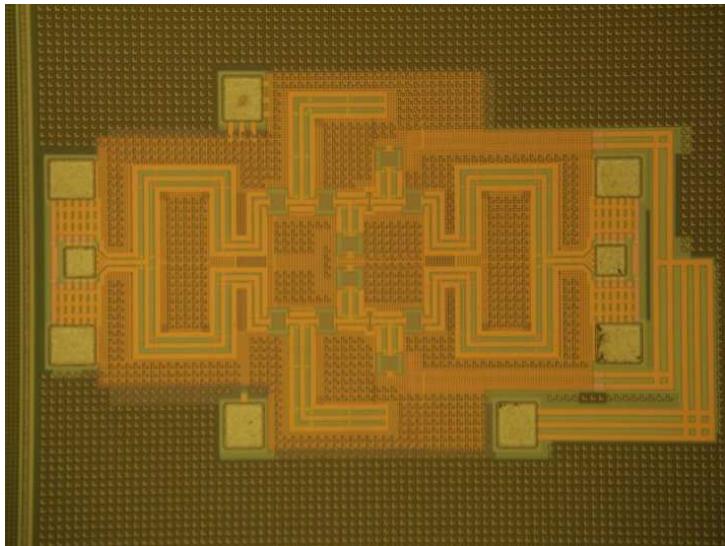


Figure 24: Layout of the 90nm 60GHz amplifier.

Within this framework, this paper has focused on power combining technique. Three different approaches were described in detail: the lumped element LC balun, a transformer-based approach and a CMOS implementation of the microwave Wilkinson combiner. Each approach was clarified with CMOS implementations to demonstrate their feasibility.

Together with other techniques such as stacked devices [22] and digital predistortion, it can be concluded that power combining will become crucial to achieve both high output power, high average efficiency and high CMOS integration.

References

- [1] M. Steyaert, F. Gobert, C. Hermans, P. Reynaert, and B. Serneels, “Digital Communication Systems: the Problem of Analog Interface Circuits,” in *Proceedings of the European Solid-State Circuits Conference*, September 2005, pp. 423–426.
- [2] M. Steyaert, M. Borremans, J. Janssens, B. D. Muer, and N. Itoh, “A Single-Chip CMOS Transceiver for DCS-1800 Wireless Communications,” in *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, San Francisco, USA, February 1998, pp. 48–49.
- [3] I. Aoki, S. D. Kee, D. B. Rutledge, and A. Hajimiri, “Fully Integrated CMOS Power Amplifier Design Using the Distributed Active Transformer Architecture,” *IEEE J. Solid-State Circuits*, vol. 37, no. 3, pp. 371–383, March 2002.
- [4] ——, “Distributed Active Transformer—A New Power-Combining and Impedance-Transformation Technique,” *IEEE Trans. Microwave Theory Tech.*, vol. 50, no. 1, January 2002.
- [5] B. Sahu and G. A. Rincón-Mora, “A High-Efficiency Linear RF Power Amplifier With a Power-Tracking Dynamically Adaptive Buck-Boost Supply,” *IEEE Trans. Microwave Theory Tech.*, vol. 52, no. 1, pp. 112–120, January 2004.
- [6] W. H. Doherty, “A New High Efficiency Power Amplifier for Modulated Waves,” *Proceedings IRE*, vol. 24, no. 9, pp. 1163–1182, September 1936.
- [7] N. Wongkomet, L. Tee, and P. R. Gray, “A 1.7GHz 1.5W CMOS RF Doherty Power Amplifier for Wireless Communication,” in *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, February 2006, pp. 486–487.

- [8] J. T. Stauth, "Dynamic Power Supply Design for High-Efficiency Wireless Transmitters," Master's thesis, U.C. Berkeley, May 2006.
- [9] P. Reynaert and M. Steyaert, "A 1.75 GHz Polar Modulated CMOS RF Power Amplifier for GSM-EDGE," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 12, pp. 2598–2608, December 2005.
- [10] W. Bakalski, W. Simbürger, R. Thüringer, A. Vasylyev, and A. L. Scholtz, "A fully integrated 5.3 GHz, 2.4 V, 0.3 W SiGe-bipolar power amplifier with 50Ω output," in *Proceedings of the European Solid-State Circuits Conference*, September 2003, pp. 561–564.
- [11] W. Bakalski, W. Simbürger, H. Knapp, H. Wohlmuth, and A. L. Scholtz, "Lumped and Distributed Lattice-type LC-Baluns," in *Proceedings of the International Microwave Symposium*, vol. 1, June 2002, pp. 209–212.
- [12] C. Lorenz, "Schaltungsanordnung zum Übergang von einer symmetrischen, elektrischen Anordnung zu einer unsymmetrischen, insbesondere bei Hochfrequenzanwendungen," German Patent no. 603816, April 1932.
- [13] P. Reynaert and M. Steyaert, "A Fully Integrated CMOS RF Power Amplifier with Parallel Power Combining and Power Control," in *Proceedings of the Asian Solid-State Circuits Conference*, November 2005, pp. 137–140.
- [14] F. H. Raab and N. O. Sokal, "Transistor Power Losses in the Class E Tuned Power Amplifier," *IEEE J. Solid-State Circuits*, vol. sc-13, no. 6, pp. 912–914, December 1978.
- [15] C. Yoo and Q. Huang, "A Common-Gate Switched 0.9-W Class-E Power Amplifier with 41% PAE in $0.25 - \mu m$ CMOS," *IEEE J. Solid-State Circuits*, vol. 36, no. 3, pp. 823–830, May 2001.
- [16] P. Reynaert and M. Steyaert, *RF Power Amplifiers for Mobile Communications*, ser. Analog Circuits and Signal Processing. The Netherlands: Springer, 2006.
- [17] G. Liu, T.-J. K. Liu, and A. M. Niknejad, "A 1.2V, 2.4GHz Fully Integrated Linear CMOS Power Amplifier with Efficiency Enhancement," in *Proceedings of the Custom Integrated Circuits Conference*, Septemeber 2006, pp. 141–144.
- [18] P. Haldi, G. Liu, and A. M. Niknejad, "CMOS compatible transformer power combiner," *IEE Electronics Letters*, vol. 42, no. 19, pp. 1091–1092, September 2006.

- [19] C. H. Doan, S. Emami, A. M. Niknejad, and R. W. Brodersen, "Millimeter-wave CMOS design," *IEEE J. Solid-State Circuits*, vol. 40, pp. 144–155, January 2005.
- [20] ——, "Design of CMOS for 60GHz applications," in *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, February 2004, pp. 440–538.
- [21] T. S. D. Cheung and J. R. Long, "Shielded Passive Devices for Silicon-Based Monolithic Microwave and Millimeter-wave Integrated Circuits," *IEEE J. Solid-State Circuits*, vol. 41, no. 5, pp. 1183–1200, May 2006.
- [22] B. Serneels, T. Piessens, M. Steyaert, and W. Dehaene, "A high-voltage output driver in a standard 2.5V 0.25 μ m CMOS technology," *IEEE J. Solid-State Circuits*, vol. 40, no. 3, pp. 576–583, March 2005.

Switched RF Transmitters

Willem Lafleure, Michiel Steyaert and Jan Craninckx ¹

KULEuven-ESAT-MICAS

Kasteelpark Arenberg 10

3001 Leuven, Belgium

willem.lafleure@esat.kuleuven.ac.be

Abstract

To overcome the limitations of traditional linear amplifiers: a reduced efficiency when operated with amplitude modulated signals, a switched linearization technique is presented. This technique is analog to the linear amplification by a class D-amplifier at baseband. By pulse-width modulation of the signal envelope, the class-D operation is extended towards RF systems too.

The spurious emissions generated by the switching operation, are well separated from the signal band by the use of an asynchronous pulse-width modulator. Measurements on a test-chip prove the feasibility of the switching technique.

1. Introduction

The ever-increasing demand for application with a higher datarate and a higher degree of autonomy has led to the use of complex modulation techniques. The design of the circuitry should be power efficient to save battery lifetime. The combination of the complex signals and the need for efficiency is a severe problem in the design of the analog circuitry. The presented solution is a possible efficiency-improving technique for these kind of signals.

The first section describes the evolution of the modulation techniques to deal with the increasing demand for bandwidth. In the next section, the influence of the properties of these signals on the design of the circuitry is discussed. Some solutions to overcome the problem are the subject of the third section.

Next, the design of the switching modulator is explained in detail. In the last paragraph, measurement results on a prototype are discussed.

¹J. Craninckx is with IMEC, Kapeldreef 75, 3001 Leuven, Belgium

2. Envelope variations

The maximal amount of data that can be transmitted through a frequency channel in free air is limited by: the available bandwidth, the maximal transmit power, and the noise floor. This is expressed by the Shannon limit:

$$R = B \log_2 \left(1 + \frac{S}{N} \right)$$

Elementary modulation techniques like BPSK (Binary Phase Shift Keying), OOK (On-Off Keying), have spectra with relatively large side-lobes, determined by a $\sin(x)/x$ -like shape. This results in an inefficient use of the channel capacity. The only way to increase the datarate is to improve the efficiency of the modulation. Since the available bandwidth is to be shared among all the users, the noise floor is a fixed physical constant, and the transmit power should be kept as low as possible for reasons of battery lifetime, spectral efficiency improvement is the only solution.

Two main categories of modulation efficiency improvement techniques are distinguished: OFDM-like techniques (orthogonal frequency division multiplexing) adds several low-datarate components orthogonally together. By the orthogonal property, these signals do not interfere with each other, but can be spectral very dense. The Shannon limit is approximated much better compared to the simple modulation schemes. Another approach uses pulse-shaping filters to eliminate the higher-order, out-of-band harmonics that would occur in a high-datarate, simple modulation scheme. In this way, higher datarates can be achieved in the same channel bandwidth.

The common effect of every high-efficient modulation is in the time domain. Due to addition of several, uncorrelated signals (OFDM-like techniques) or band-limited filtering (pulse-shaping, ...), a non-constant envelope signal arises from otherwise constant-envelope signals.

Figure 1 presents some wireless communication systems. It demonstrates non-constant envelope modulation becomes common for more advanced systems.

The amount of envelope variation is evaluated in the probability density function (PDF) of the instantaneous output power. A rough idea of the main properties of this curve is also expressed by the ‘peak-to-average power ratio’ or Crest factor. Figure 2 shows the probability density for a UMTS-signal. The envelope modulation in this system is due to a pulse-shaping filter. The variations in output power are not only due to modulation. In order to save battery power, the standards also provide the possibility to transmit at reduced power levels when transmit quality is too good. The data for the black line in figure 2 is obtained from a weighted

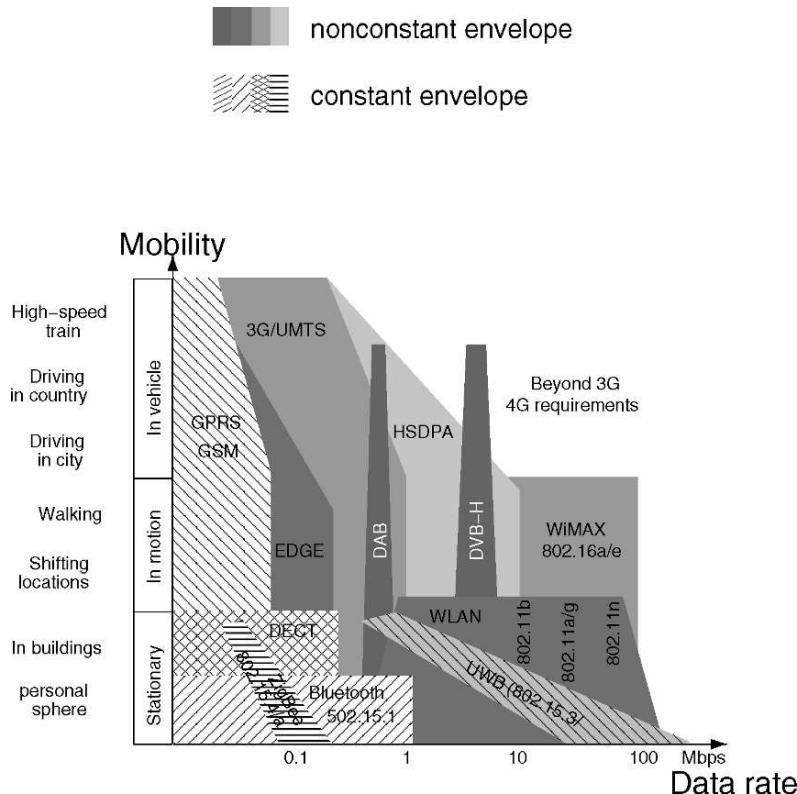


Figure 1: Constant- and nonconstant envelope systems

averaging of the full-power PDF (gray line), using real-world probabilities for each power level. [1] This is the probability density of the long-term behavior of the transmitter.

3. Power amplifier classes

First, a distinction between linear and switching (non-linear) amplifiers is to be made. Amplifiers that are continuously biased in a linear region (class A, AB, ...), feature a linear input- output relationship, but the efficiency drops due to the constant bias current. This is especially true for small output signals, since the relative importance of the bias current is larger in that case. For a class A amplifier (constant current bias), the efficiency as a function of output voltage is given by:

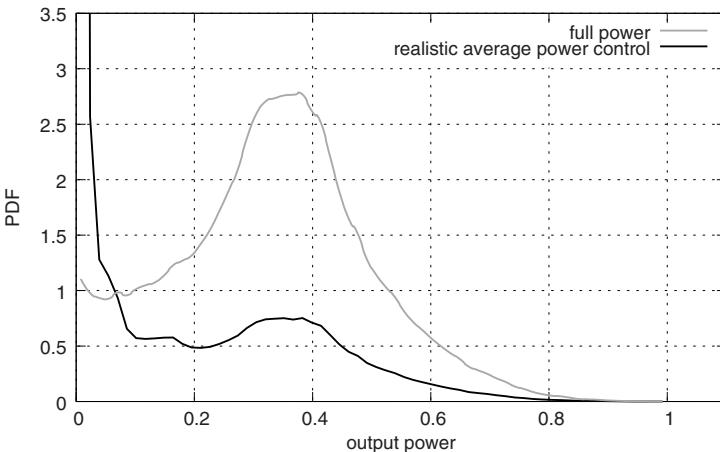


Figure 2: Probability density of UMTS transmitted power

$$\eta_A = \frac{V_{out}^2 / R_l}{I_{bias} V_{dd}} \sim V_{out}^2 \sim P_{out}$$

For class B amplifiers, where the bias current is proportional to the signal, this becomes:

$$\eta_B = \frac{V_{out}^2 / R_l}{\alpha V_{out} V_{dd}} \sim V_{out} \sim \sqrt{P_{out}}$$

In both cases, the efficiency for low output levels is proportionally low.

The other extreme, where theoretically no power is dissipated due to non-overlapping current- and voltage signals, are the class F, E, ... amplifiers. The nonoverlapping signals are obtained by a resonant circuit. Optimal operation is only observed within a relative small frequency band. These amplifiers benefit from a good efficiency, but no linearity can be achieved, since only the switching instant is transferred to the output. The combination of these properties make the class E, F amplifiers an excellent choice for constant-envelope, high-frequency signals.

For baseband amplifiers with a high peak-to-average power ratio, a highly efficient solution is the class D, or pulse-width-modulated (PWM) amplifier. By only switching of the output devices, there is never a voltage drop over the device when delivering current. So theoretically no power is dissipated in the amplifier.

	sinewave-test	full power UMTS-PDF	real-world UMTS-PDF
Class A	50 %	17%	< 5%
Class B	78 %	50%	43%
Class D	100 %	100%	100%

Table 1: Theoretical maximal efficiency with UMTS-envelope-signals

The pulse-width modulated technique is very commonly used in every low-cost, baseband systems. It's switching behavior eases also the integration together with the digital circuitry. The efficiency η_D is theoretically independent of the output power.

RF-amplifiers for systems with both an amplitude- and phase component, as needed for the system described in the first paragraph, can either be designed as a linear amplifier (class AB), to benefit from the intrinsic linearity, at the cost of a lower efficiency. A higher efficiency is offered by switching amplifiers, but additional techniques must be provided to improve the linearity. The existing solutions are discussed in the next section, and a new topology is introduced.

For a known PDF of a certain signal, the theoretical maximal efficiency when implementing a certain class of amplifier can be calculated. The table below is calculated in the assumption of theoretically perfect amplifiers (so 100% efficiency at full output power). This poses an upper limit on the practically achievable efficiency. Due to the high probability of the lower signal amplitudes, the average efficiency, and so battery lifetime, is largely reduced when linear amplifiers are used. The choice of the amplifier class is thus extremely important.

4. Existing solutions

Since the switching class E amplifier is unable to deal with envelope variations, solutions are to be found to overcome this limitation. Both analog and digital solutions are possible.

4.1. Supply modulation

As the class E amplifier has no linear relationship from the input gate to the output node, other nodes are to be modulated to introduce the AM signal. The relationship between the power supply voltage and the envelope of the output signal, is

rather linear, except for low voltages. Since all of the drain current of the power amplifier is to be supplied by the supply modulator, the efficiency requirements, along with linearity specs, are transferred to this block. The overall efficiency is the product of the efficiency of the supply modulator and the efficiency of the switching amplifier. Due to the modulation of the PA parameters, its operation shifts slightly from the unmodulated behavior, resulting in a distorted version of the output signal. As a result of these hard requirements for this supply modulator, practical implementations of this technique achieve linearity, but typically only with the efficiency of a class B amplifier. [2–5]

4.2. Doherty technique

The technique of combination of the output signals of different amplifiers is complementary to the efficiency-improving techniques of elementary amplifiers. By means of the addition of the outputs of two (or more) amplifiers, benefit is taken from the increased efficiency closer to the amplifiers saturation point. This can be applied to baseband signals, but as well to RF-signals. The problem however is in the efficient, linear addition of several high-power signals.

Summation by the active elements of the output stage, as performed in [6], reduces the design of this stage with one degree of freedom, so resulting in a less efficient design. A passive combination by means of a poly-phase network requires additional passives, but is a more efficient way to combine independent signals together. [] When a lossless addition of two independent amplifiers is possible, theoretically a large efficiency improvement is possible. [7]

The Doherty technique is considered very useful, but is actually a parallel approach to the efficiency-improvement of the basic amplifiers.

4.3. Digital modulation

A full digital approach of amplitude modulation is switching the signal on- and off for longer or shorter periods, proportional to the instantaneous envelope power. This technique is much like the class D-technique used in baseband systems. The class D technique is suited to drive a linear signal through the gate of a switching amplifier, eliminating the need for a linear modulator and its disadvantages as described above. However direct oversampling of the signal is only possible when technology's f_T permits. In RF applications, the signal frequency is typically in the order of magnitude of some GHz. Classical class D amplifiers require a large oversampling ratio, so in commercial CMOS technologies for mass production, this is only feasible for signal frequencies up to some 100MHz. [8]

Pulse width modulation of the low-frequency amplitude signal only, as will be explained further in this work, is possible within some limits. The mean switching frequency of the modulator should be high enough, to avoid switching spurs into the system band. Out of the system band, the switching products can be filtered. Systems with a mean pulse frequency within the band are practically useless. Since also the efficiency is determined by the filtering (see below), spurs in the signal band result in a class B efficiency. Very high operating frequencies of the pulse width modulator on the other hand, relax the filter specs at the cost of a tremendous increase of the power needed in the pulse width modulator. Since power saving was the original goal, these systems will practically have no use.

The switching spurs are the main source of concerns in digital modulated systems. The total magnitude of these spurs can easily be calculated from the signals PDF, independent of the modulation technique used. For any pulse-with modulated signal, the total energy in the coded signal is proportional to the instantaneous duty cycle (α). The restored signal has an energy proportional to the duty cycle squared.

The difference between these two signals is the energy present in the quantization noise. This is illustrated by figure 3. The energy of the quantization noise remains the same whether expressed in the time domain (as above) or in the frequency domain (spurs). The sum of the product of the signal's PDF and $\alpha - \alpha^2$ expresses the total energy present in the quantization noise. Figure 4 shows this total noise vs. modulation percentage for a single sinewave. The exact distribution of this total amount of quantization noise over the frequency band will depend on the actual implementation of the modulator.

The signal to total quantization noise is given by:

$$\frac{P_{sig}}{P_{noise}} \sim \frac{\alpha^2}{\alpha - \alpha^2} = \frac{\alpha}{1 - \alpha}$$

For low duty cycles (α), the quantization noise becomes relatively more important. This explains why the efficiency decreases for smaller signals if this quantization noise is partially dissipated.

5. RF pulse-modulated amplifiers

5.1. Transmitter architecture

The vectors containing the information to be transmitted can be described by a Cartesian or by a polar decomposition. The Cartesian modulator operates using the in-phase (I) and quadrature (Q) projection of the vectors. After combination

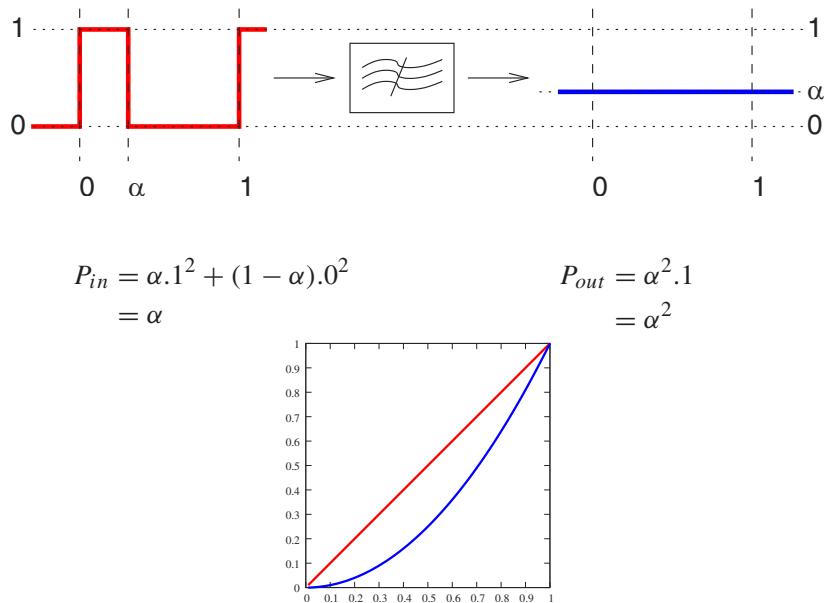


Figure 3: Signal and total quantization noise vs. duty cycle

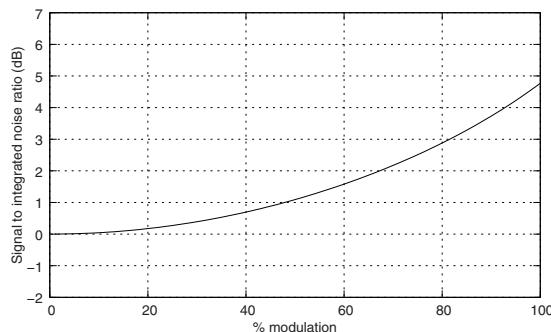


Figure 4: Total amount of quantization noise for a sinewave signal of different amplitudes

of the I and Q path, the (non-constant envelope) RF-signal is obtained. In a polar modulator, the amplitude modulated (AM) and phase modulated (PM) signals

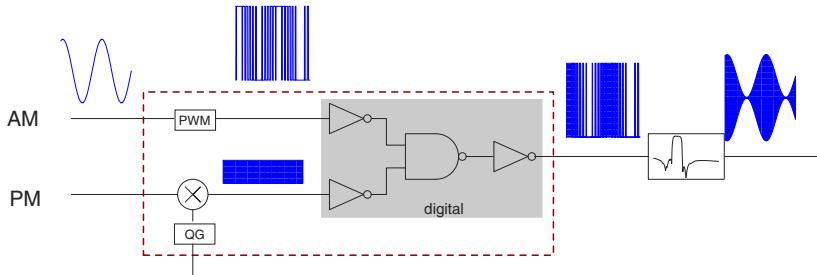


Figure 5: Principle of RF PWM

are separated. The constant-envelope PM signal can be efficiently amplified by switching circuitry. The envelope information is restored afterward on it, resulting in the non-constant envelope RF-signal.

Opposite to the Cartesian technique, where the output signal is synthesized early in the transmit chain, the main advantage of this technique is the transformation of the phase part of the signal into a binary signal, thus eliminating the need for inefficient class A or B amplifiers in this path. The polar representation is suitable for integration with an efficient, switching amplifier. Reconstructing the RF-signal at the end of the modulator however requires some additional circuitry, but this does not compensate the intrinsically low efficiency of a Cartesian modulator.

A direct modulator is used in this work. Direct upconversion of the baseband signals removes the requirement for filtering the image-IF signal, while low-frequency noise is not an issue in upconvertors dealing with high signal swings.

5.2. Amplitude modulator

The proposed solution expands class D operation to frequencies in the GHz range, by means of pulse-width modulation of the envelope of the output signal, instead of the infeasible direct oversampling of the signal. This signal is afterwards modulated on the PM-signal. The desired non-constant envelope signal is reconstructed by the filters of the transmitter. This reconstruction is the removal of the unwanted switching products at multiples of the mean switching frequency of the pulse-width modulator. For this reason, this frequency is best chosen as high as possible. Anyway, the switching products should be well out of the signal band. Since the amplitude signal and pulse-width modulated signal have a relative small

bandwidth compared to the RF frequency, this technique is very well suited for implementation in standard CMOS. The frequency shifting operation of the up-conversion however, reduces the relative separation between the signal and the quantization noise band decreases compared to a baseband class D solution, since these bands are all shifted over the same frequency. A steep filter is required to eliminate the sidebands. The required small-band filters are however already present in the transmitter, only maximal use of their filtering capacity is made by this technique.

So there is a trade-off between the ease of implementation and the filtering requirements. In this work, this problem is compensated by the choice for a pulse modulator with a good separation between the signal and the quantization noise band, so relaxing the need for steep filter specifications. An un-clocked, self-oscillating solution is used for this purpose.

The operation principle of this asynchronously oscillating circuit is shown in figure 6. A single SOPA-loop consists of a comparator fed back by a loop filter of order ≤ 3 . The loop is unstable and will oscillate at one single frequency (limit cycle). When a signal is applied at the input of the comparator, intermodulation with the limit cycle signal yields a highly linear pulse-width modulated version of the input signal [9]. Two SOPA-loops are coupled resistively together, in this way, the limit cycles are attracted to each other, becoming a common mode signal, and improving the spectral purity. Due to the asynchronous nature, all of the switching spurs are concentrated in a narrow band near the limit cycle frequency and its multiples.

Compared to clocked types of PWM-modulators (natural sampling, delta-sigma, ...) [10], spurs are much better separated from the signal band, due to the absence of a time discretisation. This can lower the mean switching frequency that has to be used for a certain amount of available filtering. Or alternatively, lower the necessary amount of filtering, for a given limit cycle frequency. Due to the limited amount of filtering available in the RF-band, the SOPA structure is selected for this last purpose in this work. Due to its simple structure, the intrinsic current consumption of the modulator is very low.

Digital logic, in this case an xor-gate is used to synthesize the combined asynchronous PWM-output signal from the two individual pulsed signals. The on-off signal obtained in this way is combined with the constant-envelope PM-signal using a digital nand-gate. (figure 5) In this way the amplitude information is modulated on the PM-signal. Since both signals are binary, the nand-output is also a binary signal. The output driver on this test-chip is a digital buffer. So all signal processing is done efficiently using digital, switching circuitry.

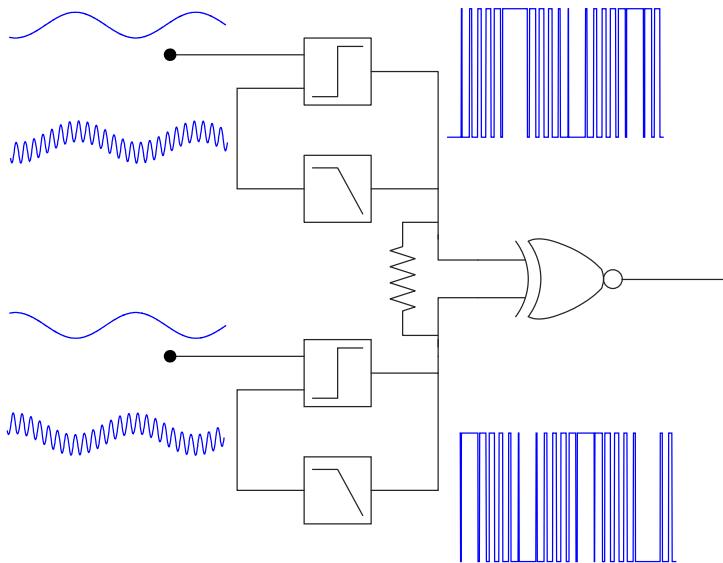


Figure 6: Principle of SOPA pulse width modulation

The spurious emissions are the main concern in this type of digital modulators. A higher limit cycle frequency takes more benefit from the available filtering, but causes more switching losses.

Techniques based on combinations of signals from different amplifiers to obtain linear behavior, like the Doherty technique, have also proven to be effective for efficiency improvement. [7] The technique presented in this work is however complementary. Since the two approaches are entirely separated, any combination is possible.

6. Circuit design

A testchip was designed to verify the above described techniques. The UMTS mobile transmit frequency band is $1.92 - 1.98\text{GHz}$, the test-modulation as described in this spec [11] is similar to 8-PSK, being filtered afterward by a RRC (Root Raised Cosine) filter in order to limit the required bandwidth. It is this filtering that gives rise to the amplitude component. The nominal chiprate of the spread and scrambled data is 3.84Mcps .

In order to be able to drive an integrated or external power amplifier, the up-converter should be able to deliver at least $-10.0dBm$ of output power.

The limit cycle frequency is designed for an offset of about $120MHz$ of the carrier frequency, as a trade-off between switching losses and filtering, so maximal benefit is taken from the already present transmit filter. In this way, the design of the on-chip SOPA can be kept rather simple, so not consuming lots of power. The spurious limit cycle products are situated well apart from the signal band, relaxing the filtering specs; using the standard filtering, these spurs can be suppressed to a level within the emission spec.

The efficiency of this technique can only be evaluated in combination with the signal restoring filter. Without filtering, the reactive energy in the limit cycle products would be dissipated in the load too. The transmit filters based on the SAW technology feature a matched impedance in the signal band, and a high input impedance for frequencies out of the signal band. Due to this high input impedance, only a small part of the limit cycle signals are dissipated, degrading the efficiency only slightly. The efficiency vs. output power for class A and class B topologies are shown in figure 11, along with the same curve for the pulse-width technique, using a commercially available SAW filter. While for a perfect band-pass filter, this curve would be entirely independent of the power, the efficiency drops for very small signals, due to the relatively higher quantization noise.

The variations over the band are explained by the frequency depending input impedance of the used filter. Figure 7 shows the resistive and capacitive part of the input impedance. For a different channel frequency, a different impedance is seen by the limit cycle products, resulting in a different transmitter efficiency.

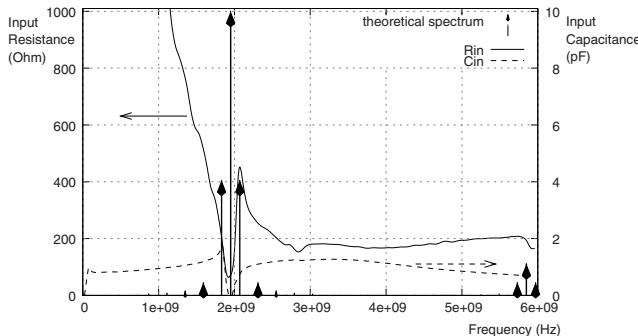


Figure 7: Real and Imaginary part of the input impedance of the filter

The curve of figure 11 shows the average efficiency over the band, the crosses indicate the maximal variations over the band. This technique clearly outperforms the class-B technique.

The entire circuit implemented on the chip is shown in figure 8. The design of the phase modulator is explained in detail in [12], an injection locked oscillator is used to obtain high gain in a single stage.

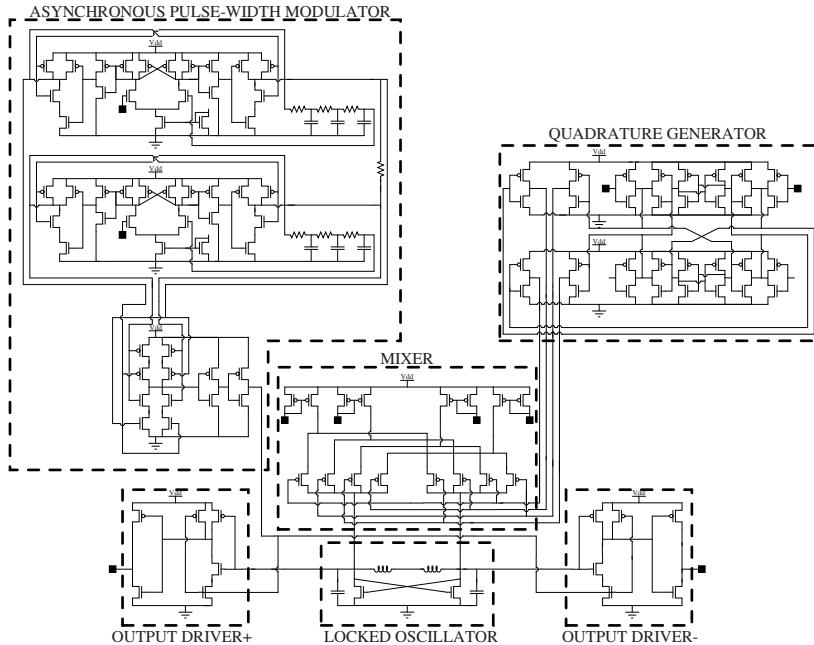


Figure 8: The implemented circuit

7. Measurements

Figure 9 shows the envelope of the output signal when a two-tone envelope test-signal with frequencies $0.4MHz$ and $0.5MHz$ is AM-modulated on a carrier.

The measured output spectrum of this signal, before any filtering, is presented in figure 10 by the gray line. The typical wide band spectrum of switched-modulation techniques is noticed: the spurs in the spectrum are the upconverted limit cycle intermodulation products. However if the signal is passed the filter these spurs are drastically reduced. In figure 10, the reduction by the filter is pre-

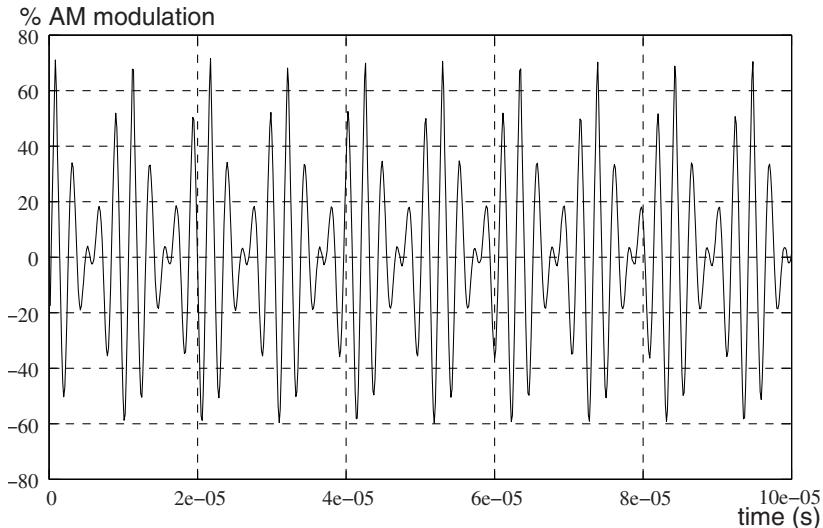


Figure 9: Measured envelope signal after demodulation - two-tone test

sented by the dot-strip line, using data from a commercially available saw-filter. From the measured output spectrum the spurious emissions are calculated and presented by the full-line. This is an underestimation of the available filtering, since in a real-world transmitter, the signal will experience more filtering from the band-limited behavior of the power amplifier, matching network, antenna, etc. It is also clear from this figure the quantization noise is concentrated in some narrow bands, keeping the close in-band spectrum near the carrier clean of quantization noise products.

The unmodulated RF-output power is 8.26dBm , with a drain efficiency of 35.6%. The insertion loss of the filter (about 2.2dB) is not relevant, since the presented technique only makes use of the already present transmit filter and does not introduce additional filters. The efficiency of the transmitter+filter combination is more constant over the various output levels, as opposite to supply-modulating techniques, where the efficiency is reduced with the efficiency of the supply modulator. Since the emitted power in a real transmitter is most of the time only a fraction of the maximal power, this greatly improves the overall efficiency, and so the battery lifetime. The power dissipation in the PWM-generator is neglectable compared to the output stage when integrated in the entire transmitter 1.8mW .

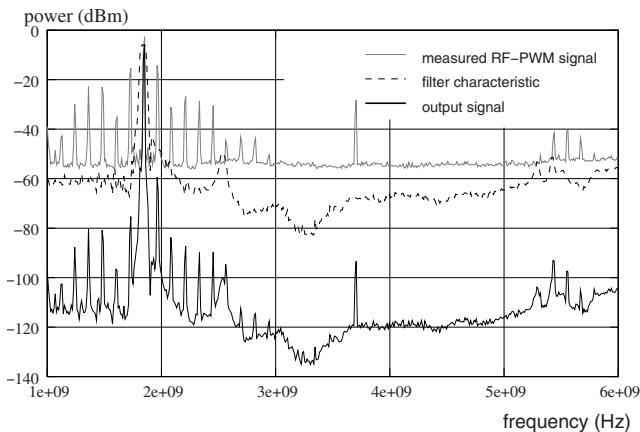


Figure 10: Signal spectra

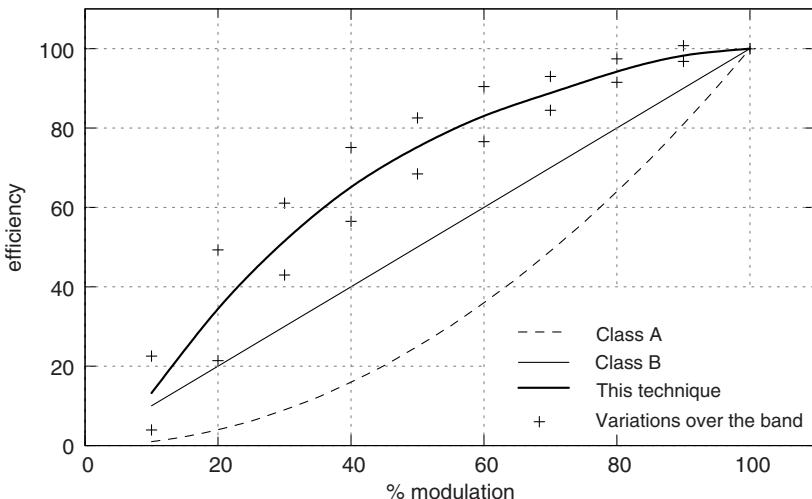


Figure 11: Efficiency vs. output power for several amplifier classes

Root Raised Cosine(RRC)-filtered (roll-off factor 0.22) 8-PSK modulation was used to test UMTS-compatibility. The Crest factor of the envelope signal is 3.2 dB, due to the RRC-pulse-shaping filter. Figure 12 shows a demodulated test-signal,

the bitrate is $3.84 Mbit/s$ with a random bit-input. The measured EVM is 7.5%, being much smaller than the required 17.5% in the spec. A part of the remaining error is believed to be due mismatch in the measurement setup. The variation of the mean error vector is shown in figure 13. It is clear there is not much variation in the performance over the entire UMTS mobile transmit frequency range.

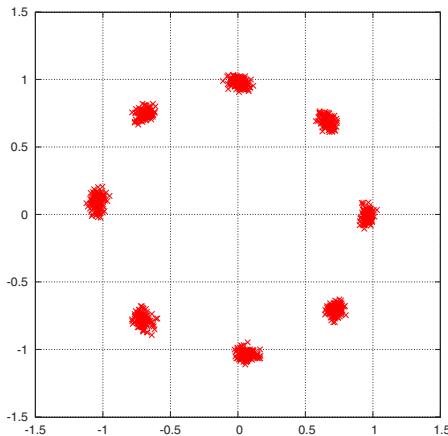


Figure 12: Measured output constellation (3.84 MHz): RRC-shaped 8PSK

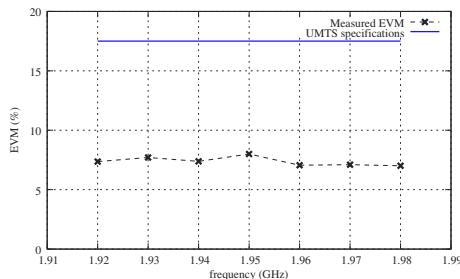


Figure 13: Measured variation of the EVM over the frequency band

The chip microphotograph is shown in figure 14, the size is $1.5mm$ by $1.15mm$. A large part of the chip area is occupied with bonding paths and decoupling capacitance in order to get rid of high-frequency interference between building blocks that are not synchronized.

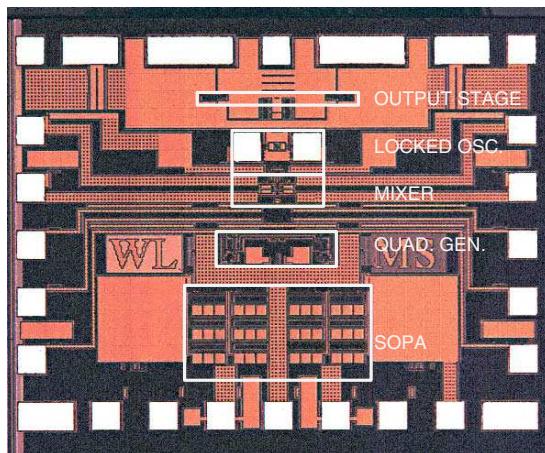


Figure 14: Chip micro-photograph

8. Conclusions

A method to implement a digital polar modulator is introduced. The feasibility of the RF pulse-width modulation technique was proved by measurements on a $0.18\mu\text{m}$ CMOS prototype. Using asynchronously generated on-off modulation of the RF signal, an efficient switching-type of PA can be used for amplitude modulated signals. No extra filters additional the already present filtering in a transmitter, are needed to filter out the spurs and reconstruct the signal while maintaining the amplifier efficiency.

References

- [1] S. Lönn, U. Forssén, P. Vecchia, A. Ahlbom, and M. Fechting, “Output power levels from mobile phones in different geographical areas; implications for exposure assessment,” in *BMJ journal of Occupational and Environmental Medicine*, 2004, pp. 769–772.
- [2] E. Mostafa, P. Jeyanandh, and K. Soumynath, “A 90-nm cmos doherty power amplifier with minimum am-pm distortion,” in *IEEE Journal of Solid-state Circuits*, 2006, pp. 1323–1332.

- [3] Y. Palaskas, S. Taylor, R. Pellerano, I. Rippke, R. B. A. Ravi, H. Lakdawala, and K. Soumyanath, “A 5 ghz 20 dbm power amplifier with digitally assisted am-pm correction in a 90-nm cmos process,” Aug. 2006, pp. 1757–1763.
- [4] M. R. Elliott, T. Montalvo, B. P. Jeffries, F. Murden, J. Strange, A. Hill, S. Nandipaku, and J. Harrebek, “A polar modulator transmitter for gsm/edge,” in *IEEE Journal of Solid-state Circuits*, Dec. 2004, pp. 2190–2200.
- [5] P. Reynaert and M. Steyaert, “A 1.75 ghz polar modulated cmos rf power amplifier for gsm-edge,” in *IEEE Journal of Solid-state Circuits*, Dec. 2005, pp. 2598–2608.
- [6] R. Staszewski, J. Wallberg, S. Rezeq, C.-M. Hung, O. Eliezer, S. Vemulapalli, C. Fernando, K. Maggio, R. Staszewski, N. Barton, M.-C. Lee, P. Cruise, M. Entezari, K. Muhammad, and D. Leipold, “All-digital pll and transmitter for mobile phones,” in *IEEE Journal of Solid-state Circuits*, Dec. 2005, pp. 2469–2482.
- [7] N. Wongkomet, L. Tee, and P. R. Gray, “A +31.5 dbm cmos rf doherty power amplifier for wireless communications,” in *IEEE Journal of Solid-state Circuits*, Dec. 2006, pp. 2852–2859.
- [8] T. Johnson and S. P. Stapleton, “Rf class-d amplification with bandpass sigma-delta modulator drive signals,” in *Transactions on Circuits and Systems I*, Dec. 2006, pp. 2507–2520.
- [9] T. Piessens and M. Steyaert, “Highly efficient xDSL line drivers in 0.35 um cmos using a self-oscillating power amplifier,” *IEEE Journal on Solid-State Circuits*, vol. 38, no. 1, pp. 22–29, jan 2003.
- [10] C. Berland, I. Hibon, J. Bercher, M. Villegas, D. Belot, D. Pache, and V. L. Goascoz, “A transmitter architecture for nonconstant envelope modulation,” *IEEE Transactions on circuits and systems II*, vol. 53, no. 1, pp. 13–17, jan 2006.
- [11] E. T. S. Institute, “Universal mobile telecommunications system, etsi ts 125 101 v3.3.0.”
- [12] W. Laflere and M. Steyaert, “An injection-locked upconversion mixer,” *Proceedings of the European Microwave association*, vol. 2, pp. 167–172, June 2006.

High-Speed Serial Wired Interface for Mobile Applications

Gerrit W. den Besten

NXP Semiconductors

Research Eindhoven (NL)

gerrit.den.besten@nxp.com

Abstract

This paper presents a power-efficient high-speed serial interface solution, which is targeted for use in battery-operated mobile or handheld devices. Both transceiver topology and signaling protocol are described. High-speed differential and CMOS signaling are merged on the same wires. The signaling scheme contains link power management features and provides multiple communication modes to adapt efficiently to bandwidth needs. This interface has been implemented in 65nm CMOS and measurement results are shown.

1. Introduction

Increasing interface bandwidth demands and the desire to reduce pin count drive the migration from classic CMOS busses towards high-speed serial interfaces. Especially routing data connections through complex mechanical constructions, like a phone hinge, requires a compact and robust interface. For mobile devices, power consumption is extremely important, which necessitates new solutions instead of applying existing standards. Besides the technical merits, new interface standards in this domain help to reduce interface diversity, stimulate multi-vendor IC interoperability, and enable a modular system approach. The enabling factor for these developments is the definition of a suitable standardized physical layer technology [1].

This paper starts with an analysis of signaling types, bandwidth, power, IO structures, and implementation trade-offs (Sections 1-4). The resulting preferred IO topology is the starting point to present the complete physical layer solution including signaling protocols (Sections 5-8). Results and conclusions are given in sections 9 and 10, respectively.

The presented serial interface solution is optimally suited for applications like camera's, displays, and other bandwidth-demanding chip-to-chip and board-to-board connections in mobile and handheld devices.

2. Quasi-static versus transmission-line signaling

For many years, unterminated CMOS rail-to-rail signaling has been the predominant interface technology in mobile applications for reasons of zero static power and hardware simplicity. However, the underlying quasi-static assumption [2], which does not take propagation time into account, cannot be fulfilled anymore for the required bandwidth with a reasonable number of wires. Furthermore, for CMOS IO the current return path is not always close to the single-ended signal, which may create a substantial current loop. This can generate EMI problems especially for the strong current spikes related to CMOS signaling; see figure 1.

When signal propagation time becomes of the same order of magnitude (1/5 or larger) as the signal transition time (not bit time!), the interconnect lines for signal transport must be considered as transmission lines and these lines need to be terminated for signal integrity. Hence, the need for line termination can either arise from higher speed and/or longer distance requirements. For a strongly slew-rate limited signal, the limit for the quasi-static assumption is roughly indicated by [3]: Speed[Gbps] x Distance[cm] = 2. For example for a serial link >100Mb/s with a length >20cm, termination is at least desirable if not necessary. Unfortunately, line termination implies static power consumption. Power consumption can be reduced with a lower signal swing. Differential signaling enables a very small swing, thanks to the easily detectable differential zero-crossings, while it also minimizes the physical size of the current loop which is beneficial with respect to EMI.

3. Power

In conventional unterminated CMOS buses the energy cost is ‘pay-per-bit’ (actually per transition), determined by charging and discharging of interconnect lines and connected devices. Low-Voltage CMOS (LVC MOS) IO is energy efficient for short links, but energy usage scales linearly with distance. Furthermore, because achievable speed is inversely proportional to distance, this solution becomes less attractive for high bandwidth interfaces, as it enforces massive parallelism.

For power efficiency comparison purposes, it is useful to consider the Energy/bit metric, which equals the Power/speed ratio: $pJ/bit = mW/Gbps$.

For terminated high-speed links, a substantial part of the power consumption is fixed, due to line termination and permanently biased circuits, so the energy cost is more like ‘pay-per-time’. Figure 1 illustrates the two cases.

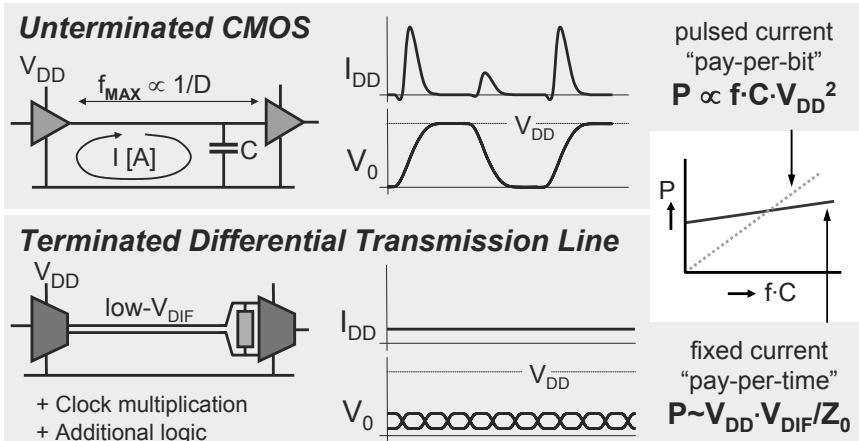


Figure 1: Comparison of unterminated CMOS signaling and high-speed differential transmission

In order to obtain good power efficiency for terminated (differential) links, this fixed energy should be optimally exploited by transmitting the maximum number of bits per time when operational. This means utilizing the maximum available bandwidth during high-speed data bursts and shutting down the link for the rest of time, which implies a packet-based transmission scheme. Because the additional overhead periods for starting and stopping transmission also impact the overall power efficiency, it is important that these periods are short, especially if very frequently small packets have to be transported. Clustering of data may reduce overhead cost and improve power efficiency.

4. High-Speed IO technology

The IO structure of many high-speed interfaces consists of one or more current-steering differential pairs. However, this is not optimal with respect to power consumption and continuing process scaling. This can be easily proven by starting from the single-ended Thevenin and Norton equivalent circuits for the line driver as shown in figure 2a (top). These two circuits by definition show the same behavior on their output pins, but it is important to notice some (internal) differences, even if in both cases the lines are appropriately terminated:

Inside the Thevenin equivalent a voltage source with the double voltage swing $2 \cdot V_o (=2 \cdot I \cdot R)$ exists, while there is only one current (I) that flows through the series-connected source and far-end termination resistors.

Inside the Norton equivalent there is a current source ($2 \cdot I$) that drives against the voltage potential, which is practically not achievable. This current is split by the parallel-connected source and far-end resistors, causing only $I \cdot R (=V_o)$ as highest node voltage.

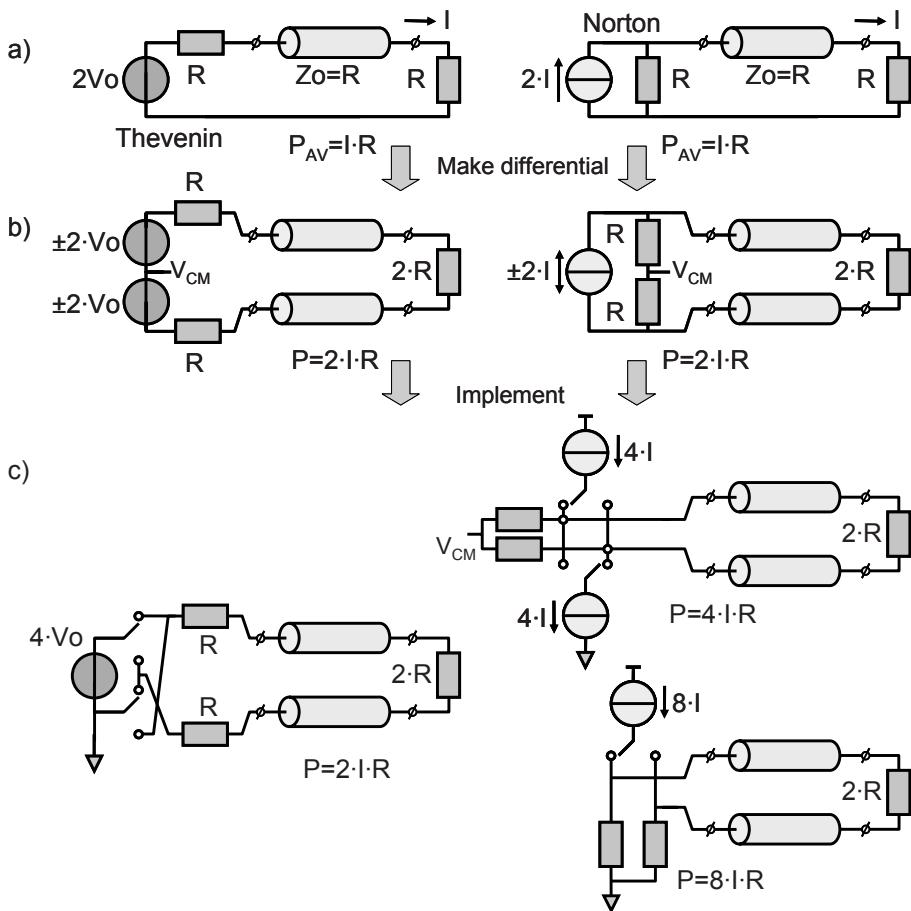


Figure 2: Switched-voltage IO (left) versus switched-current IO (right):
 a) Single-ended Thevenin and Norton equivalent networks, b) Differential equivalent networks, c) Implementation of switched-voltage versus switched-current IO; switched-current both as complementary and single-sided version

Figure 2b (mid) shows the differential Thevenin and Norton equivalent because the final implementation needs to be differential.

Figure 2c shows abstracted implementations. The Thevenin-based implementation is simple; it is a switched-voltage bridge with series resistance for proper termination [4,5]. If the $4V_o=4I\cdot R$ voltage is not directly available, it can be derived from the supply with a regulator (assumed that $4I\cdot R < VDD$, which is practically almost always true).

The implementation based on the Norton equivalent results in the commonly seen differential pair based drivers. For power efficiency reasons, a complementary driver structure with both a sourcing and sinking differential pair in anti-phase is preferable, because in that case the full current is utilized to generate signal swing. This driver topology has been popular for several standards [6,7,8]. However, it mandates a common-mode level somewhere halfway the supply, which increases the required supply voltage for both transmitter and receiver circuits. This makes these structures hard (or even impossible) to implement in advanced CMOS processes due to the limited supply voltage.

Therefore, in other standards the common-mode level has been chosen either high or low by terminating the lines one of the supplies [9,10,11]. This implies for a switch-current implementation that the current can be driven in only one direction: push or pull (source or sink), depending whether the terminations are to VSS or VDD.

Knowing the implementations globally, the power consumption of all solutions can be compared. Assuming that $2 \cdot I \cdot R$ fits into the supply and all circuits are implemented using the same supply, this means that the (Norton-based) complementary switched-current solution consumes 2x more power, because it takes twice the current from the supply compared to the switched-voltage implementation.

In case of a push- or pull-only switch-current driver the efficiency of it is reduced with another factor 2. Taking into account that for proper circuit biasing a switched-current driver implementation easily requires a supply of 1.8V or even higher, while the switched-voltage driver needs a few hundred mV that might even be generated with an efficient DC-DC converter, the switched-voltage solution becomes even more advantageous.

This makes that for mobile applications a differential switched-voltage driver is most attractive. It saves at least a factor 4 of power compared to the differential one-side switched-current drivers, with opportunities to increase this to a factor 10 or more. Furthermore it is much more future proof with respect to process scaling.

The switched-voltage driver is essentially a full-bridge structure switching between two voltages. It is beneficial to use ground for the lower level, because that is always available and shared between IC even if they do not have the same supply voltage [4]. The total series resistance of switch and resistor can be designed to match the characteristic line impedance to obtain a proper source impedance matching.

Figure 3 shows the basic structure of a switched-voltage driver, together with the transmission lines, terminations, and receiver input stage. The transmission lines have a typical characteristic impedance of 50Ω to ground each, with limited mutual coupling. The far-end is differentially terminated, while the common-level is transmitter-defined in order to avoid additional bias current. The center point of the termination has been AC-shorted to ground with a capacitor in order to provide common-mode impedance match at very high frequencies.

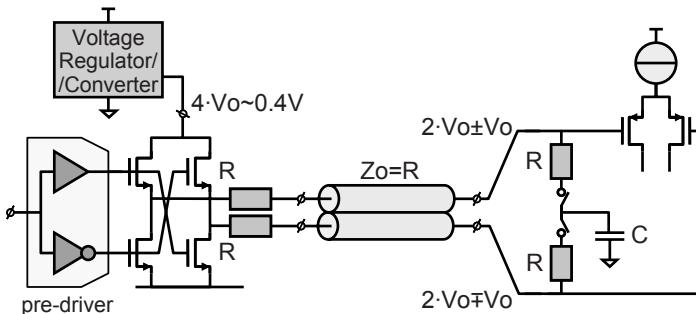


Figure 3: High-speed IO structure

5. Source-Synchronous High-Speed Communication

For high-speed transmission there are basically two possibilities: source-synchronous or embedded-clock. Source-synchronous operation requires an additional clock channel, but allows instant synchronization, provides a reliable clock to the far-end, and provides freedom to choose any clock/data rate. The latter property is especially useful to prevent interference in sensitive parts of the spectrum and to make this solution suitable for many process generations. Embedded-clock is less skew sensitive and saves some IO pins and interconnect wires, but it makes hardware significantly more complicated. Furthermore, the overhead time for link start-up from complete power-down takes longer for embedded clock solutions. Because the IO technology itself is highly power efficient and links are reasonably short, a source-synchronous solution is preferable for many applications.

For source-synchronous solutions there are two commonly used signaling schemes: Data-Clock and Data-Strobe. Although Data-Strobe signaling has a better skew tolerance, it is not easily extendable to multiple data lanes [8]. Therefore Data-Clock signaling has been chosen, which is illustrated in figure 4 for a configuration with two data lanes. The Master side provides the clock, while the Slave side receives it. The half-rate Clock signal ensures similar

characteristics for Clock and Data signals. In this Double-Data-Rate (DDR) transmission scheme both rising and falling edges are used for data slicing. Data and Clock have a quadrature phase relationship such that the data can be straightforwardly sliced with the Clock signal at the receiver side. The remainder of this paper describes the features of a single data lane but any realization may include multiple data lanes.

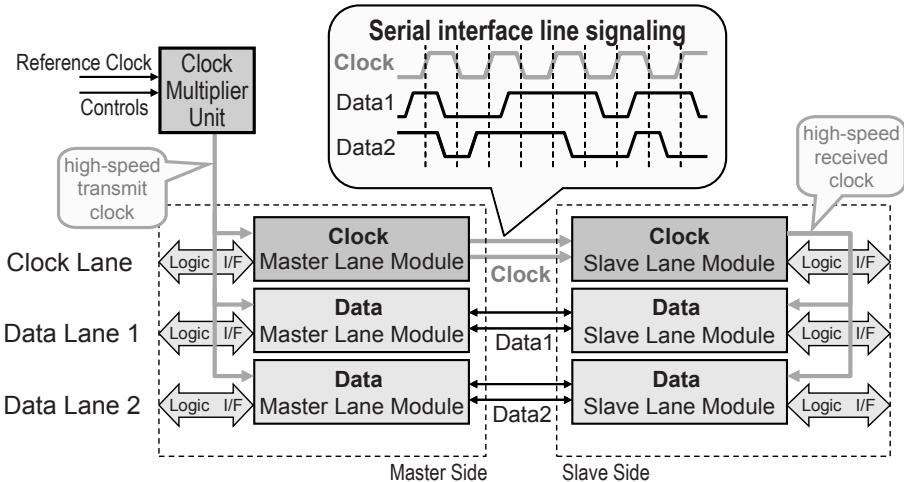


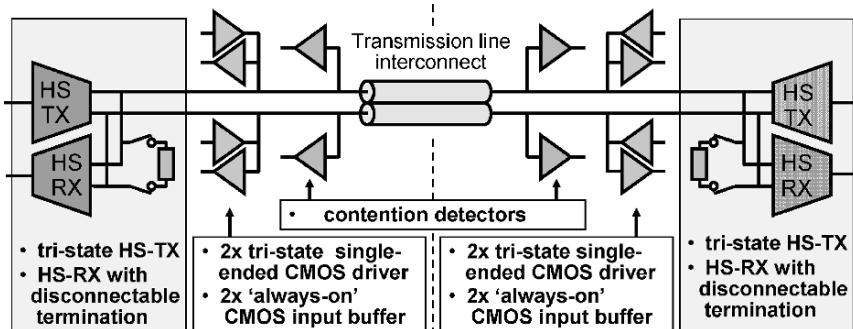
Figure 4: Source synchronous high-speed configuration with two data lanes

6. Hybrid IO: High-Speed plus CMOS: the best of both worlds

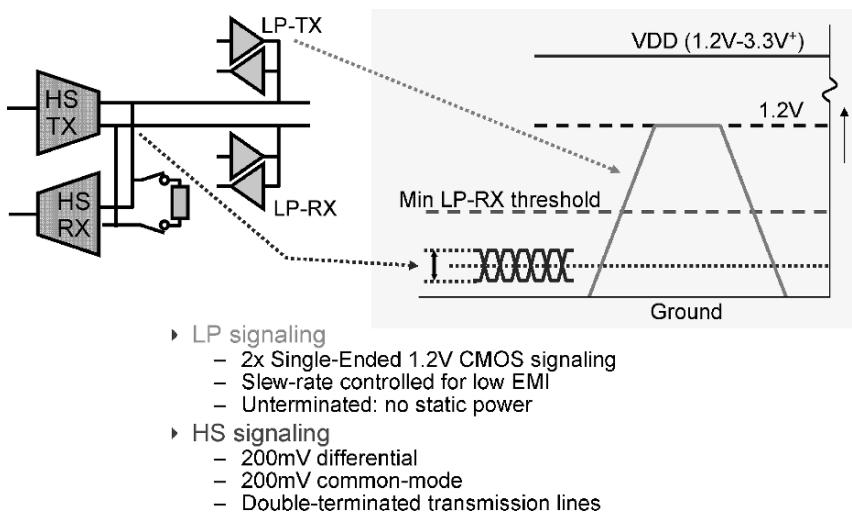
Low-speed features can be more efficiently accomplished with CMOS communication. Therefore each interface also includes CMOS driver and receiver circuits for single-ended 1.2V CMOS signaling via the same wires. These are used when high-speed driver and terminations are disabled, whereas the CMOS line drivers are tri-stated during high-speed communication.

Signals are single-ended and un-terminated in CMOS mode. This provides 4 low-power states for the two lines: LP-11, LP-00, LP-10, and LP-01.

Optionally the link can be turned around and high-speed and/or CMOS communication can be supported in the opposite direction. In that case, driver and receiver functions for high-speed and/or CMOS mode need to be added for the opposite direction. In this bidirectional case, line driving conflicts can occur and therefore detection of bus contention becomes important.



The full-fledged IO configuration is illustrated in figure 5. The CMOS input stages, which are always monitoring the lines, intentionally have a trip-level above the high-speed line signaling levels. Therefore, high-speed signaling is observed by these receivers as LP-00. Figure 6 shows the signal levels for high-speed differential and low-power CMOS signaling.



The low-power CMOS drivers are slew-rate controlled, which minimizes EMI generation. For low EMI sensitivity, the low-power receivers apply filtering and hysteresis to their inputs to suppress high-frequency (RF) noise and glitches. This is illustrated in figure 7.

In order to make all the described IO functions co-operating correctly, an advanced signaling protocol is required, which will be described in the following sections.

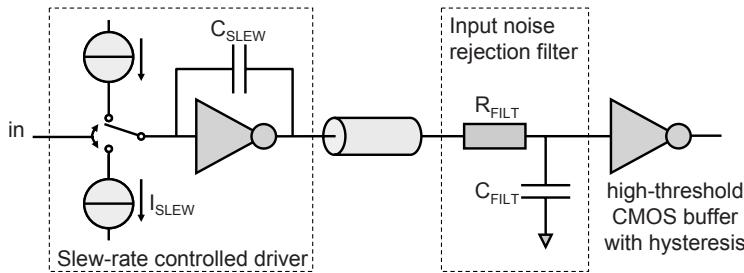


Figure 7: Noise immune LP signaling path with slew-rate controlled driver and receiver input ‘noise’ filtering

7. Operation scheme

During periods when no high-speed transmission is needed, the lines must maintain a well-defined and recognizable state. Preferably the link should have negligible power consumption in these situations.

For this stand-by state, which is called Stop state, both lines are put in the logical-high CMOS state (LP-11) that can be recognized in a receiver with its CMOS input stages without DC power consumption. This LP-11 state is exclusively reserved for Stop state and can therefore be exploited as physical reset for the state machine, which increases robustness.

The 3 remaining CMOS states of the two wires are used for link control and communication purposes. Link control includes mode-switching, link turnaround in case of bidirectional operation, and entering the Escape mode. The Escape mode provides special communication features like low-power data communication. Almost all Low-Power signaling uses Gray-coded CMOS line state sequences, which makes it skew tolerant and allows asynchronous operation.

Figure 8 shows the operation diagram. The Stop state occupies the central position where any action starts from. There are three main actions possible; High-Speed Data Transmission, Escape mode, and Turnaround. The following subsections describe these three actions in detail. Turnaround and some features of Escape mode are optional. These are not needed for all applications and their absence does not hamper baseline functionality.

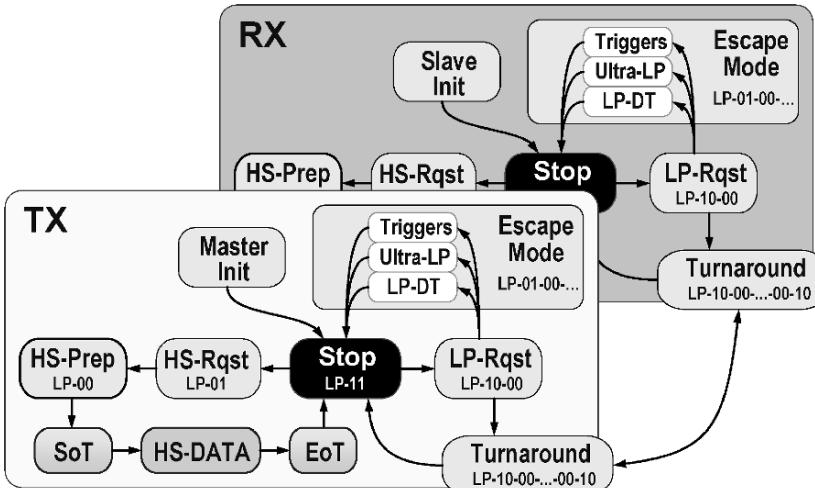


Figure 8: Operational flow diagram

7.1. High-speed data burst

The sequence of events during a high-speed data burst is depicted by the loop in the lower-left corner of the operation diagram in figure 8. The signaling during a high-speed data burst is illustrated by figure 9. Using the Gray-coded low-power state sequence LP-11-01-00, the receiver is notified of the start-up of a high-speed transmission burst. The LP-driver is disabled and the HS circuits are enabled simultaneously during LP-00 state in order to avoid severe current glitches or floating lines. A short period of undefined line state is ignored in the receiver by means of a time-out.

The high-speed signaling starts with a lead period of differential zero state to exceed the time-out, followed by an 8-bit sync word. This word ‘00011101’ is suitable to detect the word boundary and tolerant to any single bit error. After the sync word, raw or encoded payload data can be transmitted. After initial word synchronization, the source-synchronous transmission scheme is assumed to stay synchronized for the rest of a data burst. The transmission burst is ended by swapping the polarity of the signal immediately after the last payload bit and keeping that differential state for a while, before switching to low-power state LP-11. Back-tracking where the last differential transition occurred after the

detection of LP-11 enables the receiver to identify the last valid data even without payload data encoding. In case of line-coding the receiver can be notified of the end of valid data by means of a control word, which eliminates the need for back-tracking.

At the end of a data burst, the link returns to the Stop (stand-by) state by driving the lines simultaneously ‘high’. This exception to Gray-coded low-power signaling is necessary, because the receiver does not know that high-speed communication has ended until LP-11 is detected, which implies that initially the differential termination is still connected. This is not a problem as long as both wires are not driven in opposite directions. Only after the LP-11 detection the termination is disconnected. This is illustrated in figure 9.

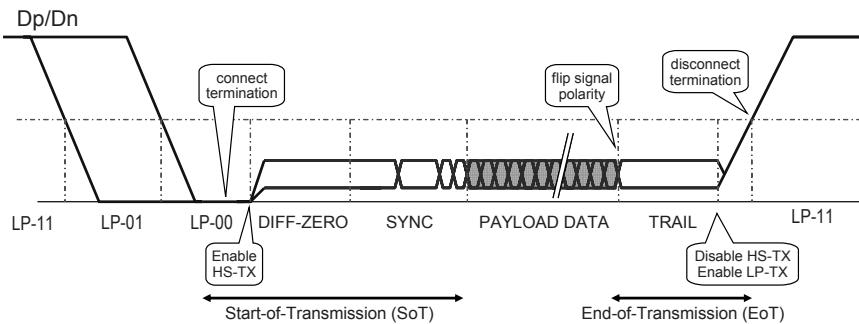


Figure 9: Line signaling during a high-speed data burst

7.2. Escape mode

The Escape mode is requested via the Gray-coded low-power sequence LP-11-10-00-01. This mode contains several features which are selected by an 8-bit entry code after the request. From the set of 256 possible words, only 8 pre-defined entry codes are allowed for maximum robustness against errors. Some entry codes represent triggers that raise a certain flag at the receive side. There are two entry codes which imply a follow-up after the 8 entry bits. These are low-power low-speed CMOS data transmission and the Ultra Low-Power State. These two are described in the following subsections.

During Escape mode, ‘Spaced-One-Hot’ bit encoding is used for transmission. This means that a logic bit value ‘1’ is communicated as an LP-10 state followed by a LP-00 (Space) state, and a bit value ‘0’ consists of LP-01 followed by LP-00. This modified One-Hot encoding where each bit is ‘Spaced’ by an LP-00 state avoids simultaneous switching of the lines and circumvents the inadvertent occurrence of the LP-11 state, which would bring the PHY back to Stop state.

This bit encoding has the property that it is self-clocking, because the clock can be derived from the two signal lines with an EXOR function. Therefore communication in Escape mode can be asynchronous as there is no need for states to the same length. This eliminates the need for an accurate clock during low-power modes which is advantageous to reduce power consumption. Only the minimum duration of states is restricted to avoid interference problems.

7.2.1. Asynchronous Low-Speed Data Communication

The 2-wire asynchronous logical CMOS signaling also enables low-speed low-power data communication. This can be advantageous when the amounts of data to transport are very small, such that starting the high-speed link would be less efficient due to start-stop overhead. After the Escape mode request and the 8-bit Entry code ‘11100001’ to select the data transmission feature, bits are asynchronously transported until the line state returns to LP-11. Due to the asynchronous nature of the communication the transmission can be paused at any time. After low-power data transmission the link always returns to Stop state. An example of a two-byte data message is shown in figure 10.

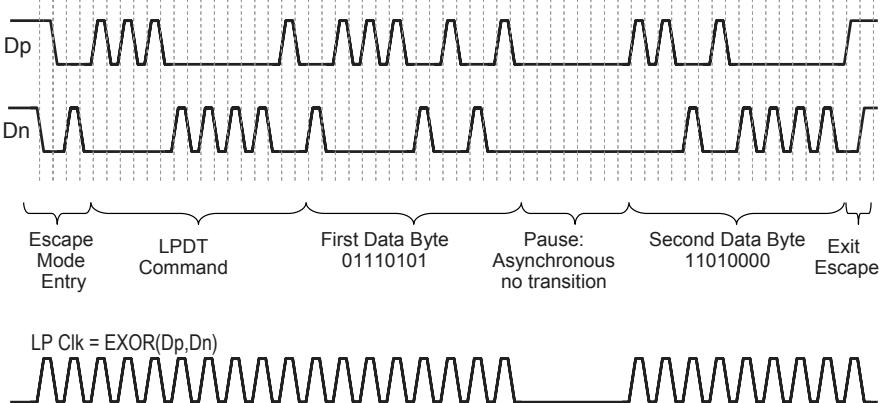


Figure 10: Example of low-power self-timed data communication

7.2.2. Ultra-Low Power State

The CMOS signaling has been chosen at 1.2V independent of IC supply voltage for interoperability reasons. This means that for ICs with higher supply voltages the 1.2V supply may have to be generated with a regulator. This typically implies some static power consumption. Furthermore, other circuits that require bias current might be powered-down if it is known that the link will not be used for a long period of time. For this reason an Ultra-Low Power State (ULPS) is included where the lines are in LP-00 (Space) state. ULPS is obtained via the Escape request and an 8-bit ULPS entry code. In order to ensure proper recovery from ULPS, there is an intermediate LP-10 state before returning to Stop state that has a minimum duration of 1ms to re-establish all biasing and to allow circuits to settle.

7.3. Bidirectional operation

An optional link feature is data lane Turnaround to enable communication in the opposite direction across the same wires. In this reverse mode similar low-speed signaling is available. This is especially attractive for links with highly asymmetrical payload, where this feature may eliminate separate wires for small amounts of return traffic.

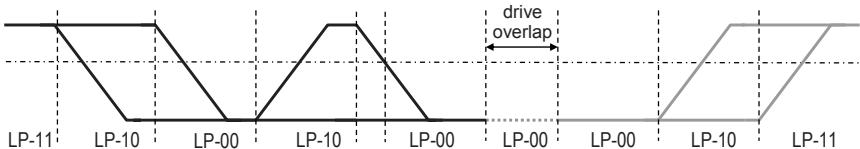


Figure 11: Turnaround signaling

Figure 11 illustrates the event of link turnaround. The initial source side transmits LP-11-10-00-10-00, which is the request for link turnaround to the initial receiver. If the receiver observes this sequence, it takes over control by also driving LP-00. The initial transmitter stops driving the lines after a while and monitors the lines for response from the other side. The initial transmitter has become the new receiver and the initial receiver has become the new transmitter. The new transmit side completes the turnaround by driving LP-10 before returning to Stop state.

If lane turnaround is supported, reverse high-speed transmission is an optional feature that can be used, but only at reduced ($\frac{1}{4}$) rate. The source-synchronous property is lost in this direction, because the clock is only transmitted from master to slave side. Therefore data recovery for reverse traffic is based on oversampling. Operation and signal sequences are similar to forward direction high-speed bursts, but at a quarter rate.

8. Clock lane behavior

The clock lane is in many aspects similar to a Data lane, but there are a few important differences. The clock lane cannot be turned around; this simplifies operation and increases robustness. The Clock lane does not include regular Escape mode with all the before-mentioned features, but only supports ULPS via the simplified state sequence LP-11-10-00. This is possible because turnaround and regular Escape mode are eliminated. High-speed mode on the clock lane means transmission of a clock burst, which is required for the source-synchronous data transfer on the data lane. The clock lane must transfer the high-speed clock signal as long as any data lane is busy with high-speed data transmission. In all other cases it can be put in Stop state or even go to ULPS.

9. Silicon implementation and measurements

This PHY has been implemented in 65nm CMOS technology. Running on a 1.2V supply the power consumption in high-speed mode is about 10mW/lane at 800Mb/s. In low-power mode the power consumption scales linearly with speed and distance and ranges from nearly zero up to a few mW's (for 10Mb/s across a 30cm link). In ULPS power consumption reduces to leakage levels. Figure 12 shows prototyping boards and figure 13 shows the measured line signals for a short high-speed data transmission burst at 800Mb/s.

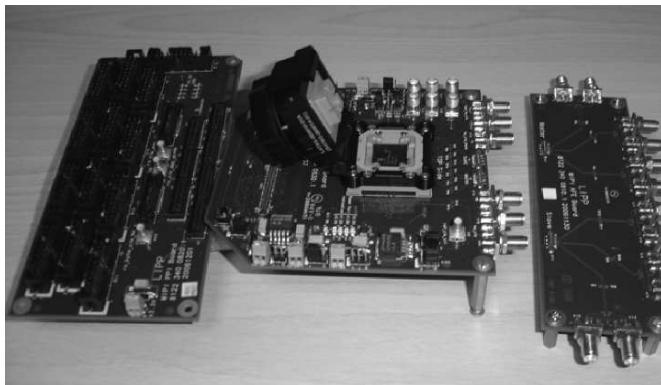


Figure 12: Prototyping boards

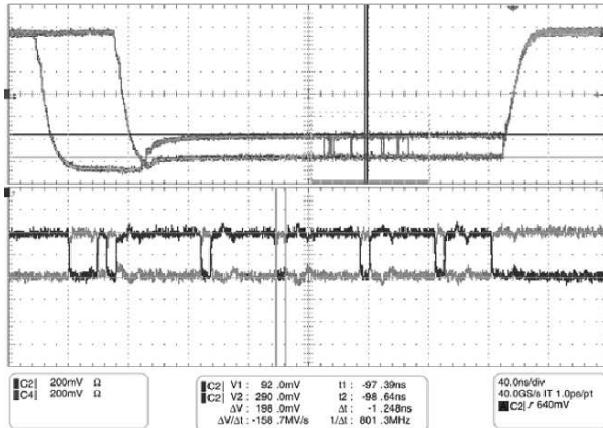


Figure 13: High-speed data burst at 800Mb/s. Top: single-ended measurement of both signaling lines. Bottom: zoom in on sync and 6 bytes of payload data.

10. Concluding Remarks

A power efficient high-speed serial interface solution is presented, which serves the need for increased interface bandwidth between ICs within mobile devices. Merging high-speed differential and single-ended CMOS signaling on the same wires enables opportunities to optimize overall performance for varying bandwidth demands. The presented solution provides multiple operation modes and features, is flexible in speed, and compared to classic CMOS buses it saves pins and reduces EMI. Compared to other high-speed interface industry standards the power efficiency is state-of-art.

Acknowledgements

I would like to thank several NXP colleagues and MIPI standardization working group members for many fruitful discussions, but it is impossible to mention all names here. Special thanks to Harold Perik, Tim Pontius, and Jeannet van Rens for sparring discussions, cooperation in standardization, and coordinating implementation efforts. Thanks to Vijayendra Paga for the measurements.

Notes and References

- [1] Mobile Industry Processor Interface (MIPI) Alliance www.mipi.org
- [2] Static field approximation of Maxwell equations if circuit dimensions are small compared to wavelength. Explained in several EM text books, for example: Field and Waves in Communication Electronics, S. Ramo, et al., John Wiley & Sons, 3rd edition, 1994, ISBN 0-47158551-3.
- [3] Assuming that proper signal settling requires the signal transition time to be $>4x$ the propagation time, and that bit time for a strongly slew-rate limited signal is only 2x the signal transition time, this results in:

$$4 \times \text{Distance[m]} / (\text{VacuumSpeedOfLight[m/s]} / \sqrt{\epsilon_R}) = 1 / (2 \times \text{Speed[bps]})$$
 Reworking for a typical relative dielectric constant $\epsilon_R < 4$ gives:

$$\text{Speed[Gbps]} \times \text{Distance[cm]} \sim 2$$
- [4] G.W. den Besten, "Embedded Low-Cost 1.2Gb/s Inter-IC Serial Data Link in 0.35um CMOS technology," *Proceedings of IEEE International Solid-State Circuits Conference*, pp 251-252, February 2000.
- [5] SLVS standard, JEDEC, JESD8-13, October 2001 (www.jedec.org).
- [6] Low-Voltage Differential Signaling (LVDS), IEEE standard P1596.3-1996 (www.ieee.org)
- [7] sub-LVDS, industry defined reduced voltage-level derivative(s) of LVDS, especially popular for mobile camera and display applications.
- [8] Firewire, IEEE standards 1394-1995, 1394a-2000.
- [9] USB 2.0 standard, April 27, 2000, www.usb.org
- [10] PCI Express, PCI-SIG, www.pcisig.com/specifications/pcieexpress
- [11] High-Definition Multimedia Interface (HDMI), www.hDMI.org

High Voltage xDSL Line Drivers in Nanometer Technologies

Bert Serneels, Michiel Steyaert, Wim Dehaene
KU.Leuven, ESAT-MICAS
Kasteelpark Arenberg 10
3000 Leuven, Belgium

Abstract

New and improved xDSL standards are developed to bridge the last mile between the Central Office (CO) and the end user. The aDSL2+ standard doubles the bandwidth of a basic aDSL system to 2.2MHz and hence increases the bitrate up to 24Mb/s. The high Crest Factor (CF) of Discrete Multi Tone (DMT) modulated signals, however, poses serious problems on an efficient and low cost implementation of the line driver in the CO [1]. With the emerging nanometer technologies the line driver remains more than ever the bottleneck for lowering the cost and power. The low supply voltages originating from nanometer technologies increase the current density in the line driver, which affects its efficiency and reliability. This paper presents a high voltage high efficient aDSL2+ CO line driver in a mainstream nanometer CMOS technology.

1. Introduction

Over the last years, there is an increasing interest in high voltage design techniques due to the rise of deeper-submicron and nanometer technologies. These technologies provide an answer to the increasing integration density of VLSI circuits and the low-power requirements of complex signal processing applications. However, a major drawback of the gate length scaling is the low nominal supply voltage of the devices. The supply voltage has to scale with the gate length for keeping the electric field across the channel of the devices within limits to ensure reliable operation. As a consequence, the design of analog circuits with high output power requirements is running into its limits. A relative unexplored area in the field of high voltage design is the use of standard CMOS devices due to reliability issues like breakdown mechanisms and hot carrier

effects. However, high voltage design techniques in mainstream CMOS technologies are gaining a lot of interest nowadays for their cost advantage and integration prospects. The idea is to find the correct operating point such that the voltage across the terminals of the transistors is limited during operation.

The techniques of high voltage design in standard CMOS are applied on the Self-Oscillating Power Amplifier (SOPA) architecture, a high efficiency line driver for xDSL. A SOPA line driver can be quite successful in a deep-sub-micron technology, since it can drive DMT signals with a high efficiency [2]. But, as any power amplifier, its efficiency and reliability drops with decreasing supply voltages. An aDSL2+ system requires an average of 20dBm being delivered to a 100Ω twisted pair telephone wire. Lowering the supply voltage results thus in an increased current density for a constant output power. This triggers the generation of hot carriers and electro-migration mechanisms which negatively affects the reliability of the driver. Moreover, the driver has to put signals with a high voltage swing, which are caused by the high CF, on the line. Since the output voltage swing of the line driver is limited by its supply voltage, a transformer with a high transformer ratio has to be used. This has two drawbacks. First, the noise, generated by the driver, is also up-transformed, limiting the noise specifications. The second, and most important drawback, is the attenuation of the received signal. This signal passes the same transformer but in the other direction causing a reduction of the signal with a factor equal to the transformer ratio. This signal has to be detected in a very noisy environment and, not to forget, the large transmit signal, which places extremely hard noise specifications for the receiver circuit.

In the next section, the basic principles of the SOPA line driver are discussed. In section 3, the influence of the low supply voltage on the SOPA design is discussed. Section 4 describes the principle of stacking standard CMOS devices as a technique for high voltage design, which is followed by a design example of a high voltage SOPA line driver in section 5.

2. SOPA Architecture

In this section, the basics of a SOPA line driver are explained. A thoroughly description can be found in [3].

Figure 1 depicts a differential SOPA structure. A basic architecture consists of three building blocks: a comparator, a loop filter and a digital buffer. The comparator is not clocked, so the loop is asynchronous. The loop filter is constructed in a way such that the loop is unstable. Therefore, a limit cycle oscillation exists in the loop, resulting in a square wave output with a determined frequency. When this unstable system is forced by an external signal V_{in} , with a frequency lower than this self-oscillating frequency, the limit cycle acts as a dither and linearizes the system as long as the error signal e at the

inputs of the comparator is smaller than the limit cycle amplitude at the comparators input. Since this amplifier is a switching type amplifier, a high efficiency can be obtained, even when buffering signals with a high CF.

Two SOPA's are coupled by means of a signal transformer. In this way, the two oscillation frequencies are attracted towards each other forcing the limit cycles to oscillate in phase. For the primary of the transformer the limit cycle thus becomes common mode, which is decoupled towards the load by means of the transformer itself. So the input signal is transferred to the low-ohmic load, while the mean switching frequency sees a high ohmic impedance. Since the limit cycle frequency is not transferred towards the load, there is no need for steep filtering to cancel the mean switching frequency.

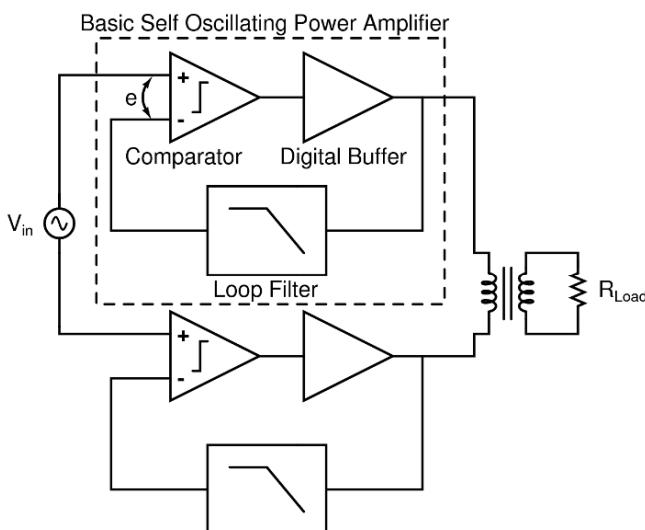


Fig.1: Block schematic of a differential SOPA.

3. Influence of CMOS scaling on SOPA design

Figure 2 shows a schematic representation of a single ended SOPA. R_L , y and R respectively represent the load resistance, the transformer ratio and the resistance seen by the SOPA amplifier after transformation by the impedance matching network, which is, in this case, the line transformer. I_{rms} is the rms current that the line driver has to deliver in order to provide an average output power of 100mW. R_{on} equals the on-resistance of the output buffer and nV_{DD} is the supply voltage with V_{DD} the nominal supply voltage of the used technology.

n is the voltage multiplication factor. It represents a scaling factor for the nominal supply voltage to determine the dependence of the supply voltage on the SOPA output stage parameters.

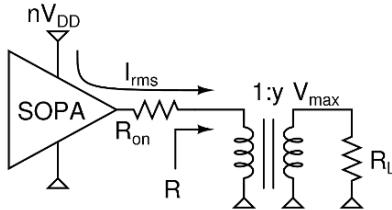


Fig.2: Schematic of a single ended SOPA architecture

The following set of equations determines the transformer ratio, the load resistance after transformation and the rms current.

$$y \approx \frac{V_{\max}}{nV_{DD}} \quad (1)$$

$$R = \frac{R_L}{y^2} = \frac{100\Omega}{y^2} \quad (2)$$

$$I_{rms} = \sqrt{\frac{P_{out}}{R}} = \sqrt{\frac{100mW}{R}} \quad (3)$$

With $V_{\max}=26.6V$, the maximum voltage that the SOPA has to deal with. For the calculation of the transformer ratio, the voltage drop over R_{on} is ignored. These equations result in the following relationship:

$$nZ \Rightarrow y \Rightarrow RZ \Rightarrow I_{rms} \quad (4)$$

Increasing the supply voltage results thus in an increased load resistance seen by the SOPA amplifier and hence a decreased rms current. Figure 3 shows a graphical representation of the relationships stated by equation (4). The nominal supply voltage is set at 1.2V. The x-scale is divided into multiples of V_{DD} to clearly see the dependency on the voltage multiplication factor n . Figure 3d shows the static efficiency of the class D type output buffer in function of its supply voltage. The static efficiency results in a simple resistive division between the on-resistance of the output buffer and the transformed load resistance:

$$\eta_{stat} = \frac{R}{R + R_{on}} \quad (5)$$

Equations (4) and (5) show that an increased supply voltage results in a higher static efficiency for the same on-resistance of the output buffer. One could also lower the on-resistance to improve the static efficiency. However, lowering the on-resistance will result in increasing switching losses which will degrade the overall efficiency.

From figure 3 one can conclude that SOPA architectures with a supply voltage lower than 3.6V becomes impractical and highly inefficient. This is argued by the following statements:

- Supply voltages lower than 3.6V require a transformer ratio larger than 7.4. The output signal is thus up-transformed with a factor of 7.4, but the received signal is attenuated with the same factor. This puts severe noise requirements on the receiver circuit.
- For supply voltages lower than 3.6V, a resistance lower than 1.8Ω has to be driven by the line driver. This results in an rms current of more than 236mV to obtain the average output power of 100mW. High current densities lead to long term reliability issues like hot carrier degradation and electro-migration.
- Due to the low resistance that the amplifier has to drive, the static efficiency and as a consequence, the total efficiency will be lower than 64% for an on-resistance of 1Ω of the output buffer.

4. Stacking devices: The high voltage CMOS solution

4.1 Introduction

Most high voltage devices use modifications of standard technologies to handle high supply voltages. They can be integrated in a mainstream CMOS technology at a higher cost, since extra mask sets and process steps are required. However, there exists an alternative low-cost solution by using only standard CMOS devices. The principle is shown in figure 4. The maximum voltage across the terminals of a transistor is limited by its nominal supply voltage such that the expected lifetime of the device falls within the foundry targets. If two or more transistors are stacked, which means that the source of one transistor is connected with the drain of the next one, the drain-source voltages add up to a multiple of the nominal supply voltage. In figure 4b an example of two stacked devices is given. The drain-source voltage of a single device is limited by V_{DD} , but the resulting drain-source voltage, from the drain of the upper most device to the source of the lower most device can rise up to two times the nominal supply voltage without affecting the transistors reliability. Needless to say that the gate biasing of these devices becomes extremely important. After all, during operation, each transistor has to be biased in such way that the voltage across its terminals remains within the nominal supply voltage.

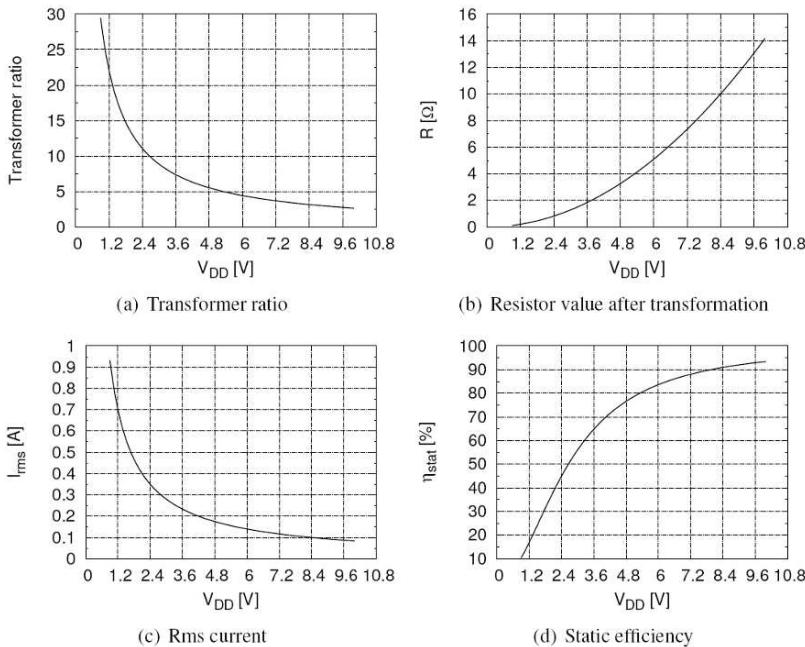


Fig.3: Dependencies of n on the SOPA's output stage parameters

Since the SOPA is a switching type amplifier, another point that has to be carefully investigated is the presence of transient voltage peaks, which can easily go beyond the nominal supply voltage during the switching of these devices.

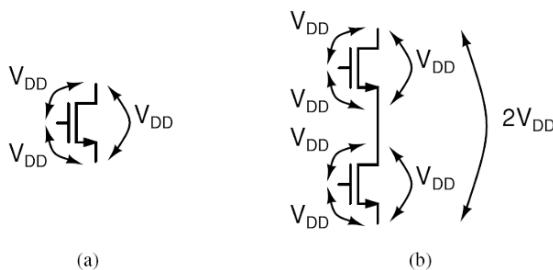


Fig.4: Principle of stacking standard CMOS devices

The principle of stacking devices is applied on the SOPA output stage. Figure 5 shows a principle schematic of the SOPA amplifier. The Pulse Width Modulation (PWM) output signal of the SOPA is applied to large switches which deliver the current to the load. These switches can either be implemented as n stacked transistors comprised from a modern CMOS technology with a nominal supply voltage of V_{DD} , as single CMOS transistors from a previous generation or as single specialized high voltage devices. The different implementations are schematically shown in figure 5. The high supply voltage is defined as n times V_{DD} , the nominal supply voltage of the modern CMOS technology. The upper switch, which pushes the output to the high supply voltage, is implemented with pMOS devices. The lower switch on the other hand, which pulls the output to the ground, is implemented with nMOS devices. If the transistor or stacked transistors are in the linear region, thus representing a small resistance, the switch is on. If they are in the cut-off region, the switch is off.

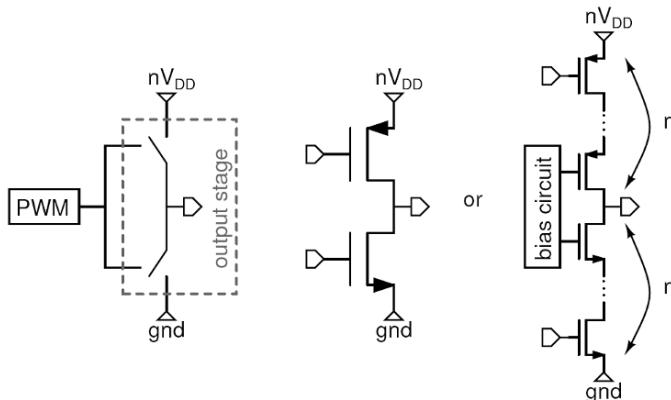


Fig.5: Principle schematic of the SOPA amplifier (left figure) with two implementations of the high voltage output stage (middle and right figure)

In the next section the power dissipation of a high voltage switching output stage comprised of standard stacked CMOS devices will be calculated. The technology used for this calculation is a mainstream 1.2V 130nm CMOS technology. A comparison is made with an implementation in a 2.5V 130nm thick oxide CMOS technology. The parameters of these two technologies are shown in table 1. The thick oxide transistor resembles a $0.25\mu m$ technology, two generations behind the 130nm technology. The next section thus will point out if scaling is beneficial for the output stage concerning the power dissipation. In the remaining of the section, the 130nm thick oxide technology will be shortly

called the thick oxide technology to avoid confusion with the mainstream 130nm technology.

Table 1: Technology parameters

	130nm	130nm thick oxide
V _{DD}	1.2V	2.5V
L _{min}	130nm	0.28μm
t _{ox}	2.3nm	5.0nm
C _{ox}	1.5e-2F/m ²	6.8e-3F/m ²
nMOS: R _{on} @ W=1000μm, L=L _{min}	0.5Ω	1.0Ω
C _{ox} @ L _{min}	1.9nF/m	1.9nF/m

4.2 Power Dissipation

The Power dissipation can be split into two terms: the static power dissipation and the dynamic power dissipation. The static power dissipation covers the losses due to biasing currents, voltage drops over parasitic resistances, etc. The dynamic power dissipation covers the losses due to the charging and discharging of capacitances.

Static Power Dissipation

The static power dissipation of a switching power output stage is equal to the power dissipated in the on-resistance of the switch. It is also called the conduction losses:

$$P_{cond} = I_{rms}^2 R_{on} \quad (6)$$

Following table 1, it is known that that the on-resistance scales with a factor 2 between a 0.25μm and a 130nm CMOS technology for minimum length transistors with the same width. Figure 6 shows a simulation of the on-resistance for a 130nm and a thick oxide n- and pMOS minimum length transistor in the linear region in function of the width. The factor 2 can be clearly observed. This is a very important result, since it means that two stacked minimum length 130nm transistors in the linear region have the same total on-resistance as one minimum length thick oxide transistor in the linear region with the same width. From figure 4 it is known that the supply voltage of two stacked transistors can be doubled. This means that two stacked 130nm transistors can handle the same supply voltage as one thick oxide transistor. According to equations (1), (2) and (3), the two structures thus have to deliver the same rms current to the load. As a consequence, two stacked minimum length 130nm transistors have equal

conductions losses as one minimum length thick oxide transistor with the same width. Or more general, n stacked 130nm minimum length transistors with a width W have equal conduction losses as $n/2$ minimum length stacked thick oxide transistors with the same width W .

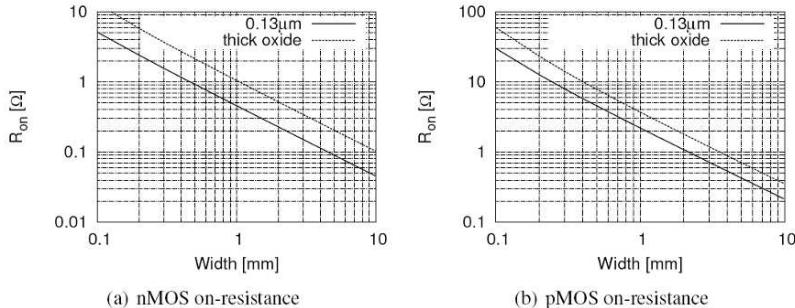


Fig.6: Simulation of the on-resistance

From figure 6 it is clear that the on-resistance drops with increasing width. As a consequence, the conduction losses could be made very low just by increasing the transistors widths. This is also shown in figure 7, where the conduction losses are simulated for $n = 1..5$ stacked nMOS 130nm and thick oxide transistors. One can notice the good resemblance between the curves $n = 2$ and $n = 4$ of the 130nm technology conduction losses and the curves $n = 1$ and $n = 2$ of the thick oxide technology conduction losses respectively. However, increasing the transistors dimensions for lowering the on-resistance will increase the dynamic power dissipation as will be seen in the next section.

Figure 7 also shows that stacking more devices, thus increasing the supply voltage will lower the conduction losses. This means an improvement in static efficiency as was already demonstrated in figure 3d.

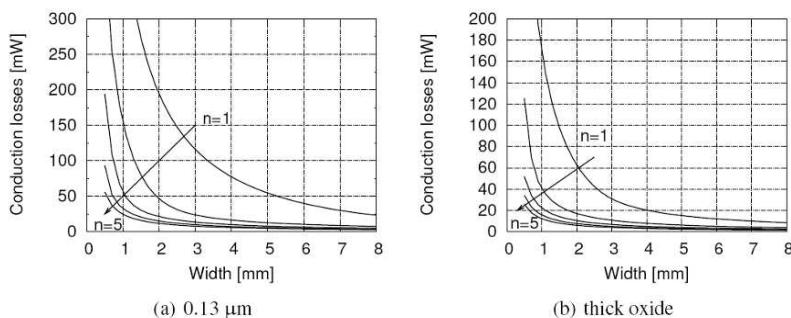


Fig.7: Simulation of the conduction losses

Dynamic Power Dissipation

The calculation of the dynamic power dissipation of a high voltage output stage comprised of stacked transistors covers the charging and discharging of parasitic capacitances of the transistors and the power dissipated in the tapered buffers driving the output stage. The global expression for the dynamic power dissipation equals: CV^2f , with C the capacitance value, V the voltage range over which the capacitor gets (dis)charged (V is also the voltage from which the capacitor gets (dis)charged) and f the switching frequency. The switching frequency in the SOPA system equals the limit cycle frequency f_{lc} . For the simulation of the switching losses, a limit cycle frequency of 40 MHz has been chosen.

Since this kind of power dissipation is caused by switching transistors it is also called by the term switching losses. Figure 8 shows n stacked nMOS transistors and their parasitic capacitances that are taken into account for the calculation of the switching losses. The expression for the switching losses is split into three terms:

- The first term is the power dissipated in the tapered buffers which drives transistor M1. For a tapered buffer, optimized for speed, this term equals:

$$P_{buffer} = \frac{C_{inl}V_{DD}^2f_{lc}}{e-1} \quad (7)$$

- The second term covers the losses due to the charging and discharging the input capacitances of the stacked transistors. Since the voltage across the terminals of the stacked transistor does not exceed V_{DD} due to the dedicated bias circuit, the maximum power loss by switching the inputs of the stacked transistors equals:

$$P_{C_{in}} = V_{DD}^2f_{lc} \sum_{i=2}^n C_{in_i} \quad (8)$$

The power loss by switching transistor M1 is already taken into account in the expression for the power dissipation in the tapered buffer.

- The third term covers the losses due to the charging and discharging the well capacitances of the stacked transistors. These capacitances originate from the fact that each stacked nMOS transistor has their own separate p-well. This p-well is lying in an n-well, to isolate the device from the substrate. As such, a large junction capacitance is present between the bulk of the transistor and the substrate. This capacitor is called the well-capacitor C_{well} . Without this triple well structure a DC-path would exist from the bulk-source connection of every stacked transistor to the ground resulting in substrate currents which can lead to latch-up. The power loss due to the charging and discharging the well capacitances equals:

$$P_{well} = f_{lc} \sum_{i=1}^n C_{well,i} [(i-1)V_{DD}]^2 \quad (9)$$

Since the drain-source connections of the stacked transistors can be charged-up to a multiple of V_{DD} , this term will have a large contribution in the total power loss and will eventually be the limiting factor for stacking more transistors. The well capacitance of a transistor is defined as:

$$C_{well} = \alpha c_j WL \quad (10)$$

With c_j the junction capacitance between the bulk of the nMOS transistor and the n-well. α is a layout dependent correction factor such that the value of the well-capacitance can be calculated from the width and the length of a transistor.

The total switching losses thus equals:

$$P_{switching} = P_{buffer} + P_{C_{in}} + P_{well} \quad (11)$$

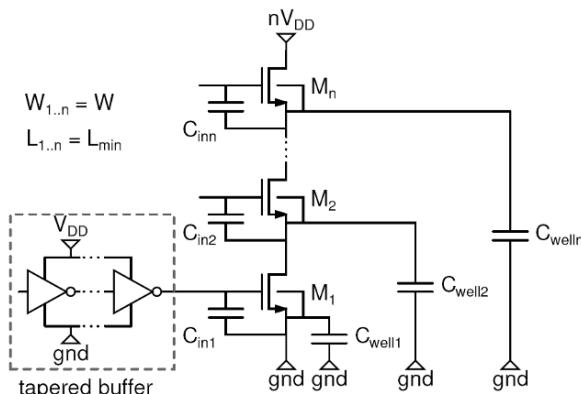


Fig.8: Schematic of n stacked nMOS transistors with the elements that lead to the calculation of the switching losses

Figure 9 shows the total switching losses in function of the width of the transistors for the 130nm and the thick oxide technology for different numbers of n nMOS stacked transistors. From this figure some logic conclusions can be drawn. Increasing the width of the transistors results in larger parasitic capacitances. Following the equations (7), (8) and (9) this leads to increasing switching losses. Secondly, stacking more devices results in more parasitic capacitive elements, leading again to higher switching losses. Moreover, the well-capacitances at the drain-source connections of the stacked transistors will be (dis)charged to higher voltages, resulting in a strong rise of the switching

losses. These two conclusions are exactly the opposite as those of the static power dissipation. Therefore, an optimum can be found as will be seen in the next section.

Figure 9 also shows that a configuration of n stacked minimum length nMOS 130nm transistors have comparable switching losses as $n/2$ stacked minimum length nMOS thick oxide transistors. The switching losses are slightly higher for the configuration with 130nm devices. This is due to the large junction capacitance c_j between the bulk of the nMOS devices and the n-well which isolates the nMOS devices from the p-substrate. Therefore, the P_{Cwell} losses have the largest contribution in the total switching losses.

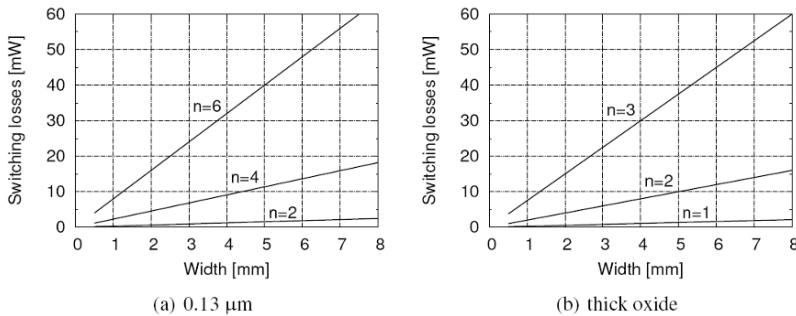


Fig.9: Calculation of the switching losses

Total Power Dissipation

Figure 10 shows the combination of the conduction and the switching losses leading to the total power losses for different numbers of n stacked nMOS transistors in the 130nm and the thick oxide technology. Careful observation of these figures leads to the following remarks:

- As the conduction losses drop and the switching losses rise with increasing transistors width, a minimum in the total power losses can be found, which is clearly visible.
- There is a good resemblance between the losses of n stacked 130nm nMOS transistors and $n/2$ stacked thick oxide transistors.
- The minimum of the total power losses increases for increasing the number of stacked transistors. This is due to the strong increase in switching losses for higher numbers of n in spite of the lower conduction losses.
- The minimum of the total power losses is reached for lower values of the width of the stacked transistors for higher numbers of n . Stacking more transistors results in more parasitic capacitances

which leads to a higher and steeper curve of the switching losses. As a consequence the point where the switching losses become higher than the conduction losses is reached for lower transistor widths. On the other hand, this is a good result in terms of area, since it means that the area will not increase exponentially by stacking more transistors.

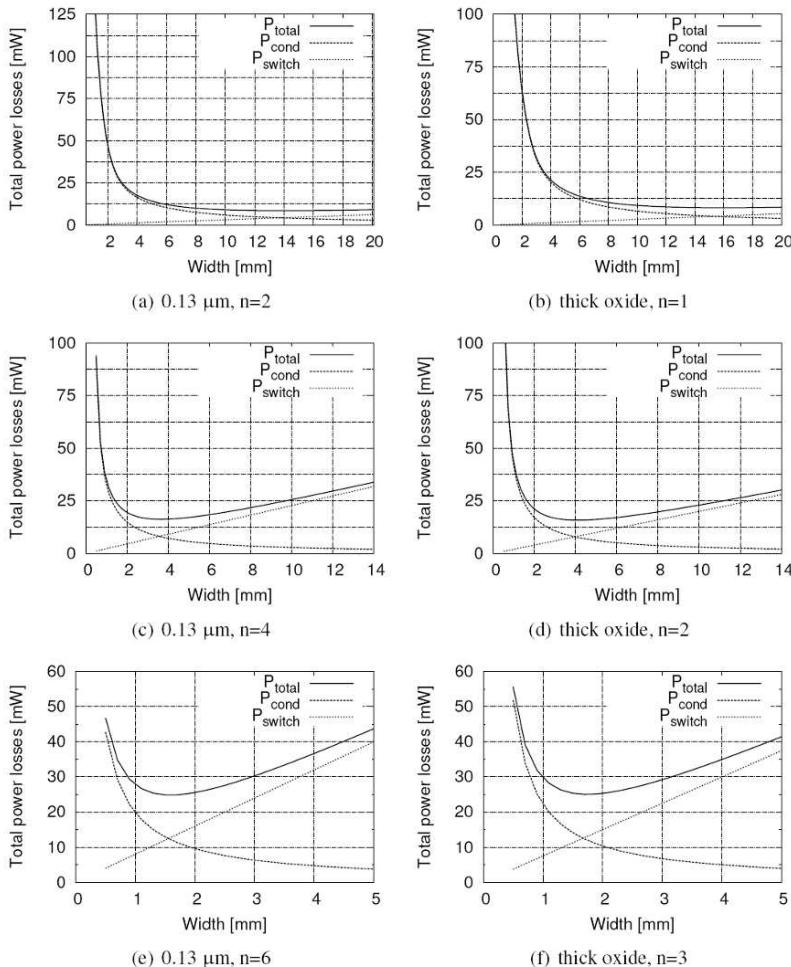


Fig.10: Simulation of the total power losses for different values of n

Until now the power losses were discussed for nMOS stacked transistors. The same results are obtained for pMOS stacked transistors. The output stage of

the SOPA amplifier is comprised of nMOS and pMOS stacked transistors to implement the switches, which was shown in figure 5. Figure 11 shows the minimum total power losses in function of the number of nMOS and pMOS stacked transistors for an implementation in a 130nm technology. In fact, the figure shows the total power loss of the SOPA output stage, without inclusion of the dedicated gate bias circuit of the stacked transistors. To make a comparison with an implementation in the thick oxide technology, the total power losses of the output stage comprised of $n/2$ stacked thick oxide transistors is plotted (squares) on the n scale instead of $n/2$. For example: An implementation of eight stacked 130nm devices has a total power loss of 60mW, whereas an implementation of four stacked thick oxide devices has a total power loss of 61mW.

Figure 11 shows that an implementation of the output stage with n stacked 130nm devices is slightly more efficient than an implementation with $n/2$ stacked thick oxide transistors. The real advantage lies in the fact that the low voltage issues due to scaling the technology to ultra deep-sub-micron or nanometer technologies can efficiently be circumvented by stacking devices. It is thus possible to design a high voltage output stage in a low voltage technology, resulting in a compact low-cost single chip solution, without degradation compared to an implementation in a technology from a previous generation which has a higher nominal supply voltage.

A final remark that can be made with this figure and which was already clear from figure 10 is that the minimum total power loss increases when more transistors are stacked. One can conclude then to use only two stacked 130nm devices for this structure has the lowest power losses. However, one must take in mind that only two stacked devices results in a rather low supply voltage. As a consequence, previously discussed problems like large transformer ratios, high distortion, high rms currents, reliability, etc. will arise. For the design of the SOPA line driver one must thus carefully choose the number of stacked transistors to alleviate these problems and not to over-design the output stage.

5. Design example

The stacking of devices is a very promising technique for designing high voltage circuits in a low voltage standard CMOS technology. However, to prove its credibility, the technique needs to be verified by practical implementations. Therefore the stacking technique is used to design a high voltage output stage for the SOPA line driver to relax the transformer ratio. The technology used for this design was a 1.2V 130nm standard CMOS process. In the next section the line driver architecture will be discussed. The focus is set on the high voltage output stage.

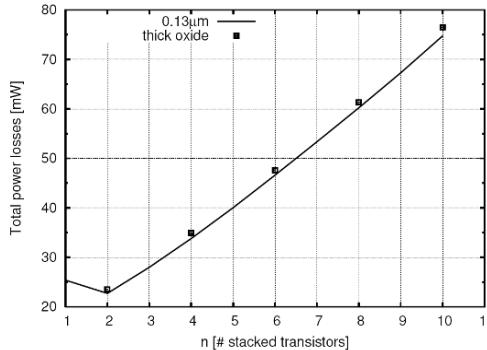


Fig.11: Minimum total power loss of the SOPA output stage in function of the number of stacked transistors

5.1 Line Driver Architecture

Figure 12 shows the block diagram of the proposed high voltage line driver. The single SOPA is constructed with a RC-integrator in the forward path followed by a non-clocked comparator. The SOPA is thus of first order. By adding an integrator in the forward path, the modulation shifts from asynchronous delta modulation to asynchronous delta-sigma modulation. The advantage of the inclusion of the integrator is the higher linearity of the switching system for the same over-switching ratio (= bandwidth/mean switching frequency).

The high voltage buffer converts the output of the comparator to high voltage levels. The output of this buffer is fed back to the integrator using a loop filter. Since the integrator and comparator operate at the low, nominal supply voltage of the used technology, the output of this buffer is needs to be down converted within the voltage limits of the technology. This is implemented in two stages. The first stage is a voltage division set by the resistors R3 and R4 in the loop filter. The second is a current division set by the resistors R1 and R2 at the input of the comparator. The output voltage is thus converted to an output current, which is subtracted from the input current and the resulting current is integrated over the capacitor. This resembles the advantage of an RC-integrator used in a low voltage technology, namely, its high linearity for large input voltage signals since the input node can be considered as a virtual ground.

The frequency of the self-oscillation is set by the loop filter in combination with the filter in the forward path. This self-oscillation provides thus a dithering effect for the input signal and hence linearizes the comparator and the high voltage output buffer for frequencies lower than the oscillation frequency.

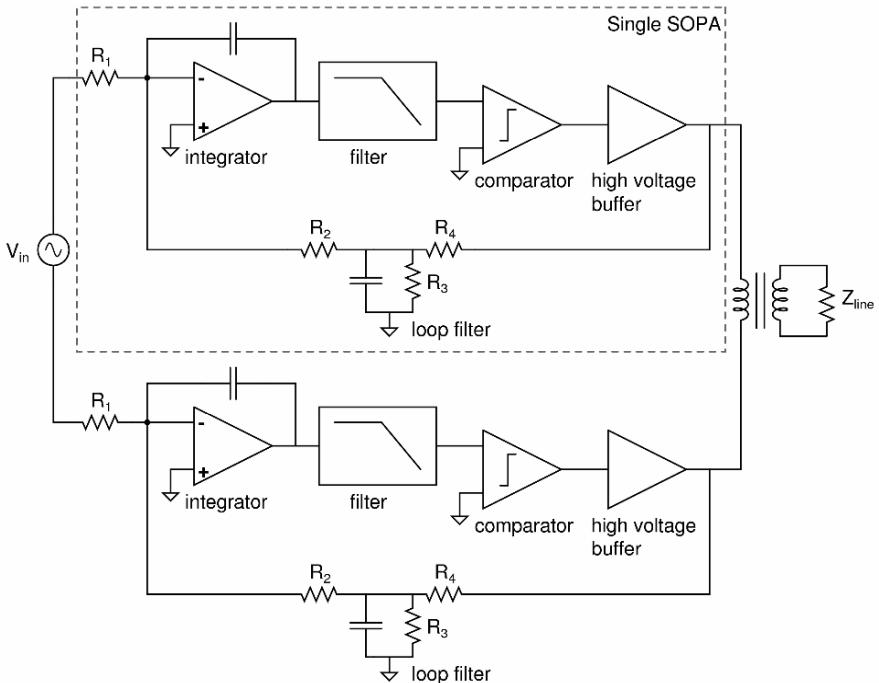


Fig.12: Block diagram of the proposed line driver

5.2 High Voltage Output Stage

Figure 13 depicts the block schematic of the high voltage buffer. The output stage is composed of five stacked transistors for the pull-up switch and the pull-down switch respectively. The theoretical maximum supply voltage of the buffer is thus five times V_{DD} . A novel dedicated bias circuit was designed to keep the voltage across the terminals of the stacked transistors within the technology limit, the nominal supply voltage V_{DD} . The output of the buffer is controlled by the outer stacked transistors, which are driven by a tapered buffer. Distortion of the output waveform is minimized by providing matched delay paths from the input of the driver to the outer stacked transistors. This approach uses a symmetrical supply with two level shift circuits for setting the offset voltages of the pMOS- and nMOS buffer. The level shifters are preceded by a Non-Overlapping Switching (NOS) circuit, to minimize the power dissipation due to short circuit currents. The combination of the NOS, the tapered buffers and the

level shifters form the pre-driver circuit of the outer stacked transistors. In the next section, the design of each building block will be discussed.

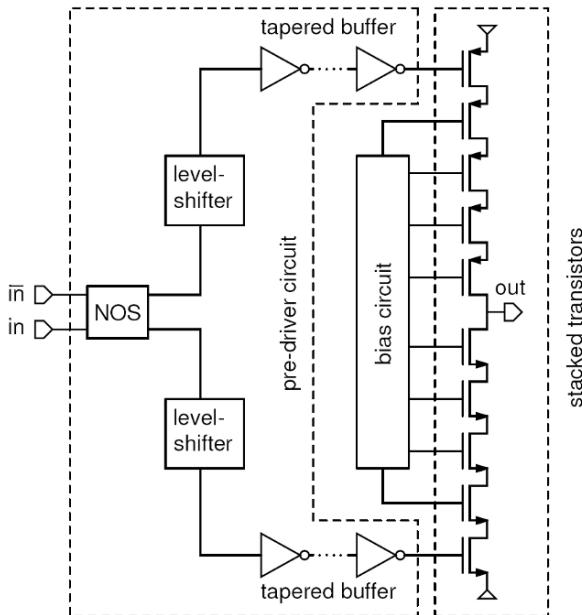


Fig.13: Block schematic of the high voltage output buffer

Bias Circuit

Figure 14 shows the schematic of the stacked transistors with the bias circuit. The working principle of this output stage is based on the techniques used in [4], where the gates of the stacked transistors are set by a resistive ladder network. The capacitors parallel with the resistors counteract the overshoot during transients on the gates of the transistors due to their parasitic gate-drain capacitance. In the presented high voltage buffer, transistors MB1 to MB6 and MBB1 to MBB6 perform the function of the resistive ladder network.

All transistors are used as switches. This means that in the on-state they are in the triode region resulting in a low impedance. In the off-state, they are in the cut-off region, resulting in a high impedance. Assuming that V_{DD} is the nominal supply voltage of the technology used, all voltages on the internal nodes of the buffer are written in function of V_{DD} , as shown in figure 14. For reasons of clarity, the supply voltage is defined between 0 and $5V_{DD}$. Consider the transition of the output from high ($5V_{DD}$) to low (0). The signals arriving from the tapered buffers switch M10 off and M1 on. Node n1 gets discharged and M2 is switched on, which in turn discharges node n2. Now, M3 and MB1

are switched on. This results in a discharging of respectively node n₃ and node g₃. The same reasoning accounts for M4 – MB2 and M5 – MB3 discharging the nodes n₄–g₄ and out–g₅. Also, with the discharging of node g₄, MBB1 is switched off and MBB2 is switched on. The discharging of node g₅ has three effects. First, MBB4 is switched off. Secondly, M6 is switched off when node n₅ is discharged till V_{DD}, which in turn switch MB4 off. Finally, MBB3 is switched on. MBB3 discharges together with MBB2 node g₆. This switches MBB6 on and MBB5 off. M7 is switched off when node n₆ is discharged till 2V_{DD}. MBB6 discharges node g₇, which switches M8 off when node n₇ is discharged till 3V_{DD}. The same reasoning can be made for the transition of the output from low (0) to high (5V_{DD}).

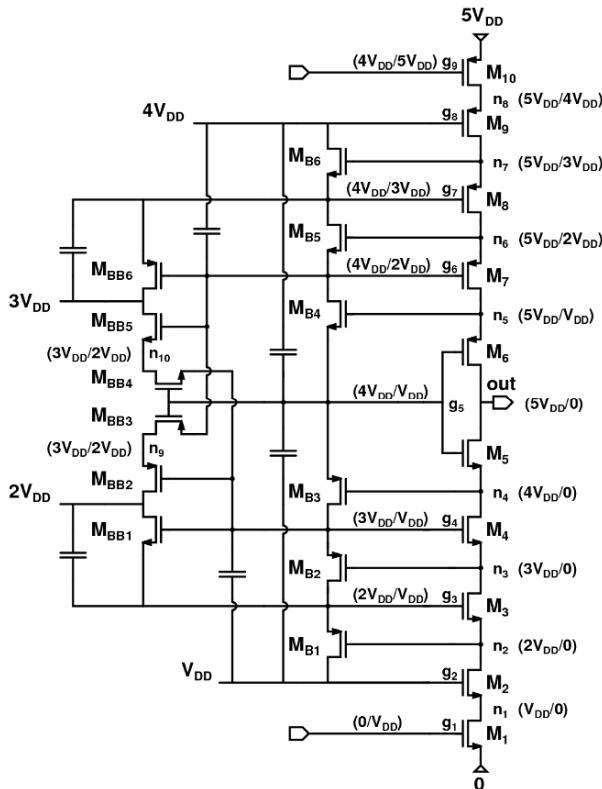


Fig.14: Schematic of the stacked transistors with the bias circuit

Since this switching does not occur immediately for all transistors at the same time, it is decided to use only 90% of the nominal supply voltage for the

devices. The lowering of the nominal supply voltage results in a voltage headroom for the transistors during transients.

Pre-Driver Circuit

The output level of the high voltage buffer is controlled by switching the outer stacked transistors M1 and M10 on and off. Therefore, two control signals are necessary. The pre-driver circuit generates these two control signals. The following functions must be fulfilled by the pre-driver circuit:

- Since the on-resistance of the stacked transistors needs to be low, their input capacitance is very large. Therefore, the two control signals must be buffered before driving these transistors.
- The pre-driver circuit must contain a level shifting function. To switch the transistors M1 and M10 on or off, the following control signals are necessary: gnd and V_{DD} for M1 and $4V_{DD}$ and $5V_{DD}$ for M10 if the supply voltage is defined between gnd and $5V_{DD}$. Or more general, a voltage offset of $(n-1)V_{DD}$ between the two control signals is necessary for driving a nV_{DD} voltage circuit.
- To minimize the distortion of the output of the high voltage driver, matched delay paths are necessary from the input of the pre-driver circuit to the gates of the outer stacked transistors.
- Since the high voltage driver is used as the output stage of a SOPA line driver, its delay must be low enough for handling limit cycles up to 40MHz. A Non-Overlapping Switching (NOS) circuit should be added in the pre-driver structure if timing permits, since short circuit currents can lead to a considerable amount of power dissipation in high voltage circuits.

Figure 15 shows a schematic representation of the pre-driver circuit, which fulfills the previous stated functions. It is comprised of three main parts: a NOS circuit, two level-shift circuits and two tapered buffers. Matched delay paths are obtained by using symmetrical voltage supplies, defined between $-nV_{DD}/2$ and $nV_{DD}/2$, combined with two level-shift circuits, one with an upwards voltage offset of $2V_{DD}$ and one with a downwards voltage offset of $2V_{DD}$. The resulting total voltage offset is thus $4V_{DD}$, which is necessary for a $5V_{DD}$ voltage circuit. The pre-driver circuit converts thus two complementary square waves, defined between $-V_{DD}/2$ and $V_{DD}/2$ to two control signals, one defined between $-5V_{DD}/2$ and $-3V_{DD}/2$ for the nMOS input and one defined between $3V_{DD}/2$ and $5V_{DD}/2$ for the pMOS input. The level-shifters are placed in between the NOS and the tapered buffers. The NOS uses feedback. Therefore, two extra level-shift circuits should be necessary if the level-shifting is done before the NOS. To keep the dimensions of the level-shift circuits small, their driving capacitances should be small. Therefore, the level-shifters are placed before the tapered buffers. In the next paragraphs, each of these building blocks are shortly discussed.

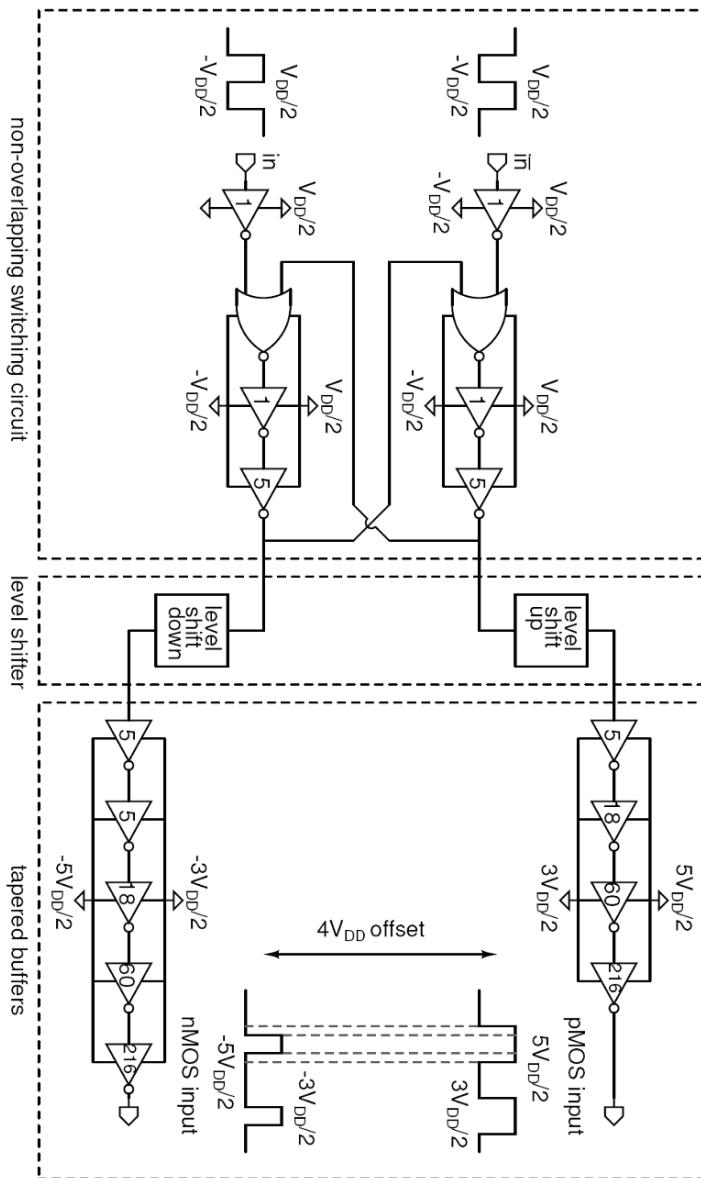


Fig. 15: Schematic of the pre-driver circuit

Non-Overlapping Switching Circuit: The NOS scheme is implemented by using feedback over two nor-gates. The switching signals are shifted with the delay of the inverters in the loop.

Tapered buffer: The scaling factors of the tapered buffer are shown in figure 15. These factors are larger than the scaling factors optimized for speed. In this way the area and power dissipation is reduced at the cost of extra delay, as described in [5]. However, the delay of the buffer equals 170ps, which justifies the use of a NOS circuit.

Level-Shifter: Thanks to the implementation strategy of using two level-shift circuits instead of one, the voltage offset per level-shifter is reduced to $2V_{DD}$, whereas an implementation with one level-shifter requires an offset of $4V_{DD}$ for the level-shift circuit. This reduces the complexity of the level-shift circuit dramatically.

Figure 16 depicts the circuit schematic of the developed level-shifter for an upwards offset of $2V_{DD}$. The supply voltages of these circuits correspond with the symmetrical supplies stated in figure 15. It consists of two differentially switched transistor ladder networks. The ladder networks divide the high supply voltage such that the voltages across the terminals of each transistor in the circuit are limited. The inverter A provides, together with the coupling capacitor Cu, faster switching of the current sense circuit comprised of the transistors Mp1 and Mp2. The speed of the level shift circuit is thus made almost completely independent of the time needed to charge and discharge the internal nodes of the transistor ladder network. The delay of the circuit depends now on the delay through the extra inverter. Moreover, this extra inverter can now be sized for charging and discharging the output load. The transistor ladder network is implemented with stacked transistors. Two extra diode-configured transistors M_D are added per transistor ladder network to set off voltage transients that could occur during switching.

5.3. Measurements

Figure 17 shows a realization of the presented high voltage aDSL2+ line driver in a 1.2V 130nm CMOS technology.

Figure 18 shows the measurement of the self-oscillation frequency of a single SOPA. The self-oscillation frequency of 25MHz is set by the filters and an output voltage swing of 4.7V is achieved with a load of 12.5Ω . A DMT signal consisting of 512 tones with a tone-spacing of 4.3125kHz is applied to the driver to derive the Missing Tone Power Ratio (MTPR) performance. Tones 1-64 are left unused to form the upstream tones and tones 102, 183, 286, 381 and 462 are left out as antenna tones. Figure 19 shows an MTPR measurement of the

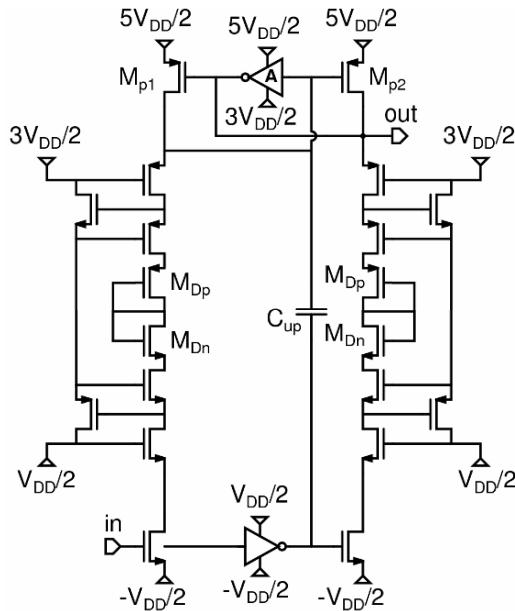
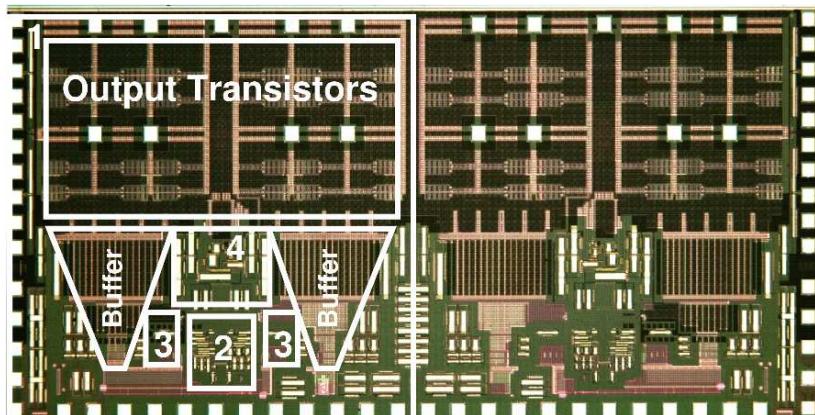


Fig.16: Circuit schematic of the level-shifter



1. Single SOPA
2. Integrator and comparator
3. Level shifter
4. Bias circuit

Fig.17: Die micrograph of the aDSL2+ CO line driver

antenna tone at the highest and most critical frequency. An MTPR of 58dB has been achieved for a DMT signal with an average output power of 20dBm and a CF of 5.6. The total power dissipation is 237mW resulting in an efficiency of 42%. Table 2 summarizes the process and performance specifications.

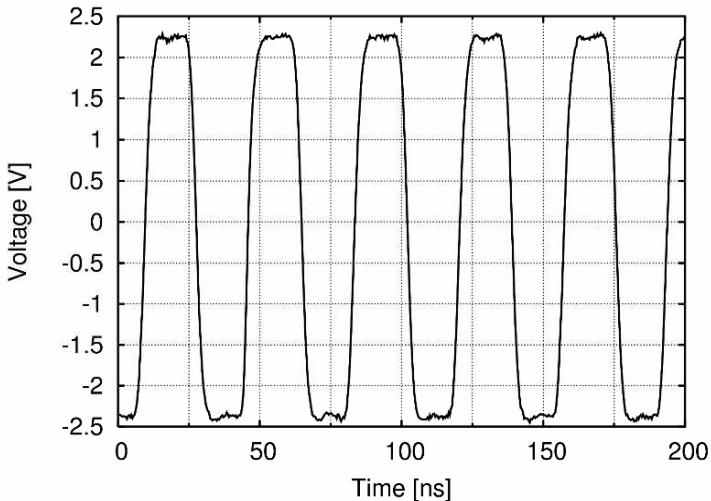


Fig.18: Measurement of the output square wave of a single SOPA

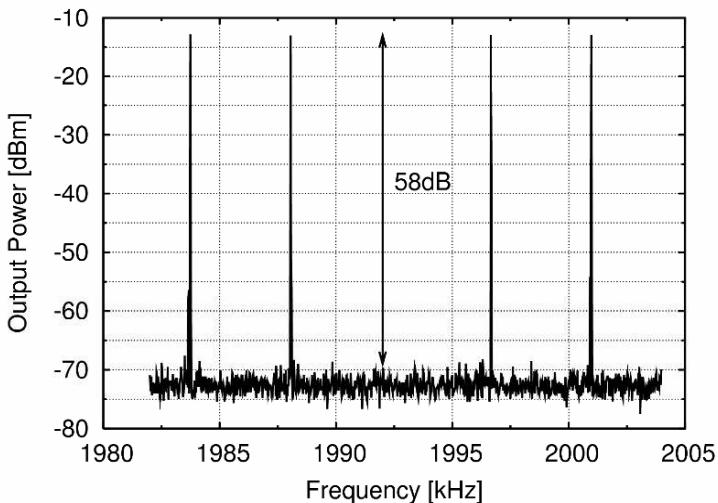


Fig.19: A 58dB MTPR line measurement around tone 462

Table 2: Measured line performance

Parameter	This work	aDSL2+ specification
Technology	130nm CMOS	
V_{DD}	1.2V	
V_{DD} high	5.5V	
V_{ptp} @ 12.5Ω	4.7V	
Bandwidth	>2.2MHz	2.2MHz
P_{out}	20dBm	20dBm
CF	5.6	>5
MTPR	58dB	55dB
P_{total}	237mW	
Efficiency	42%	

6. Conclusions

With the advent of the nanometer CMOS era, sub 1V supply voltages are not an exception anymore. These low supply voltages limit the specifications of analog power delivering circuits dramatically. In this paper the effect of the low supply voltages on the SOPA architecture is described. The technique of stacking standard CMOS devices, for designing high voltage circuits, is discussed and such a high voltage output buffer is integrated into the SOPA system. The result is a fully integrated high efficiency aDSL2+ CO line driver in a mainstream CMOS technology. This proves the working principle of stacking devices. Moreover, the technique can be used for designing high-end application like xDSL line drivers. In this way, a solution is provided, allowing fully integrated single chip designs, to circumvent the low supply voltages of nanometer CMOS technologies.

References

- [1] L. Cloetens, "Broadband Access: The Last Mile", ISSCC Digest of Technical Papers, IEEE, Feb. 2001, pp. 18-21
- [2] T. Piessens and M. Steyaert, "SOPA: A High Efficiency Line Driver in 0.35 μ m CMOS using a Self-Oscillating Power Amplifier", ISSCC Digest of Technical Papers, IEEE, Feb. 2001, pp. 306-307
- [3] T. Piessens and M. Steyaert, "Design and Analysis of High Efficiency Line Drivers for xDSL", Kluwer Academic Publishers, 2004
- [4] B. Serneels, T. Piessens, M. Steyaert and W. Dehaene, "A High Voltage Output Driver in a standard 2.5V 0.25 μ m CMOS Technology", IEEE JSSC, Vol. 40, No. 3, March 2005, pp. 576-583
- [5] H.J. Veendrick, "Short Circuit Dissipation of Static CMOS Circuitry and its Impact on the Design of Buffer Circuits", IEEE JSSC, Vol. 19, No. 4, Aug. 1984, pp. 468-473

VoIP SLIC OPEN PLATFORM

THE WIDEBAND SUBSCRIBER LINE INTERFACE CIRCUIT FOR VOICE OVER IP (VoIP) APPLICATIONS

Luc D'Haeze,

Jan Sevenhans, Herman Casier, Damien Macq, Stefan van Roeyen,
Stef Servaes, Geert De Pril, Koen Geirnaert, Hedi Hakim,

Communication High voltage Products

AMI Semiconductor, Belgium

luc_dhaeze@amis.com

Abstract

The large scale deployment of the voice over IP, or voice over packet technology and the migration of this technology towards the edge of the network stimulates a revival of SLIC and CODEC design. The basics of this existing technology are refreshed and the architecture of a specific implementation is described in this paper.

New smart power technologies with single chip low voltage CMOS and high voltage DMOS and bipolar devices allow for full single chip SLIC & CODEC integration. Another option is to put the low voltage CODEC in a deep sub micron system on chip with a mini-SLIC, integrating only the high voltage drivers.

In the field new voice circuits get deployed now to bring voice to the phone over the wideband network and do the conversion from wideband IP coded voice to the analogue real world speech signals as close to the phone as possible via an Analogue Telephone Adaptor [ATA]. The conversion can be done in the phone as a real IP-phone, in a PC (Skype) via the audio of the PC or via an ATA, driving a regular phone. The conversion can also be done in the house via a VoIP to POTS converter next to the wideband modem, in your garage e.g. to drive your existing home phone cabling. The conversion can also be done with an ATA in the street cabinets driving only the last kilometer to your home phone.

In this paper a Short Haul Line Integrated Circuit [SHLIC] is described, used to transmit Voice over Internet Protocol, also called VoIP, IP Telephony, Internet telephony, wideband telephony, wideband phone, and the routing of voice conversations over the Internet or through any other IP-based network. The SHLIC is a pure analogue circuit processed in a high voltage technology.

Also the CODEC, performing the wideband 16Ksamples/s or A/ μ law narrow band signal processing, is briefly described.

1 Introduction to SLIC's, CODEC's, and POTS circuits

The subscriber line integrated circuit [SLIC] also called sometimes the “Silicon transformer” is implementing what the telecom speech transformer did 30 years ago in the Plain Old Telephone System [POTS] (Figure 1). It brings the AC speech and signalling on the line as a signal source with an impedance equal to the central office impedance (Z_{CO}). In addition it has to supply the phone with a battery voltage, also called the DC feed voltage (VDCfeed or Vbat). The SHLIC is controlled by the CODEC.

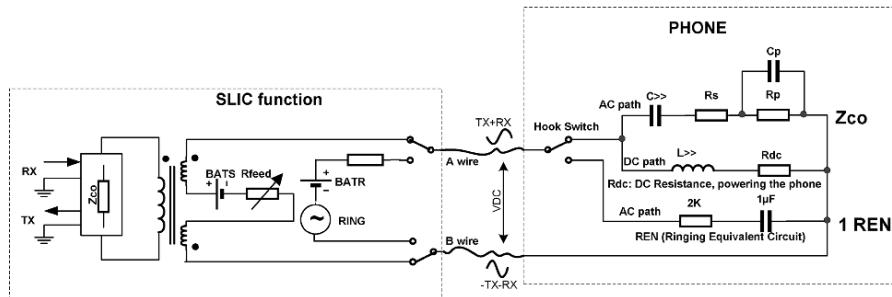


Figure 1

Transformer based Phone connections and equivalent model of an analogue phone set

1.1 Long versus short haul POTS

Short Haul denotes “short lines” or short distance connections (up to 500m). VoIP for home and small offices belongs to the short haul applications.

One example of a VoIP configuration is given in Figure 4. The SHLIC drives a short line from the ATA to the analogue phone. The ATA shown in Figure 4 has an USB interface, but also ATA’s with PCI or Ethernet interfaces exists.

A longer line, up to kilometers long, connecting the central office to the phone, is called a Long Haul connection. Different specifications apply for long and short haul e.g. battery and ringing voltages, lightning protection ...

The long lines in long haul connections require higher battery voltages, larger ringing and metering signals to be transmitted because of the voltage drop on the several kilometers of copper twisted pair between the SLIC and the phone of the subscriber. The DC voltage to power the phone is often 48V or more for long haul public lines and < 32V for short haul applications. Ringing voltages on long haul public lines go as high as 90Vrms on one wire with the DC battery on the other wire. Some countries require symmetrical ringing, some applications require asymmetrical ringing. In public exchanges this high ringing voltage is delivered by a separate ringing generator, which is shared over a plurality of lines in the Central Office (CO) rack. Speech or ringing is switched to the twisted pair telephone line by relays.

Most short haul applications use sinusoidal or trapezoidal symmetrical 50Vrms or 71Vp ringing amplitude to reduce power dissipation in the SHLIC.

In the public long haul line infrastructure, the line conditions can go to extremes as explained in the standards GR-1089-CORE, ITU K11, K20, K39...

Lightning on public lines mainly comes on the old phone wires in the air. Many countries have less severe lightning conditions with the phone wires under the pavement. Also for VoIP applications and short haul in the home and small offices these requirements are less severe.

Primary 700V and secondary 200V lightning protection are external to the SLIC in the long haul public line equipment. The 200V secondary protection is chosen to avoid clipping of the asymmetrical ringing amplitude of 90Vrms plus 48VDC battery.

For short haul, lightning secondary protection could be reduced to <80V to stay above the 50Vrms ringing amplitude with the protector's breakdown voltages. In this case we can consider integrating the secondary protections in the Bipolar-CMOS-DMOS (BCD) SHLIC chip.

In Figure 2 a typical protection scheme is given for short haul applications. The primary protection devices are the Crowbars. The secondary protection consists of a 10Ω protection resistor and a thyristor. The 50Ω resistor is used to sense the line current.

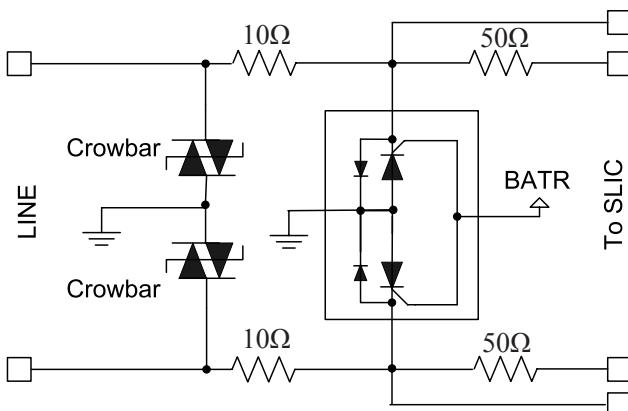


Figure 2
Primary and secondary protections and line current sensing resistors

1.2 The hybrid

The phone is equipped with a hybrid circuit to convert the 2 wire full duplex signal on the twisted pair to a 4 wire signal with separate RX and TX channel and visa versa.

Figure 3 shows an example of a resistive hybrid circuit. The impedance from the line side is $2x(ZL/2)=ZL$ ($R \gg ZL$). A signal VL coming from the line, assume $VTX=0$, is converted to a voltage $VRX=(2/3)*VL$.

On the other hand, a voltage VTX , driven by low impedance line drivers [LD] in the phone, will cause $VRX=0$ and $VL=VTX/2$ for a line terminated with an impedance equal to ZL . With this hybrid circuit, a speech signal incoming from

the line is converted to an RX signal, which is driving the phone earpiece. A signal coming from the TX side, which is the microphone, is transmitted to the line but nothing is transmitted to the RX side or earpiece.

Notice that when the line does not show the correct impedance, a part of the VTX voltage will be reflected and the VTX voltage will no longer be completely cancelled on the RX side. In this case we are talking about an echo on the line, which is very annoying as the user will hear his voice back on his earpiece with a certain delay. On the other hand, in a real phone, a part of the speaker signal is deliberately transmitted to the earpiece otherwise; the callers will be convinced that they are not heard at the other end of the line and that the phone line is “dead”.

Echo can be at the near-end or the far-end. Near-end echo is the result of mismatch between the effective line impedance and the Z_{co} of the SLIC or SHLIC.

Far-end echo is a result of the mismatch of the phone impedance and the effective line impedance or comes from reflections on connections or open branches (stubs) on the line in the phone network.

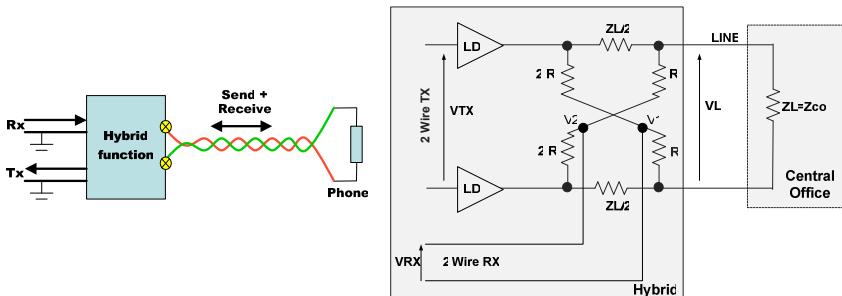


Figure 3
(a) Hybrid function and (b) hybrid schematic example

1.3 HV Supplies

Two high voltage supplies are required for the SLIC or SHLIC.

In speech mode, BATS (typical -32V) determines the supply of the SHLIC line drivers. This DC battery feeds the voltage between A-wire and B-wire, also called VDCfeed.

In ringing mode, BATR determines the supply of the SHLIC line drivers. This voltage depends on the ringing requirement. In ringing mode the applied voltage on the line makes the bell in the phone ring. The AC ringing voltage amplitude can go from 50Vrms in standard short haul to 90Vrms in standard long haul POTS. The short haul typical standard 50Vrms can be realised with -72V for BATR, but some customers with long haul background continue to require 90V-120V-150V ringing. This pushes the BCD technology to higher breakdown voltages and boosts the SHLIC power dissipation.

The integrated SHLIC, described in this paper uses the AMIS I3T80 technology and delivers maximum 50Vrms for the active ringing without external devices.

Both high voltage batteries (BATS and BATR) are negative voltages with respect to the earth potential to avoid electrolytic corrosion of the copper pair.

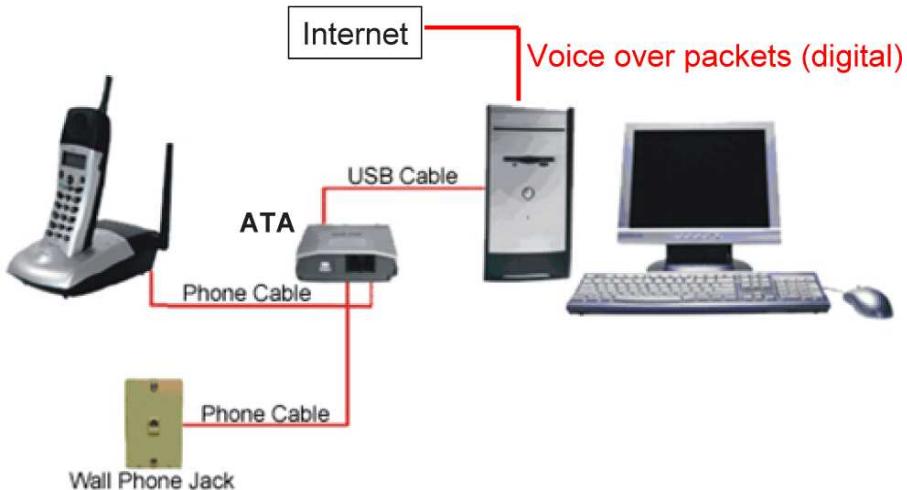


Figure 4
VoIP configuration (ATA with USB interface)

1.4 SHLIC + CODEC functions of a VoIP connection

A typical VoIP connection between the analogue phone network and the digital network is shown in Figure 5. Figure 5 shows what is inside the ATA. The SHLIC and CODEC are converting the 2-wire duplex speech on the twisted pair to a separate receive [RX] and transmit [TX] channel for the 2 persons talking in a voice call between 2 phones. Each phone transmits its TX signal and receives its RX signal.

Phone1 has TX1 and RX1 and phone2 has TX2 and RX2. Moreover TX1=RX2 and TX2=RX1, independently of the switching system: CO (Central Office), PABX (Private Automatic Branch eXchange) or VoIP system between the 2 phones.

The SHLIC is sensing the line current (signal coming from the line) and converts it to an analogue voltage which is passed to the CODEC for signal processing. The SHLIC is receiving the analogue receive signal of the CODEC (signal to be driven on the line) and is driving the line with this analogue signal superimposed on a DCfeed voltage.

The AC impedance presented to the twisted pair by the SHLIC in the speech spectrum from 300Hz to 3.4KHz for narrow band systems and from 50Hz to 7KHz for wideband systems, is the Z_{CO} or central office impedance in the standard POTS vocabulary.

The VDCfeed or battery supply voltage has to feed the phone from the power in the Central Office (CO) or Private Automatic Branch eXchange (PABX) or VoIP modem via the twisted copper pair.

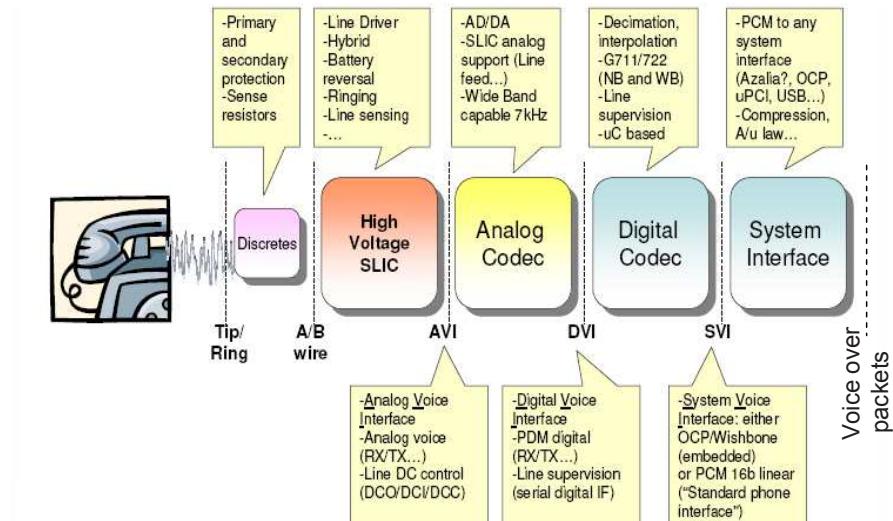


Figure 5
SLIC + CODEC functions of VoIP connection

The DC line current is monitored in the CODEC. A current limitation, in the 20 to 50mA range, is programmed in the CODEC. As soon as the programmed current limit is reached the wire DC voltage is lowered to control the line current. The SHLIC acts as a signal source with controlled DC impedance (Rfeed).

1.5 Line supervision

A phone in the ‘on-hook’ state is for DC an open line with no DC-current flowing. The impedance seen on the line is the bell impedance (Figure 1).

When the handset is lifted (off-hook), a DC load is connected on the line and DC line current flows in the SLIC. When this current is larger than the off-hook detection threshold, the off-hook mode is detected. (Figure 6). The DC load is determined by the cable length and the DC impedance of the phone.

Switch hook event occurs when the user takes the phone off the hook to start or answer a call.

Ring trip detection is the process of an incoming call making the phone ring until the user takes the phone off the hook to answer the call.

Line supervision is the common name for both processes of switch hook and ring trip detection.

Line supervision is done by line current monitoring in the SHLIC and the information is passed to the CODEC for further processing.

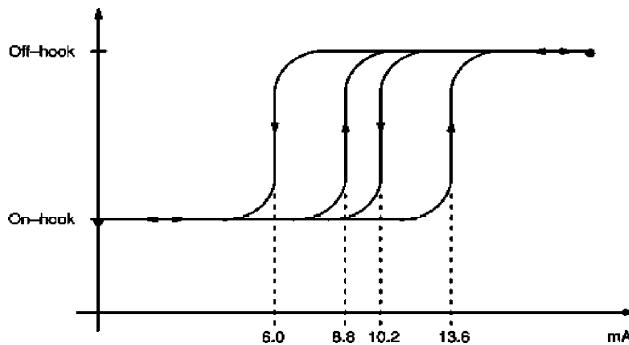


Figure 6
Line current determines on-hook versus off-hook

During ringing the DC current is monitored during the silent phases and the AC current is monitored during the active ringing phases.

1.6 Signals on the line

The signals on the line (Figure 7) contain following information:

1.6.1 Off-Hook transmission

1.6.1.1 Speech signal

The voice is a balanced signal superimposed on the A-wire and B-wire DC battery feed voltage levels.

1.6.1.2 Integrated metering

Metering pulses are signals sent by telephone exchanges to metering boxes and payphones aimed at informing the user of the cost of ongoing telephone calls. The properties of these signals differ between countries, but they typically have a frequency of 12KHz or 16KHz, and duration of several tens of milliseconds. Each pulse represents a certain incremental cost. Therefore, during more expensive calls the exchange will generate more metering pulses per minute than during cheaper calls. Metering pulses are injected via the input RX. Thus for the SHLIC the metering is treated exactly in the same way as the speech AC-signal coming from the CODEC.

1.6.1.3 DTMF: Dual Tone Multi Frequency

The SHPOTS system allows the injection of tones, for signalling or user test purposes. A tone comprising of two programmable sine wave frequencies (Multi Frequency Shift Keying) and programmable amplitudes can be generated. In this way, the most common call-progress and information tones, melody notes, or DTMF tones can be synthesized. The resulting tone signal is added to the speech signal. DTMF signalling is used for telephone signalling over the line in the voice-frequency band to the call switching centre. The version of DTMF used for telephone tone dialling is known by the term Touch-Tone.

A DTMF keypad is laid out in a 4×4 matrix, with each row representing a low frequency, and each column representing a high frequency. Pressing a single key such as '1' will send the sum of two sinusoidal tones with frequencies 697 and 1209 hertz (Hz). The original keypads had levers inside, so each button activated two contacts. These tones are then decoded by the switching centre to determine which key was pressed.

Other DTMF event examples are: busy signal, dial tone and ring-back tone.

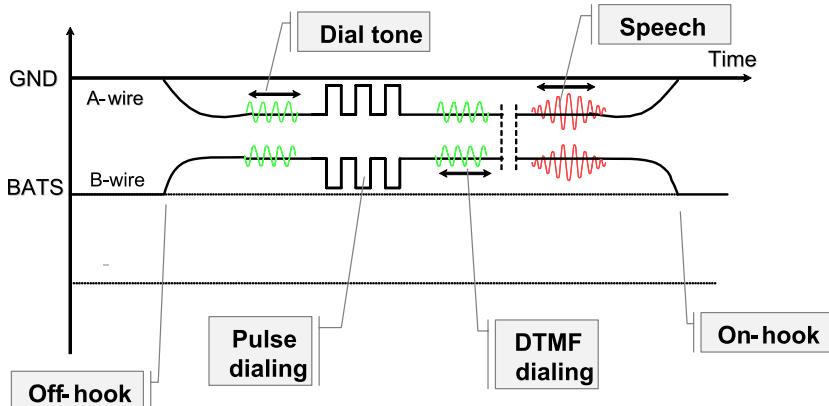


Figure 7

Signals on the line after an off-hook event occurs and a call is set-up

1.6.2 On-Hook transmission

The Analogue Display Service Interface (ADSI) mode is the same operation for the chipset as the off-hook mode except that this mode is programmed via the system software and as such doesn't require DC line current. This state is activated between the two first ring bursts in order to send information on the phone display. CLID (Calling Line Identification) is one of the services of ADSI. It is used to transmit the identity (telephone number or name) of the caller to the receiver telephone set. ADSI can be done with BATS or BATR battery voltage on the line.

1.6.3 Ringing

Ringing can be superimposed to the DC-feed voltage, this is called asymmetrical ringing.

In case of symmetrical ringing, A-wire and B-wire DC voltages are both polarised on the half ringing battery DC level (BATR/2). A balanced low frequency (16Hz-50Hz) sinusoidal or trapezoidal burst of 50V RMS is superimposed on top of this DC signal.

Ringing requirements are measured in REN (Ringer Equivalent Number).

A ringer equivalency number of 1 represents the loading effect of a single "traditional" telephone ringing circuit. Note that modern telephone equipment may have a REN significantly lower than 1. Externally-powered digital-ring phones may have a REN as low as 0.2, while modern analogue-ring phones (where the ringer is powered from the phone line) typically have a REN around

0.8. The total REN for a subscriber's line is simply the sum of the REN's of all devices connected to the line. This number expresses the overall loading effect of the subscriber's equipment on the central office ringing current generator. The local telephone company usually sets a limit on the total REN, typically 5 or less. If the total allowable REN load is exceeded, the phone circuit may fail to ring (or otherwise malfunction). In some countries the REN is better known as the Ringer Approximated Loading number (RAL). In the United States 1 REN is equivalent to a 6930Ω resistor in series with an $8 \mu F$ capacitor. In Europe 1 REN used to be equivalent to an 1800Ω resistor in series with a $1 \mu F$ capacitor. The latest ETSI specification (2003-09) however, calls for 1 REN to be greater than $16 k\Omega$ at 25 Hz and 50 Hz.

The ring signal is a burst signal containing an active (ringing) part and a silent part (Figure 8). The ring signal is generated in the CODEC and sent via the RX path to the SHLIC. The SHLIC RX gain is adjusted to generate more than 40Vrms at the end of the line (GR-909-CORE). The TX/DCO path gain is reduced during ringing mode to avoid saturation of the sense amplifier.

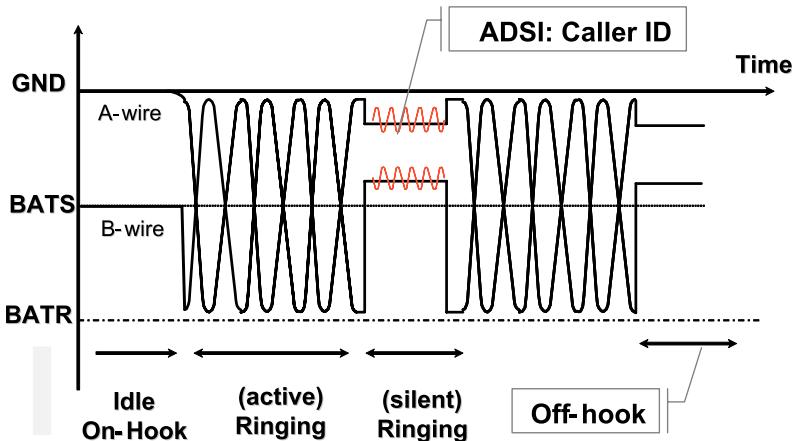


Figure 8
Symmetrical ringing signal

1.7 A/ μ law compression

Only for narrow band speech, the speech TX signal is companded to 8 bit to comply with the 8bit pulse code modulation (PCM) standard. Companding is done according to the A-law or μ -law algorithms. Both are logarithmic based companding algorithms and have similar characteristics.

Symmetrical expanding is done in the RX-channel.

μ -law is a companding algorithm, primarily used in the digital telecommunication systems of North America and Japan. It is similar to the A-law algorithm used in Europe. By convention, A-law is used for an international connection if at least one country uses it.

The μ -law and A-Law algorithms encode respectively 14-bit and 13-bit signed linear PCM samples to logarithmic 8-bit samples at 8Ksamples per second or 64 Kbit/s bit stream.

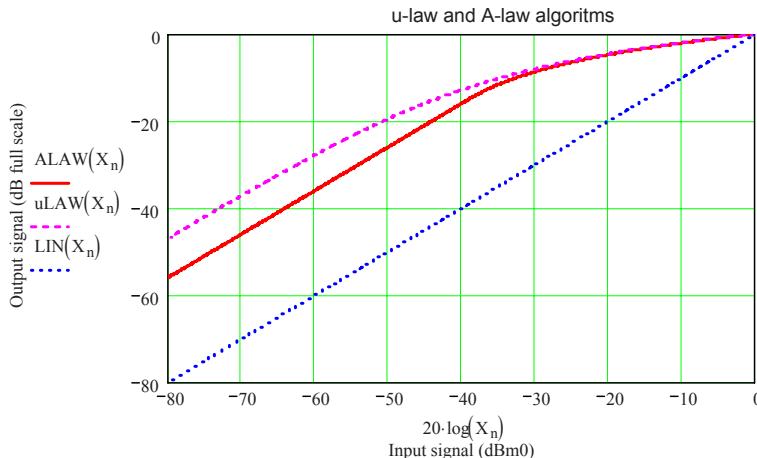


Figure 9
A-law and μ -law Compression versus linear signal

VoIP or wideband telephony does not compand the voice signal and is communicating a linear 16bit signal at 16Ksamples per second or 256 Kbit/s bit stream. Figure 9 shows the transfer function for the different compression algorithms.

1.8 Impedance matching

To have a proper functionality of the phone hybrid, the hybrid has to look into a matched impedance. This impedance is the characteristic impedance of the transmission line, terminated by another phone system with the same characteristic impedance. This ensures that all energy flowing along the line is absorbed at the terminations and no signal is reflected back over the line.

The first widespread users of matched, balanced transmission lines were the early telephone companies. Their early systems had no electronic amplification and delivered maximum audio power from one telephone to another up to 20 miles (32 km) away.

Because propagation time through these 20 miles is significant even at voice frequencies, equipment at each end had to match the line impedance to avoid severe frequency-response errors due to reflections and standing waves. Very old lines typically used AWG#6 (American Wire Gauge) wires spaced 12 inches (305 mm) apart, having a conductor diameter of 0.165 inches (4.1mm). This made their characteristic impedance exactly $Z = 276 \cdot \log\left(\frac{2D}{d}\right) = 600\Omega$ where d is the diameter and D the spacing of the wires (in same units of measure), and Z is expressed in ohms. Therefore 600Ω became the standard impedance for these balanced telephone lines and all early telephone equipment in general.

Interference, mostly coming from the AC power lines running parallel to the phone lines for miles, was largely eliminated through balanced operation of the lines. Balanced (for noise rejection) and impedance-matched (for power

transfer) transmission lines were clearly necessary for acceptable operation of the early telephone systems, which had no amplifiers. Later, as the telephone network grew, amplifiers, filters and “hybrid” transformers were added to enable long-distance transmissions. Proper operation of these components depended critically on precise 600Ω impedances.

1.8.1 Characteristic impedance of a transmission line

The characteristic impedance of a lossy transmission line with the conductance term G_m neglected is given by:

$$Z_0(\omega) := \sqrt{\frac{R_m + 1i\omega L_m}{1i\omega C_m}}$$

The Propagation constant is then given by:

$$\gamma(\omega) := \sqrt{(R_m + 1i\omega L_m) \cdot (1i\omega C_m)}$$

For a VVT-F2 cable with diameter $d=0,6\text{mm}$ and length $l=100\text{m}$ following parameters have been measured: $R_m=128\text{m}\Omega/\text{m}$ for 2 wires (loop); $L_m=775\text{nH/m}$ for 2 wires (loop); $C_m=11\text{pF/m}$. Figure 10 shows the characteristic impedance of this VVT-F2 cable.

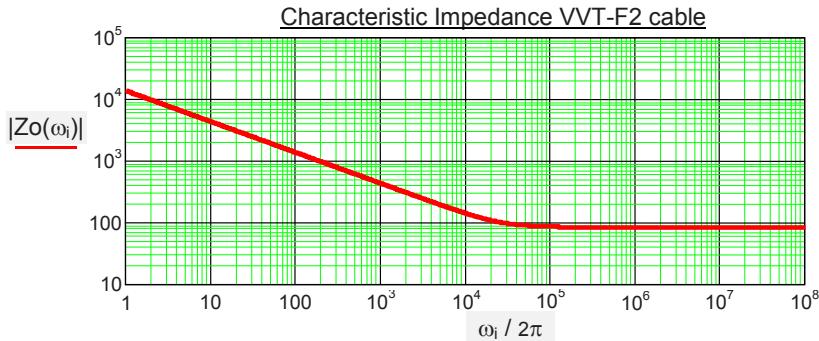


Figure 10
Characteristic Impedance of a VVT-F2 twisted pair cable

1.8.2 Input impedance of a lossy transmission line

The impedance seen at input of the cable when terminated with a load Z_L is given by:

$$Z(\omega) = \frac{A(\omega) \cdot Z_L(\omega) + B(\omega)}{C(\omega) \cdot Z_L(\omega) + D(\omega)}$$

In which the ABCD matrix coefficients are given by:

$$A(\omega) := \cosh(\gamma(\omega)) \quad B(\omega) := Z_0(\omega) \cdot \sinh(\gamma(\omega)) \quad D(\omega) := \cosh(\gamma(\omega))$$

$$C(\omega) := \frac{\sinh(\gamma(\omega))}{Z_0(\omega)}$$

Figure 11 shows the impedance seen at the input of a lossy VVT-F2 cable of 10m respectively 500m when terminated with a pure 600 Ohm and terminated with a 1REN (Europe) load (1800Ω in series with $1\mu F$).

Figure 12 shows a lumped simulation model of the 500m VVT-F2 transmission line with 100 RLC lumped elements in series.

The same figure also demonstrates that this model is correct, up to two decades above the first resonance peak, determined by the total inductance $L_m \times$ length of the cable and $C_m \times$ length of the cable.

The SHLIC needs to terminate the line with the proper line impedance Z_{co} that can vary in value for different country options.

Basically reflections are the basis for choosing the line impedance but for historical reasons different countries have different requirements going from 600Ω to 900Ω for the real part of the line impedance. In addition some countries went as far as adding a reactive part put on paper in the national specification to match the characteristic impedance of the line in that country. For this reason, we have in the SHLIC several impedance synthesis loops. The loops are closed via the CODEC.

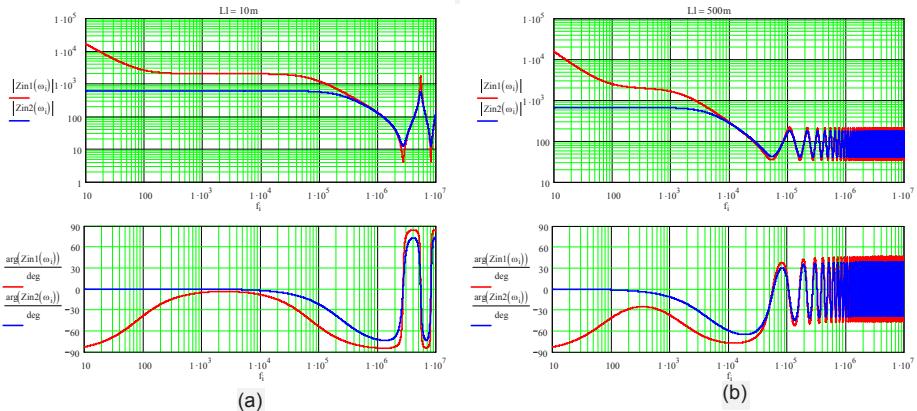


Figure 11

- (a) Input impedance for $L=10m$ VVT-F2 cable terminated with 1REN ($Zin1$) and with 600 Ohm ($Zin2$)
- (b) Input impedance for $L=500m$ VVT-F2 cable terminated with 1REN ($Zin1$) and with 600 Ohm ($Zin2$)

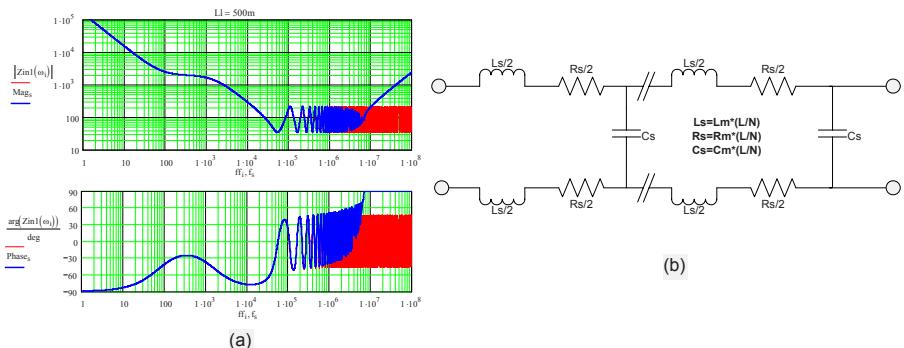


Figure 12

(a) Comparison between theoretical model and simulator output for $N=100$ segments for 1 REN load and $L=500\text{m}$ cable (b) cable model used for simulations

2 An example implementation, SHLIC block diagram explained

The telephone line contains a balanced, full duplex speech signal on top of a DC voltage between the two wires, which supplies the POTS telephone from the exchange (CO). The SHLIC controls and limits the DC voltage between the two wires and converts the balanced, full duplex speech signal on the line into two single-ended speech signals for the CODEC. The TX-signal is the speech signal from the line to the CODEC; the RX-signal is the speech signal from the CODEC to the telephone line. The SHLIC also controls its DC and AC output impedance to the telephone line. The DC output impedance is programmable and sets, together with the DC impedance of the short/long telephone twisted-pair, the maximum current delivered by the CO to the POTS. The AC impedance is a complex impedance, which matches the characteristic impedance of the telephone line. The SHLIC further delivers the ringing signals, does the line supervision and passes the dialling, metering and other signalling from the CODEC to the line.

The SHLIC block diagram is given in Figure 13.

The Line is connected to the SA and SB pins of the SHLIC via 2 protection resistors $R_{pr}=10\Omega$ (not shown in Figure 13).

In the RX-path (bottom side of Figure 13), the single ended speech information from the CODEC (pin RX) is first converted into a differential signal and then superimposed on the DC information, also controlled by the CODEC (pin DCC). The two DMOS Operations Amplifiers MOA1 and MOA2 then amplify the complete signal and drive the twisted pair telephone line via the external resistors R_s (pins AW and BW). In the TX-path (middle part of Figure 13), the TX signal is first extracted from the full duplex line signal by the Herter bridge, which consists of the two external resistors R_s between pins AW-SA and BW-SA and the four bridge resistors before the sense amplifier. The bridge is also

complemented with three extra resistors, amplifiers DCA and gm and the external filter capacitor Ci to measure the DC current in the line (pin DCO). The SHLIC further contains a battery switch between the speech battery voltage BATS and the larger ringing battery voltage BATR, a test switch (TST) to test the transmission characteristics, a bandgap voltage reference, voltage regulator, bias circuits, control logic and various protection circuits.

The signal processing of the SHLIC+CODEC is shown in Figure 14. There are four signal loops in the system. The first is the measurement loop, which extracts the DC line current and the AC TX-speech signal from the line signals. The second is the Rfeed-loop, which drives the programmable battery and controls the DC feed impedance to the line. The two other loops are impedance synthesis loops, which control the real part, respectively the complex part of the AC output impedance in the speech band.

2.1 Measurement loop

This loop consists of two external 50 ohm resistors Rs, the four internal bridge resistors (R1 in Figure 14), the “Sense” amplifier with its two resistors (R6 in Figure 14), the integrator “DCA” with its internal resistor Ri and external capacitor Ci and the transconductance stage “gm”. The external resistors and the bridge resistors convert the differential line current into the signal Tx. Tx contains both the AC speech signal and the DC line current information. To cancel out the DC component from the Tx signal, an integrator (DCA) with low frequency pole ($0.17\text{Hz} = 1/2 * \pi * R_i * C_i$) together with a GM stage is acting on the bridge resistors. In this way, the DC part of the line current is split from the speech signal and the Tx signal, which is transmitted to the CODEC does not contain a DC component. This maximizes the effective signal range for the speech A/D converter in the CODEC. The DCO signal, which is the output of the integrator, contains the DC part of the Line current and is also transmitted to the CODEC for further signal processing and control.

Not shown in the schematics is the reduction of the gain around the sense amplifier during ringing in order to avoid saturation of the measurement loop.

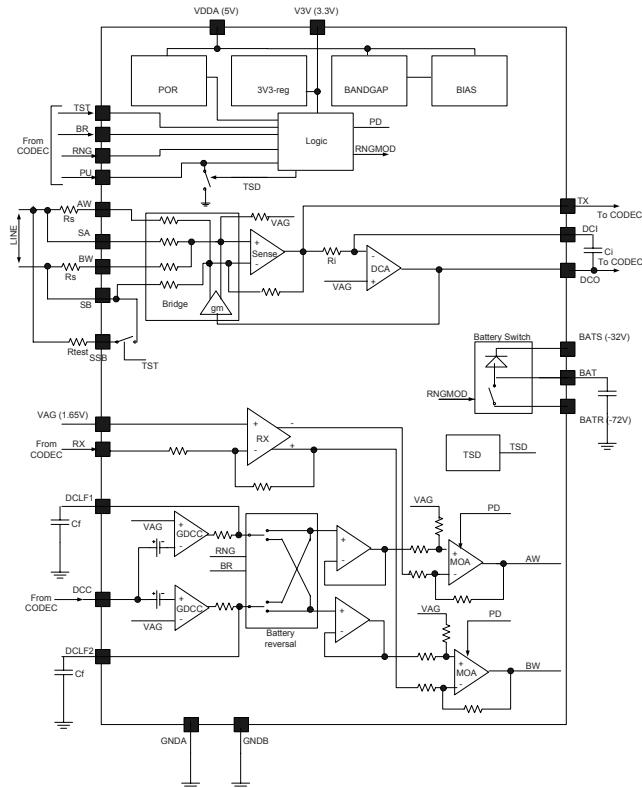


Figure 13
SHLIC block diagram with basic external components

Transfer function of the measurement loop:

Bridge:

$$V_{TX} = \frac{R_6}{R_1} \cdot (I_1 - I_2) \cdot R_s$$

I1 and I2 are the line currents. Also $I_L = I_1 = -I_2$

Integrator:

$$V_{DCO} = \frac{V_{TX}}{s \cdot C_i \cdot R_i}$$

Gm stage:

$$I = g_m \times V_{DCO}$$

Voltages on TX and DCO

Combining the 3 above equations gives:

$$\frac{V_{TX}}{I_L} = R_s \cdot N_2 \cdot s \cdot C_i \cdot R_i \cdot \frac{2}{s \cdot C_i \cdot R_i - 2 \cdot N_1 \cdot N_2}$$

And

$$\frac{V_{DCO}}{I_L} = N_2 \cdot R_s \cdot \frac{2}{s \cdot C_i \cdot R_i - 2 \cdot N_1 \cdot N_2}$$

With:

$$N_1 = g_m \cdot R_1 \quad N_2 = \frac{R_6}{R_1}$$

For $R_s=50\Omega$; $N_1=4$; $C_i=2.2\mu F$; $R_i=433K\Omega$; $N_2=5/3$:

Voltage on TX is a high pass filtered line current with pole at 2.2Hz. Its pass-band trans-impedance gain is: $2 \cdot R_s \cdot N_2 = 133.33\Omega$

Voltage on DCO is a low pass filtered line current with pole at 2.2Hz. Its pass-band trans-impedance gain is: $\frac{R_s}{N_1} = 12.5\Omega$

The TX (speech part) and DCO (DC part) have two very different trans-impedance gains to fit in the 2Vpp dynamic range of the low voltage CMOS in the CODEC. This is done in this way, as the speech signals (AC parts) are much lower in amplitude than the DC part (DC wire current).

The noise of the measurement loop is an important design constraint since the noises of the sense amplifier, the integrator and the gm stage are considerably amplified towards the TX pin, which contains the speech signal.

It can be easily shown that the input referred noise of the sense amplifier is multiplied by a factor $2x(R_6/R_1)+1$ towards the output TX which is 4.3 for $R_6/R_1=5/3$

It can also be easily shown that the input referred noise of the gm stage is multiplied by a factor $2xg_{mX}R_6$ towards the output TX that is 13.3 in our case.

As the input is of the gm stage is coupled via the large C_i capacitor towards the input of the integrator, the integrator input referred noise is also multiplied with the same factor 13.3 towards the TX output.

To minimize the noise in the speech band, amplifiers with bipolar inputs have been used for these circuits. Their equivalent input flicker noise has to be below $30nV/\sqrt{Hz}$ at 1kHz.

Besides the noise, the matching of the external 50 ohm resistors and the internal bridge resistors is critical to preserve the longitudinal balance of the signals on the two wires of the telephone line.

22 Rfeed loop

The measurement loop described above, measures the DC current flowing in the line and sends the DCO signal to the CODEC for loop supervision (LSV). This loop from DCO to DCC contains an AD, the LSV and a DA (Figure 14). The loop supervision controls the DC voltages, generated by the SHLIC through the DCC pin. DCC is the SHLIC input, which controls the DC battery voltage between A-wire and B-wire as a function of the line DC current taken by the phone(s) served by the line.

As soon as the line current goes above a programmed level, the loop supervision in the CODEC drives the DCC-node voltage back to the SHLIC to realize the Rfeed resistance of the battery.

The voltage on the A-Wire is normally GND-3V and the voltage on the B-Wire is BATS+3V as long as the current in the line is lower than the current limit threshold (Figure 17a and Figure 17b). As soon as the current goes above a programmed current level, DCC is driven by the loop supervision in the CODEC to lower the voltage on the A-Wire and to increase the voltage on the B-Wire (Figure 17a). The resulting smaller DC feed voltage lowers the DC current in the line and synthesizes the Rfeed impedance.

The input signal DCC on the SHLIC is level shifted, converted to a differential signal, amplified by 15 by the amplifiers GDCC and low-pass-filtered by the external capacitors DCLF1 and DCLF2 before being added to the differential speech signal in the MOA's (Figure 13). The Rfeed impedance is further calculated together with the AC impedance synthesis.

Rfeed Transfer function:

The DC loop from DCO through the CODEC (Figure 16) will cause a voltage on AW and BW determined by:

$$V_{AW} = 15 \cdot V_{DCO} \cdot G_{DC} \quad \text{and} \quad V_{BW} = -15 \cdot V_{DCO} \cdot G_{DC}$$

The impedance seen between AW and BW is given by:

$$Z_{o_{DC}} = \frac{V_{AW} - V_{BW}}{I_L} = 60 \cdot G_{DC} \cdot R_s \cdot \frac{N_2}{s \cdot C_i \cdot R_i - 2 \cdot N_1 \cdot N_2}$$

And $R_{feed} = Z_{o_{DC}} + 2 \times R_s + 2 \times R_{pr}$

For the given value of N_1, N_2 and $G_{DC} = 0.53$ we obtain, for $s=0$, $Z_{o_{DC}} = 200\Omega$. In series with Z_o we have two external sense resistors ($R_s = 50\Omega$) in series with 2 external protection resistors ($R_{pr} = 10\Omega$) making the generator output impedance $R_{feed} = 320\Omega$.

Figure 17b shows the Rfeed control system as function of the DC line current. For rather small line currents the loop is not activated making the Rfeed impedance equal to $2 \times (R_s + R_{pr}) = 120\Omega$. When the line current reaches a programmed value the Rfeed loop is activated, generating an impedance Rfeed. The impedance presented to the line caused by both Rfeed and Z_o loops is given in Figure 17c.

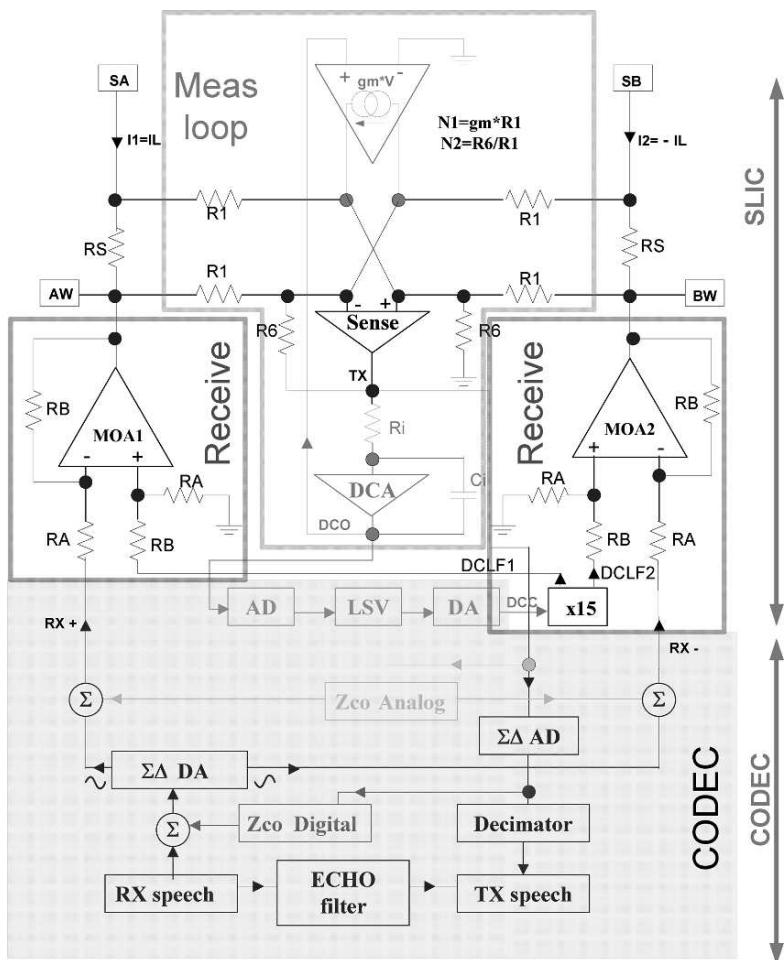


Figure 14
Line impedance synthesis loops

2.3 AC Impedance synthesis loops: ZCO loops

The telephone standards require the synthesis of a complex impedance as shown in Figure 15 within the speech band: 300Hz to 3.4KHz for narrow band and 50Hz to 7KHz for wideband systems.

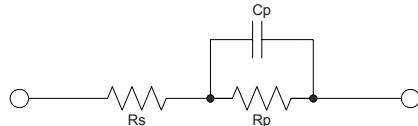


Figure 15
Standardized Zco

The impedance is synthesized with two loops: one for the real part of the impedance and one for the complex, capacitive part of the impedance.

The real part of Zco is represented by a Zco Analog box in Figure 14. It loops the speech signal TX back to the RX path with a programmable gain and controls in this way the real part of the line impedance in the speech band e.g. 600Ω or 900Ω. In order to synthesize a real impedance, there may be no phase shift between the TX and RX signals. For this reason the real part synthesis needs to be done in the analogue circuit domain without the latency of digital signal processing. The synthesizer is a soft programmable gain amplifier between the TX and the RX pin of the CODEC.

The reactive part of Zco is represented by a Zco Digital box in Figure 14. It operates on the pulse density modulation [PDM] TX signal from the sigma delta A/D converter tapped before the decimator filter. It is synthesized in the digital domain with soft programmability and fed back to the digital sigma delta modulator in the RX path to be added to the RX speech signal.

Zco analog impedance transfer function:

The AC loop from VTX through the CODEC and the two MOA amplifiers (Figure 16) will cause a voltage on A-Wire and B-Wire determined by:

$$VAW = 2 \cdot VTX \cdot G_{AC} \quad \text{and} \quad VBW = -2 \cdot VTX \cdot G_{AC}$$

Where VTX/IL is defined above.

The impedance seen between A-Wire and B-Wire is given by:

$$Z_{o_AC} = \frac{VAW - VBW}{IL} = 8 \cdot G_{AC} \cdot Rs \cdot N2 \cdot s \cdot Ci \cdot \frac{Ri}{s \cdot Ci \cdot Ri - 2 \cdot N1 \cdot N2}$$

And $Z_{co} = Z_{o_AC} + 2xRs + 2xRpr$ (Rpr are the 10Ω protection resistors given in Figure 16).

For the given value of $N1, N2$ and $G_{AC}=0.75$ we obtain $Z_{o_AC}=500\Omega$ for frequencies $\gg 2.2\text{Hz}$. In series with Zo we have 2 external sense resistors of each 50Ω in series with $2xRpr=2x10\Omega$ making the generator output impedance $Zco = 620\Omega$.

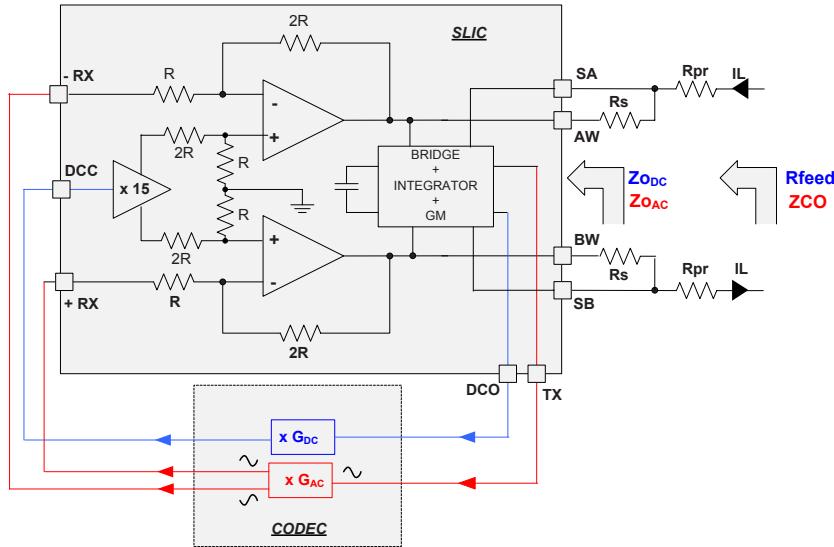
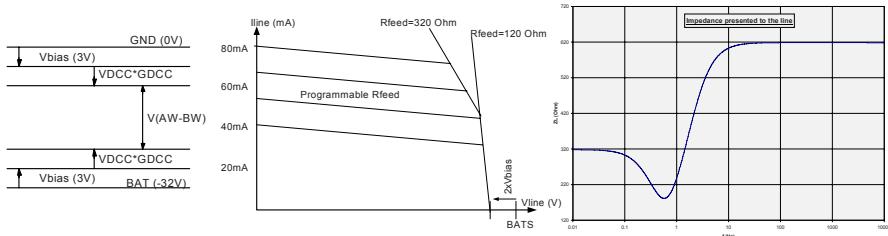


Figure 16
Zco and Rfeed loops via the CODEC



- (a) DC wire voltages: VDCC is the level shifted DCC voltage from the CODEC and GDCC=15 is the SHLIC DCC-gain.
 (b) Typical Rfeed characteristic
 (c) DC and AC Impedance, seen from the line in normal operation

2.4 Longitudinal signal suppression

The line is exposed to EMC disturbances. A common mode signal injected on the line should not be converted to a signal on TX. As can be seen when looking to the VTX equation (see transfer function of the bridge) we can see that the voltage on TX is 0 when the line currents I_1 and I_2 are equal. It can be shown that a mismatch on the external resistors R_s as well as on the internal bridge

resistors R1 and/or R6 will cause VTX to deviate from zero when a common mode signal is injected on the line.

The bridge must be very symmetrical. Both the internal resistors and the external current sense resistors have to be well matched. A mismatch between the bridge resistors or external sense resistors is converting a common mode line current into a differential line current. Also for the internal resistors Medium Ohmic Poly (MOP) resistors have been used because they have the smallest temperature gradient and are placed far away from the MOA's to avoid that a temperature gradient over the chip causes mismatch in the bridge resistors. The bridge can be seen as an instrumentation amplifier. When applying a common mode input signal the CMRR will decrease for worse matching of the resistors. To reach 40 dB longitudinal balance the resistors have to match better than 1%. A matching better than 0.1% is required for 60dB longitudinal balance.

2.5 DC-bias

In the DC-BIAS function, the voltage entering via the DCC pin is superimposed on the internally generated Vbias voltage to bring the A-Wire (B-wire in case of reversed battery polarity) a voltage Vbias below the GND ground level and to bring the B-Wire (A-wire in case of reversed battery polarity) a voltage Vbias above the VBAT battery level. Vbias is typical 3V.

The DC wire biasing voltages DCLF1 and DCLF2 are externally decoupled (Cf in Figure 13) to filter out the noise of the DCC amplifiers.

2.6 3.3V Regulator

The regulator function is converting a 5V external supply voltage to a regulated 3.3V for the CMOS CODEC.

The analogue ground VAG is derived from this 3.3V regulated supply in the CODEC and fed back to the SHLIC together with the single ended analogue signals RX and DCC, generated in the CODEC.

2.7 RX amplifier

The RX amplifier is converting the single ended RX signal to a differential signal to drive a balanced signal to the line via the MOA's. The RX gain is increased during ringing mode to generate a larger AC signal on the line. The gain control is implemented in the CODEC functions. The noise of the RX amplifier has to be low as it is multiplied via the MOA's towards the line.

2.8 Battery switch

The selection of the battery BATS or BATR is defined by the status of the RNG input. In the case BATS is selected (RNG=0), the internal battery VBAT is connected to BATS. BATR is then only used to polarise the substrate.

In the case BATR is selected (RNG=1), VBAT is connected to BATR via an internal switch. This switch is capable to conduct the total maximum line-current. The battery BATS is now disconnected from VBAT by an inverse polarised diode. BATR must in any case be the most negative battery supply.

2.9 Test switch

An internal test switch is positioned between the pin SB and pin SSB. Connecting an external load between SSB and SA allows test of the transmission characteristics in (simulated) off- and on hook conditions. The typical on resistance of the test switch is around 510 ohms.

2.10 MOA

The MOA is the line driver amplifier. The line drivers form a pair of balanced drivers to interface directly with the subscriber loop through a pair of matched external feed resistors. The line driving opamps are protected against short circuits by a current limitation and against excessive power dissipation by an over-temperature protection circuit (TSD), which sets the outputs AW and BW in high impedance mode and the SHLIC in power down.

The load on the MOA amplifier consists out of a 100pF external capacitor with a 50Ω series resistor towards the line.

The MOA consist out of a high gain input stage, a Castello stage, and 2 parallel output stages (Figure 18). One output stage is a class-AB stage, the other a class-B stage. The class-B stage is the power booster delivering most of the power with high efficiency, but moderate linearity. The class-AB stage is a “correction amplifier” which delivers a limited power to linearize the class-B stage.

The class-B stage is able to deliver line currents up to the over current limit of the MOA which is typically 150mA. The currents in the class-AB and class-B stage as function of the load current is given in Figure 19.

The class-AB output quiescent current is set to keep the MOA stable for the max capacitive load and the min output current (for on hook signalling). The class-B stage is inactive for small line currents up to ~ 1.5 mA.

V_{in2} , and the difference in voltage between V_{ref1} and V_{ref2} are setting the currents in the Castello stage. Instead of distributing the MOS diodes on both sides of the Castello stage, all four diodes are placed in the reference branch.

If the 4 stacked devices, to make V_{ref1} and V_{ref2} are of the same size as the 4 Castello transistors then for zero output current, $I_1=I_2=I_{ref}$ and so the class-AB quiescent output current is defined. The currents in the castello stage are given by:

$$I_4 = \frac{I_1}{N} \quad I_2 = I_{o1} + I_4 \quad I_1 = I_{o2} + I_3 \quad I_3 = \frac{I_2}{N}$$

When there is no output current to be delivered by the MOA, then: $I_{o1}=I_{o2}$, $I_{ABP}=I_{ABN}$ and $I_{CP}=I_{CN}=0$. V_{in2} is at the middle of V_{ref1} and V_{ref2} .

For $N=2$: $I_{o1}=I_{o2}=I_{ref}/2$ and $I_{ABP}=I_{ABN}=M \times (I_{ref}/2)$.

Because of the ratio A:1 of the PMOS class-B pre-drivers and the ratio 1:A of the NMOS class-B pre-drivers the class-B output stage is off for small output currents.

When current has to be sourced to the load, I_{ABP} has to increase and therefore V_{in2} will decrease, causing a smaller VGS voltage and a smaller current for transistors Tc1 (I_1) and Tc2, while the VGS voltage and current of Tc3 and Tc4

(I2) will increase. This will cause a larger current Io1 and a smaller current Io2. The class-AB current IABP will increase, while the class-AB current IABN will decrease. The larger the current to be sourced, the larger Io1 and the smaller Io2. As the currents in Tc1 (I1) and Tc4 (I2) are mirrored towards the class-B pre-driver we can see that when increasing the current to be sourced, the class-B P-driver starts to operate as soon as the current through Tc4 equals A times the current through Tc1. Similar, the larger the current to be sunk, the smaller Io1 and the larger Io2. When increasing the current to be sunk, the class-B N-driver starts to operate as soon as the current through Tc1 equals A times the current through Tc4.

The MOA amplifier noise and total harmonic distortion (THD) has also to be low. Worst-case THD occurs in the transition region where the class-B amplifier takes over to deliver power from the class-AB. In this case, for a 1 KHz and 3 KHz dual tone THD values lower than -60dB are obtained.

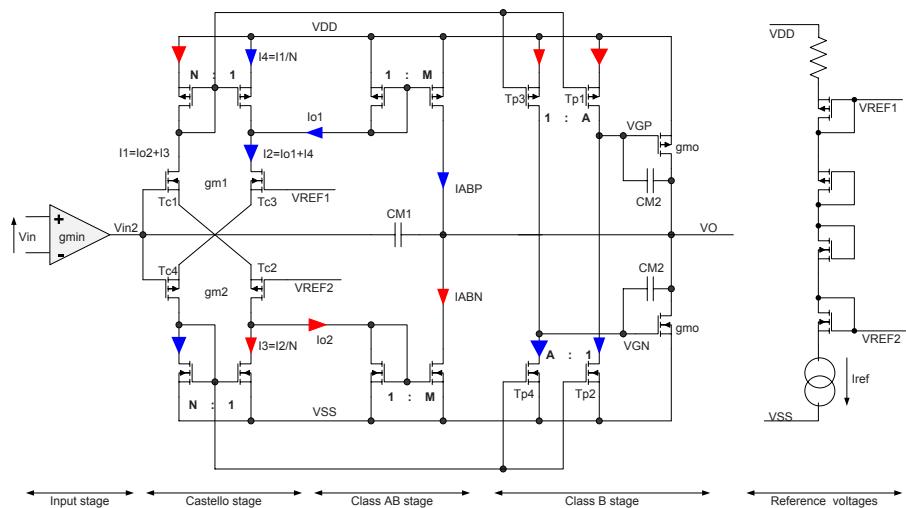


Figure 18
MOA principle schematic

2.11 Battery reversal

The Battery Reversal Circuit, polarises the AW versus the BW voltage. The BR signal determines whether the AW or the BW is the most positive signal.

2.12 POR

Power on reset function monitors the 3.3V regulator output and generates a reset for the digital part as soon as the 3.3V regulated output gets too low.

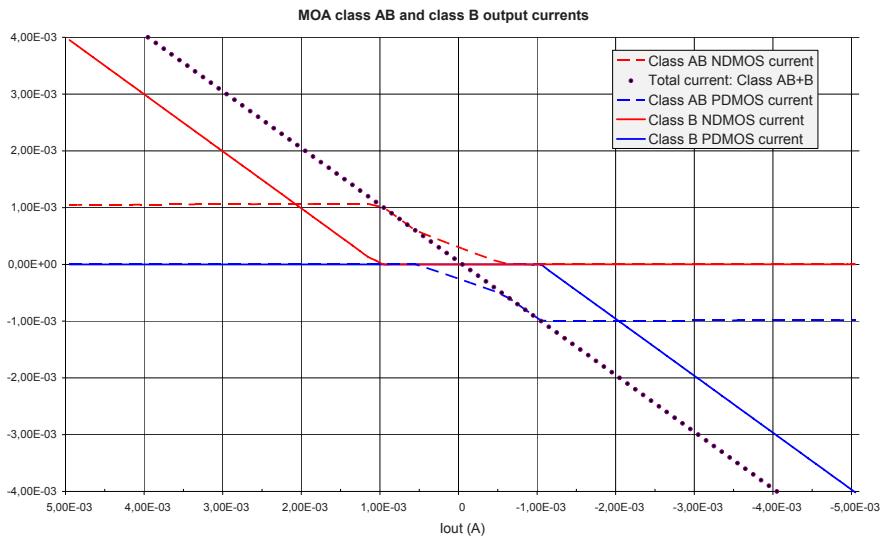


Figure 19
Simulation of the MOA class-AB and class-B currents

2.13 Bandgap

On chip voltage reference generator.

2.14 BIAS

On chip reference current generator.

2.15 CODEC function

The CODEC is performing line termination impedance synthesis for the speech band and the battery feed current over the twisted pair as well as the AD & DA function, G711/G722 filtering and echo cancellation (with 7 static programmable filter coefficients stored in the coprocessor coefficient RAM) or in general all the standard CODEC functions of the speech, metering, DTMF, CLID ... to and from the full duplex POTS twisted pair. It operates on a 3.3V regulated supply delivered by the SHLIC 3V3 pin.

The design is following the G.711 speech-filtering standard for 3.4 KHz speech according to GR-909-CORE specification and the newly desired 7 KHz wideband speech (G.722) that will allow higher quality voice calls.

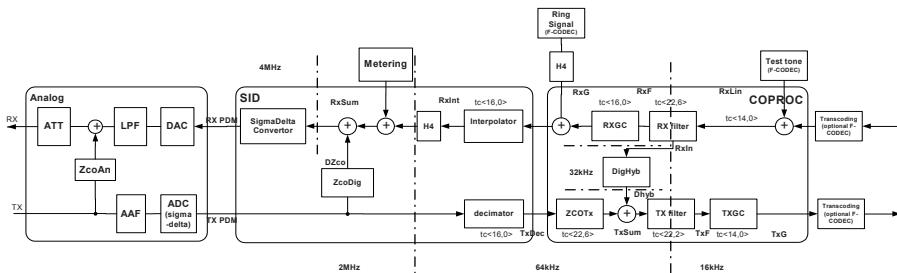


Figure 20
CODEC functional block diagram

2.15.1 The Analogue CMOS CODEC functions

Loop supervision

In the loop supervision the analogue DCO signal is sampled at 2 KHz with 8-bit accuracy and variable digital gain in the digital loop supervision allows Rfeed soft programming.

Anti alias filter

The TX speech A/D conversion is based on 2048 KHz single loop pulse density modulation $\Sigma\Delta$ AD conversion. To achieve sufficient low image noise anti alias filtering on the 7 KHz TX signal with 2048 KHz sample frequency, a second order Butterworth anti alias filter is included.

Switched capacitor AD

A second order switched capacitor sigma delta modulator AD is sampling the analogue TX speech signal at a rate of 2048 samples/sec.

Switched capacitor DA

The DA conversion is a 4096 KHz pulse density modulator followed by a second order-smoothing filter.

The DAC consists of an interpolator moving the sample frequency up in a first step to 256 KHz 16 bit via linear interpolation followed by a digital sigma delta modulator to convert it to a single bit 4096Kbit/s PDM to drive the clamp circuit. The pole of the filter is at 40 KHz to allow the 12 KHz or 16 KHz metering signals to pass over the RX path.

Low pass filter

Is a second order Butterworth low pass filter. The filter does also the conversion from a double-ended signal towards a single ended signal.

2.15.2 The digital CMOS DCODEC functions

The CODEC filter configuration, the control of the CODEC functionality itself and some SHLIC configuration is performed by the FCODEC (Firmware CODEC). This FCODEC API (Application Programming Interface) is firmware running on an external processor.

2.15.2.1 DCODEC hardware content

Decimator (TX)

The filtering of the Sigma-Delta modulator output is carried out in two steps. The decimator attenuates the out-of-band noise, produced by the Sigma-Delta Modulator, and reduces the data rate from 2046 KW/s, 1 bit/word (TxPdm) to 64 KW/s, 16 bits/word (TxDec). The decimation is performed before the filtering, allowing to perform the processing at a lower rate (64 KHz). The decimator consists of 2 consecutive digital filters with intermediate down sampling to reduce the clock speed. The first step is a 16 sample moving average and the second step is a 2 sample moving average to achieve the 32x down sampling.

ZcoTx

The decimation is followed by the conversion of line current to line voltage (ZcoTx).

Interpolator (RX)

The DAC consists of an interpolator moving the sample frequency up in a first step from 64 Ks/s to 512 Ks/s 16 bit via linear interpolation and then a digital sigma delta modulator converts it to single bit PDM 4096 Kbit/s to drive the clamp circuit.

TX/RX wideband filters

The filters characteristics are specified in the G722 standard and implemented in the DCODEC coprocessor HW/FW. The RX/TX filters reduce the TX and RX speech band to 50 Hz – 7 KHz. In TX direction a down sampling is performed from 64 KHz to 16 KHz, in RX direction an upsampling is performed from 16 KHz to 64 KHz.

The Transmit Gain Control (TXGC) and Receive Gain Control (RXGC)

TXGC allows compensating the gain and adjusts the signal levels as required by the administration. The TXGC will output the signal (the 16bit linear code word) at a rate of 16 KW/s. Linear PCM 16 bit samples at system side in the wideband case are processed at 16 KHz. The system software on the processor can do all kinds of encoding before it is transferred over the Internet.

In the receive direction, an inverse process takes place through the Receive trans-coder, RXGC (Receive Gain Control), Receive Digital Filter, interpolator and digital Sigma-Delta Modulator. The RXGC receives the linear signal at a rate of 16 Ks/s, 16 bits/word.

In the narrowband case very often A-law or μ -law encoding and decoding is performed (G711). This function realizes the logarithmic compression of the 16bit 8 KHz speech into the 8 KHz PCM format for 64kBit transmission after a 7dB variable gain setting in steps of 0.25dB.

Echo Hybrid

A programmable digital hybrid is integrated in the VoIP CODEC IP.

For echo cancellation we have a digital programmable filter matching the impulse response of the line with loop phase delay and attenuation over the spectrum.

The digital hybrid takes the signal Rxln before the DC cancellation in the RX filters; from this it computes at a rate of 32 KW/s the echo to be cancelled in the Tx path.

Digital Zco

For capacitive or inductive line termination a pole and a zero is realised to emulate the inductive or capacitive part of the Zco. That is implemented by a multiplication on the PDM TX signal and summation of the product with the RX signal before the digital $\Sigma\Delta$ modulator DA converter as shown in Figure 20.

Besides the processing of the voice signals, the CODEC is also able to generate metering, ringing and test signals. The metering signal is generated at a rate of 1024 KW/s before being added to the 4096 KW/s signal stream. The high metering sampling rate is due to the high 12 KHz and 16 KHz metering frequencies. Ringing and test signal signals are generated at 16 KW/s rate; the ring signal is held for 4 times before being added to the 64 KW/s receive signal stream RxF.

2.15.2.2 FCODEC

Next to the Digital and Analog CODEC a firmware CODEC is implemented. This FCODEC exposes an API (Application Programming Interface) to the system software, allowing easy integration and extendibility of the features.

The F-CODEC API manages the mode of operation of the SHLIC and sends next to voice samples, line control and status information to and from the system software.

The main functions of the FCODEC are:

- The implementation of a phone finite state machine (Figure 21). The line can be placed in different modes according to events coming from the system software or the DCODEC.
- The implementation of advances line test functions.

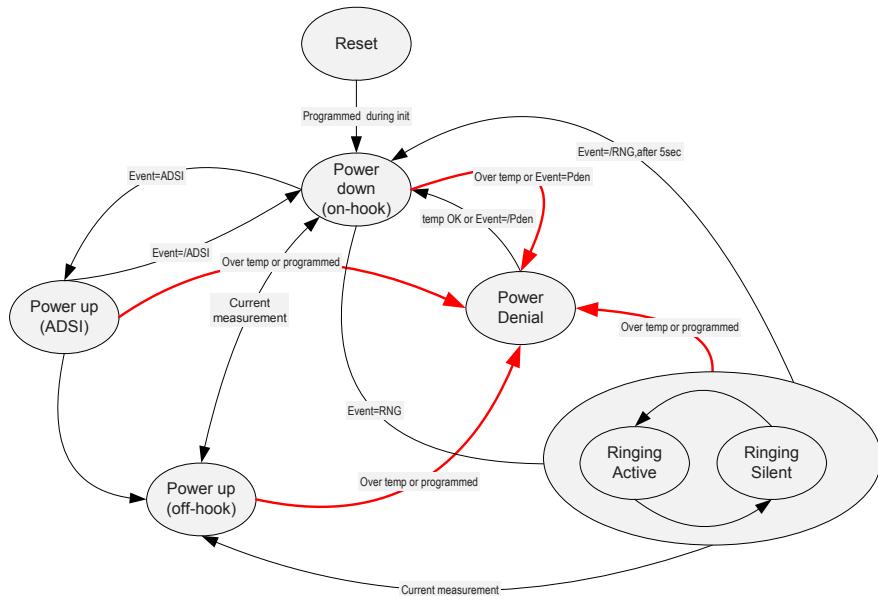


Figure 21
SHLIC management Finite State Machine

3 Norms

Depending of the bandwidth, different standards have to be respected:

Narrow band: 8Ksample/s sound BW = 340 Hz.... 3.4 KHz according to standard G.711

Wideband : 16Ksamples/s sound BW = 50 Hz...7 KHz according to standard G.722

4 Conclusion

The large scale deployment of the voice over IP, or voice over packet technology and the migration of this technology towards the edge of the network stimulate a revival of SLIC and CODEC design. The basics of this existing technology are refreshed and the architecture of a specific implementation is described in this paper.

New smart power technologies with single chip low voltage CMOS and high voltage DMOS and bipolar devices allow for full single chip SLIC & CODEC integration although another architecture placing the low voltage CODEC in the deep submicron system chip with a mini-SLIC integrating only the high voltage drivers is also a new trend in this reborn analogue skill area.

A general view on an architecture tailored towards a modern and flexible VoIP application was described, elaborating in detail:

- The top-level bloc diagram of SHLIC + CODEC.
- The impedance synthesis loops for DC and AC.
- The Line driver principle architecture.

- The different signals along a telephone line.
- Matching and noise requirements of the impedance synthesis loop.
- The usage of a CODEC controlled by an open firmware API.

5 Acknowledgment

The authors express many thanks to Wesley Verbracke and Sven Soleme for doing the layout and to Gert Naert for preparing the test boards.

The authors also thank all the colleagues of the past 3 decades working on POTS in this company. This VoIP POTS overview paper was mainly based on company experience of various colleagues since 40 years of SLIC and CODEC research, development and production in Belgium. More particularly in Oudenaarde, Antwerpen and Gent and this under various corporate company constellations over the years in ITT Bell telephone Manufacturing Company and Mietec, Alcatel and Alcatel Microelectronics, ST and finally AMI Semiconductor now as a result of company strategy.

Most of the POTS information is encyclopedic today but the revival of POTS via the VoIP technology brings new interest in this topic. It again triggers new development and innovation in electronic circuits and silicon technology.

The authors thank the encyclopedic organizations for bringing all this information on the www and encourage the readers to search the web encyclopedia for further historical background.

6 References

100-V high-performance amplifiers in BCD technology for SLIC applications
 Castello, R.; Lari, F.; Siligoni, M.; Tomasini, L.; Solid-State Circuits, IEEE Journal of Volume 27, Issue 9, Sept. 1992 Page(s):1255–1263

A 150-V subscriber line interface circuit (SLIC) in a new BiCMOS/DMOS-technology Zojer, B.; Koban, R.; Petschacher, R.; Sereinig, W.; Solid-State Circuits, IEEE Journal of Volume 32, Issue 9, Sept. 1997 Page(s):1475–1480

A switching regulator and lightning protector for a subscriber line interface circuit Chen, R.K.; Lerch, T.H.; Spires, D.A.; Solid-State Circuits, IEEE Journal of Volume 21, Issue 6, Dec 1986 Page(s):947–955

Part III: Very High Frequency Front Ends

The third chapter of this book is on Very High Frequency Front Ends, operating at 60GHz and above. It addresses high-speed design issues from application level down to technology.

The six papers follow more or less a top-down approach. The first paper of Peter Baltus et al. starts from application and system level and goes down to hardware and circuit challenges and constraints. It addresses the whole chain from EM-wave to bits and vice versa, thus including propagation, antenna, and transceiver, and beam forming through phased-array antenna structures.

The next two papers focus on the circuit-design level, for critical blocks in the chain, each for a given technology. The paper of Mihai Sanduleanu et al. addresses specific circuit designs in full-CMOS for basic blocks in the chain, including oscillators, dividers, and mixers. By proper design, the application specifications are matched with the of full-CMOS device properties. The paper of Ullrich Pfeiffer focuses on RF design concepts for mmWave power amplifiers, realized in SiGe, so with bipolar devices and MOST devices (BICMOS); it addresses also power detection, power coupling and power-combining structures.

The last three speakers start from an even deeper level, the technology level. They address technology properties and device aspects, both for actives and passives, in the general context of very-high frequency operation, and from that they go bottom up to circuit-design level. The paper of Sorin Voinigescu et al. does that for both SiGe and CMOS, gives a comparison between these technologies, and links this to several state-of-the-art IC designs. The paper of Sharon Malevsky et al. also compares BICMOS and CMOS, and applies this to LNA and mixer design for receiver front ends. Finally, the last paper, that of Herbert Zirath et al., focuses on III/V technology for various circuit blocks for integrated front ends for mm-wave applications, and puts this in relation to CMOS.

SYSTEMS AND ARCHITECTURES FOR VERY HIGH FREQUENCY RADIO LINKS

Peter Baltus^{1,2}, Peter Smulders³, Yikun Yu²

¹NXP Semiconductors Innovation Centre RF Caen, France

²Eindhoven University of Technology Mixed-signal Microelectronics Group

³Eindhoven University of Technology Radiocommunications Group

Abstract

This paper discusses very high frequency radio links from the application level down to the circuit constraints. Because of the technical difficulties and higher cost of technologies and packaging, applications at these frequencies only make sense when the special properties of these high frequencies are offering clear advantages for these applications. Such advantages can be higher system capacity, better security and privacy, or higher spatial resolution. Exploiting these advantages requires careful choices in system design and architecture, and imposes specific constraints on circuits and technologies. In most cases, it will also require beam forming through phased array antenna structures. Implementation of the signal processing for beam forming can be achieved in an efficient way in the RF domain.

1 Introduction

Recently, the interest in very high frequency radio links has increased. This is caused by applications that need wireless radio links with high data rates, better security and privacy, and/or better spatial resolution, and by the availability of mainstream IC technologies that allow relatively cheap implementations of the RF parts for such radio links.

In this paper, standardization, regulation and the radio channel will be introduced first (section 1). The motivations for using very high frequency radio links will then be discussed (section 2), as well as the requirements and constraints for transceiver architectures and circuits in radio links for these applications (section 3).

1.1 Developments in standardization

There are plenty multimedia applications calling for wireless transmission at Gbps or multi-Gbps transmission over short distances. Examples are wireless Giga-bit Ethernet (1.25 Gbps), synchronization and high-speed download (as fast as possible) and wireless transfer of high definition video (2 – 20 Gbps). These data rate figures cannot be accommodated in the traditional frequency bands below, let us say, 10 GHz without significant service degradation. However, sufficient spectral space can be found at very high frequencies, e.g. around 60 GHz where in the order of 5 GHz of spectral space has been allocated worldwide for unlicensed use [14]. The reason for this allocation is the occurrence of significant oxygen attenuation in this band which makes it unsuitable for long-range (> 1 km) transmission. Fortunately, this potential mass market for low-cost 60 GHz radios can be addressed by todays low-cost process technology. As a consequence, several efforts are underway to develop standards for radio links at these frequencies.

In March 2005 the IEEE 802.15.3 Task Group 3c was formed to develop a 60 GHz-based physical layer as an alternative for the existing 802.15.3 Wireless Personal Area Network (WPAN) Standard 802.15.3-2003. This promises a high coexistence (close physical spacing) with all other micro wave systems in the 802.15 family. The standard should be ready by May 2008. In addition, there are some ad-hoc initiatives to develop a de-facto standard for more specific products. An example is the WirelessHD consortium which is developing specifications for wireless high definition multimedia interface (HDMI). Target data rates for first generation products are 2 – 5 Gbps whereas scalability to 20 Gbps will be theoretically possible for higher resolutions and colordepth.

1.2 Regulation

Figure 1 shows an overview of the world-wide allocation of the 60 GHz band [16-19]. In this scheme the allocation for Europe is only indicative since it is still under consideration [20]. In order to achieve world-wide harmonization it would make sense to shift the European band 2 GHz downwards. An additional motivation for this is that at frequencies above 64 GHz the oxygen absorption becomes rapidly insignificant (from 7 dB/km at 64 GHz decreasing to only 2 dB at 66 GHz) so this part of the spectrum will be better suited to accommodate other type of applications such as outdoor back haul connections. For the same reason the Japanese regulation should be reconsidered. The world-wide

minimization of bandwidth-spread will also considerably simplify RF and antenna design because bandwidth requirements can only be met at significant cost of other performance figures such as antenna efficiency.

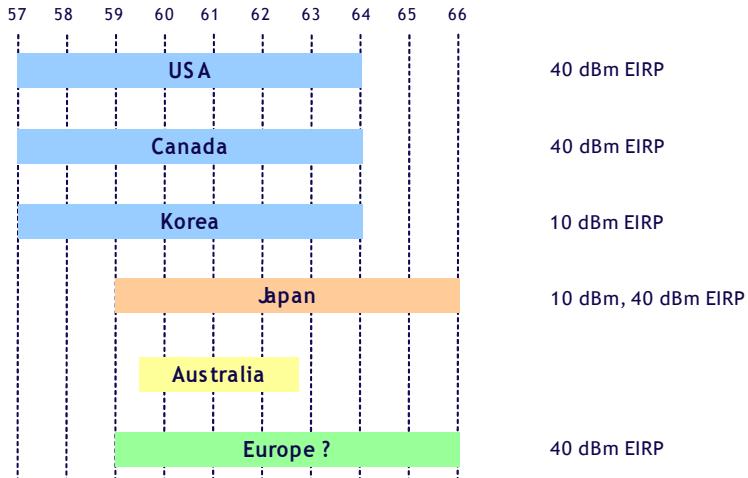


Figure 1: World-wide allocation of the 60GHz band

1.3 Propagation channel

A basic link-budget calculation as given in [14] leads to the conclusion that for the reliable transmission of Gbps, over distances up to 10 meters, antennas must have a relatively high gain. Such antennas do not provide rich multipath which rules out true-MIMO techniques at 60 GHz. On the other hand, antenna gain is easy to achieve with small structures at that high frequency which motivates the use of narrow antenna beams in combination with beam steering techniques in order to increase the flexibility of operation. In what follows, we will therefore focus on the application of high gain antennas.

1.3.1 Small scale fading

Fig. 2 shows the typical variation in received power at 60 GHz over distances that are small or comparable with the free space wavelength (5 mm) when using fan-beam antennas having an gain of 16.5 dBi at both ends of a line-of-sight (LOS) link. The measurement environment is described in [21]. Figure 2a depicts the variability for a narrowband signal whereas Figure 2b shows the

variability for a signal with a bandwidth of 1 GHz. It is observed that with such a large bandwidth the available signal power arriving at the receiving end (RX) varies insignificantly if the position is changed over a limited range, that is, within the small-scale region. This particular characteristic implies that only a very small fading margin is required in the 60 GHz radio design and that the available signal power at a certain position only depends on the large-scale properties of the environment.

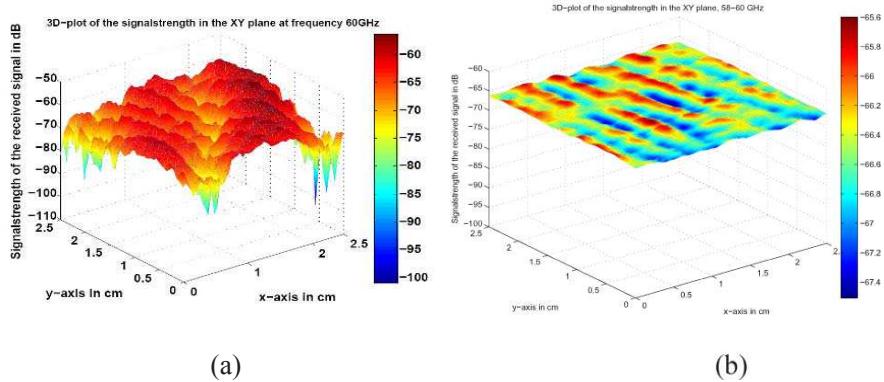


Figure 2: Small-scale variability of the signal strength at 60 GHz for (a) narrowband and (b) 1 GHz wide-band transmission

1.3.2 Large scale fading

The received power from a transmitter at a separation distance d is related to the path loss and can be represented by

$$P_r(d) = P_t + G_t + G_r - PL(d) \quad (1)$$

in decibels, where P_t is the transmit power, G_t and G_r are the antenna gains at transmitter and receiver side, respectively. The path loss is usually modeled over the log-distance as

$$PL(d) = PL_0 + 10n\log(d) + X_\Omega \text{ (dB)}, \quad (2)$$

where PL_0 gives the reference path loss at $d = 1$ m, n is the loss exponent and X_Ω denotes as zero mean Gaussian distributed random variable with a standard

deviation Ω . Wideband (1 GHz) measurements under LOS conditions with the aforementioned 16.5 dBi fan-beam antennas revealed n -values close to 2 which complies with the well-known Friis formula for free-space [21]. Standard deviation Ω is about 1 dB which confirms the low small-scale variability in received power.

With the fan-beam antennas the channel dispersion can be kept amazingly low, in the order of a few nanoseconds at maximum, even if there is considerable mispointing [21]. This implies that with high-gain antennas and under LOS conditions very high data-rates in the order of Gbps can be obtained by applying just a simple modulation scheme and without the need for channel equalization. This is confirmed by practical experiments as presented in [22].

2 Motivation for using very high frequencies

In the context of this paper, very high frequency radio links are defined as the upper part of the SHF band and the lower part of the EHF band, from about 10GHz to 100GHz. There are several motives for wanting to use very high frequencies in radio links:

- the radio spectrum at very high frequencies is still rather undeveloped, and therefore more radio spectrum with wider bandwidths is available at these frequencies;
- the system capacity is higher at very high frequencies because the range of radio signals is limited, resulting in smaller cells. Therefore the same frequency can be reused at shorter distances distances;
- the inherent security and privacy is better at very high frequencies because of the limited range and the relatively narrow beam widths that can be achieved;
- the spatial resolution is better at very high frequencies;
- the physical size of antennas at very high frequencies becomes so small that it becomes practical to build complex antenna arrays and/or further integrate them.

In the following sub-sections, each of these motives will be discussed separately.

2.1 Room at the top

The need for wide-band radio spectrum increases because of systems that require wireless transfer of very high data rates. The demand for such systems is stimulated because:

- people becoming used to higher data transfer rates in wired systems such as dedicated cables (HDMI), gigabit ethernet (IEEE 802.3-2005), Firewire (IEEE 1394) etc. These interfaces offer data rates in the 1Gbps to 10Gbps range. Once people are used to higher data rates for wired systems, they will start to expect similar data rates from wireless systems as well;
- applications that require higher data rates, such as high-definition video, synchronization of portable equipment with large internal memories, etc. are becoming more popular;
- storage systems with larger capacity, that people will want to copy, back-up, synchronize etc. within a limited time;
- signal processing systems that can deal with higher data rate streams are used to develop new applications that often need to receive or transmit these data rates through a wireless link.

Such high data rate radio links are currently being investigated, developed and partly already being used in:

- licensed high-speed microwave links (around 40GHz, 75GHz, 85GHz and 95GHz);
- unlicensed short range data links in the 60GHz band for wireless data networks (802.16, 802.15.3c) and wireless video and audio streaming (WiHD).

These higher data rates can of course be achieved with various RF frequencies, and are by themselves not a motivation for moving to very high frequency radio links. In principle, a high data rate can be achieved by a combination of signal bandwidth and signal dynamic range [1]. The limit for the data rate over a channel is set by the capacity C of the channel and is a function of the bandwidth BW and the signal to noise ratio SNR (3):

$$C = BW \log_2(1+SNR) \quad (3)$$

A high data rate can therefore be achieved with low bandwidths when a high signal-to-noise ratio can be achieved. However, a high signal-to-noise ratio requires either a short distance between transmitter and receiver, or a high transmit power, or high gain antennas, as described in the Friis Transmission equation (4):

$$P_{RX} = P_{TX} G_{RX} G_{TX} \left(\frac{\lambda}{4 \pi r} \right)^{\alpha} \quad (4)$$

In this equation, P_{RX} is the received power, P_{TX} is the transmitted power, G_{RX} is the gain of the receive antenna, G_{TX} is the gain of the transmit antenna, λ is the wavelength, and r is the distance between the antennas. The original Friis transmission equation is valid for free space environments with a value of 2 for the parameter α . It is also used to approximate the average received power in multi-path environments inside buildings, in which case the parameter α varies from 1.8 to 5.2 and is higher for higher frequencies [2] because of reduced transmission through typical walls.

By combining equation (3) and (4), the achievable data rate of a system can be expressed as function of bandwidth and frequency (5):

$$R \leq BW \log_2 \left(1 + \frac{P_{TX} G_{RX} G_{TX} \left(\frac{\lambda}{4\pi r} \right)^\alpha}{k T BW} \right) \quad (5)$$

The impact of frequency and bandwidth on the achievable data rate is shown in Figure 3 for a system with $r=10m$, $P_{TX}=0.1W$, and half wavelength dipole antennas in free space ($\alpha=2$). This figure shows the achievable data rate as a function of frequency and bandwidth at a distance of 10m with a transmit power of 100mW. The figure shows that data rates in excess of 10Gbps can be achieved for high bandwidths (1GHz) at low frequencies (1GHz).

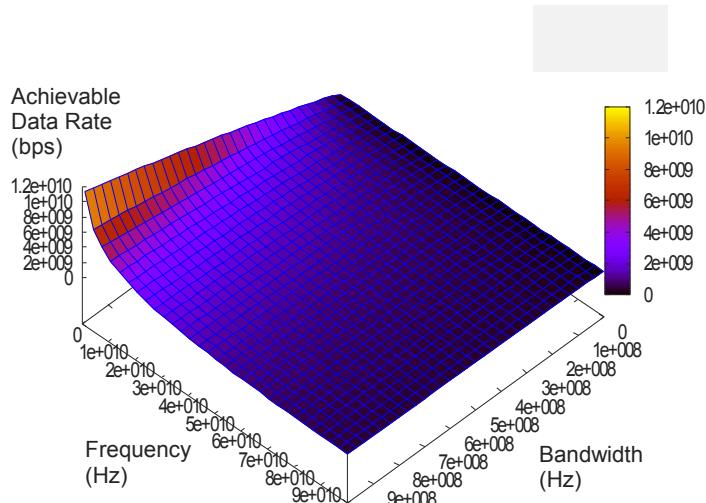


Figure 3: Achievable data rate versus frequency and bandwidth with half wavelength dipole antennas

The shape of the graph is caused by the different influences of bandwidth and SNR (and therefore indirectly frequency) on the channel capacity. Increasing bandwidth might seem like an obvious way to improve the channel capacity, but it of course also increases the noise in the channel and therefore reduces the signal-to-noise ratio at a fixed signal level. Therefore, increasing bandwidth makes sense only if the SNR is sufficiently high: for small SNR, the channel capacity is independent of the bandwidth. Therefore, going to high bandwidths at high frequencies only makes sense if the received SNR is also high.

Since in in-house environments alpha is a function of frequency, the optimum at low frequencies will be even more pronounced than shown in Figure 3. This, together with the higher transparency of walls at lower frequencies and the simpler and cheaper electronics, explains the popularity of relatively low frequencies for radio communication.

However, this inherently leads to a conflict: if all high data rate applications would prefer to use a lot of bandwidth at low frequencies, then the radio spectrum at low frequencies would quickly fill up – which it indeed does. This results in a drive towards higher frequencies, since there will be more (cheap) bandwidth available than at lower frequencies. In addition, the decrease in data rate when increasing the frequencies as shown in Figure 3 is somewhat deceptive, in that it is caused by the decrease in antenna size at higher frequencies. If we would keep the physical antenna size the same, there would be no decrease in data rate with higher frequencies, and the achievable data rate increases significantly with the bandwidth again, as shown in Figure 4.

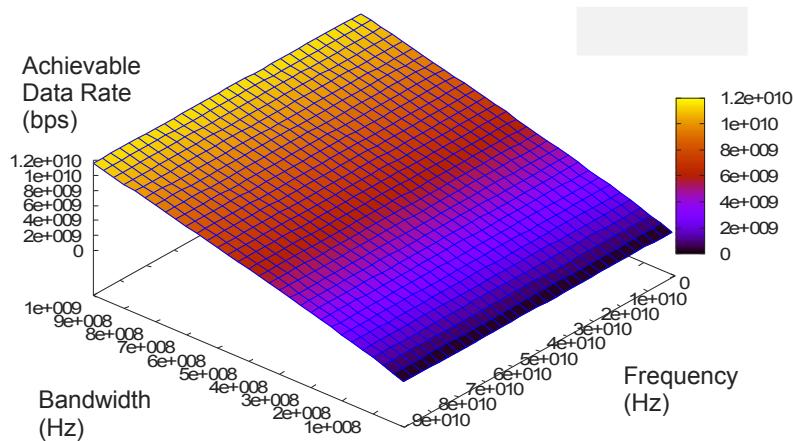


Figure 4: Achievable data rate versus frequency and bandwidth
for fixed size antennas

Please note that this is only true for line-of-sight radio transmission in empty space ($\alpha=2$). In indoor environments with higher values for α , there is a reduction with frequency even with fixed antenna sizes. However, higher values of α are a way of modeling the losses through walls. As long as a radio link exist within a single room without the need to penetrate walls, the data rate at high frequencies is still close to Figure 4.

Therefore, high data rate radio links with high bandwidths at high frequencies only make sense when electrically large antennas ($>> \lambda/2$) are used. Such physically large antennas provide antenna gain and directivity. For fixed links (e.g. LMDS), this can be implemented as a single antenna that is mechanically aligned towards the antenna on the opposite side of the radio link. For mobile links, the alignment of the main lobe needs to be achieved dynamically, usually through adaptive phased array structures. Especially in mobile systems, the performance of the phased array (the combination of the antennas themselves and the transceiver electronics) can be significantly influenced by the environment, e.g. when the antennas are in close proximity to objects such as walls, furniture, or hands of people. Since most very high frequency systems will critically depend on good performance of the phased array, adaptivity of such a phased array needs to be extended to include also proximity effects. Although at very high frequencies, objects need to be quite close to influence the antenna impedance [8], coupling between elements of a phased array can change significantly in the presence of objects near the array.

Systems that do not require high data rates and/or that need to penetrate walls are likely to stay at lower frequencies, both because of the simpler design and lower cost of the transceiver and because of the better transparency of walls and other objects.

2.2 Capacity, security and privacy

Although the relatively high opacity of walls at high frequencies can be a limitation for radio links, it also provides advantages such as a higher system capacity, since the same radio spectrum can be reused at shorter distances. Further increase of the system capacity can be achieved by exploiting the spatial dimension even more through the use of narrow radio beams created by phased array systems. This will reduce interference between different links, and therefore increase system capacity.

Higher inherent security and privacy is also easier to achieve at higher frequencies, both because the radio signal range is limited to mostly a single room, and because the signal can be confined to a narrow beam relatively easily. When a signal is difficult to pick up outside of the narrow beam and outside the room, there is already an inherent privacy protection in the system that does not

depend on correct implementations of unbreakable encryption algorithms, and is therefore not susceptible to eavesdropping from a distance.

Similarly, a receiver that is only sensitive to signals originating in a narrow beam within a single room is inherently more secure against attacks from a distance, without having to rely on correct implementations of unbreakable identity validation algorithms.

These advantages are partially offset by the need to have individual access points for every room, which translates into a higher investment and installation cost. Nevertheless, this can be an attractive trade-off depending on the value of system capacity, security and privacy for a system.

2.3 Spatial resolution

Very high frequencies allow a proportionally better spatial resolution, which can be used for active and passive imaging as well as radar applications. Passive imaging uses natural thermal emissions of objects in the 35GHz and 94GHz bands to reconstruct an image, whereas in active imaging the system transmits mm-wave signals to “illuminate” objects [5]. Currently, there is a lot of interest in these types of imaging because of security applications, e.g. screening of people for concealed weapons, but there are other interesting applications in e.g. medical, safety and testing areas, as well as in the more traditional areas of radio astronomy and space-born radios (around 95GHz).

Radar with high resolution is relevant for e.g. automotive (anti-collision and adaptive cruise control, [6]) in the 24GHz and 77GHz bands, and autonomous robot guidance [7] (terrain maps).

2.4 Antenna arrays and antenna integration

The short wavelengths allow the monolithic integration of relatively cheap antennas [9]. At 60GHz, the wavelength in vacuum is only 5mm, so a half-wavelength dipole antenna would only be 2.5mm long in vacuum. In a silicon technology the physical length of a dipole will be decreased, depending on the dielectric properties of the inter-metal and substrate materials. In a mainstream silicon IC process, the cost for integrating such an antenna can then be less than \$0.10.

There are of course limitations when using integrated antennas, such as:

- losses in the silicon substrate;
- losses due to the metalization and inter-metal materials;
- limited flexibility in distance of the antenna to the ground plane;

- limitations in the packaging, which has now to be transparent at the relevant frequencies;
- limitations in the placement of the IC.

Nevertheless, there are also significant advantages:

- avoiding cost and performance loss in bringing RF signals off-chip through bondpads, bumps and package
- avoiding cost, performance loss and potential reliability problems because ESD protection for the RF signals can be (partially) avoided;
- the dimensions and relative positions of the antennas can be accurately designed, and are known at design time so the transceiver can be optimized for the antenna properties.

2.5 Consequences for beam steering

To exploit the advantages listed in the preceding sections 2.1 to 2.3, antennas with narrow beams and high gain are required, and, with the exception of stationary line-of-sight links, these properties of these beams need to be adaptive to the position of the transceivers and to the environment. Therefore, adaptive beam steering is going to be an essential part of most very high frequency radio links. This will impact both architectural and circuit level requirements for very high frequency transceivers.

3 System level, architectural and circuit level requirements

Although they are in many aspects similar to transceiver architectures at lower frequencies, transceiver architectures for monolithically integrated transceivers at very high frequencies have to meet several different boundary conditions and requirements.

3.1 System design

As discussed in section 2.1, one of the main drivers for implementation radio links at very high frequencies is the availability of empty bands that allow the use of wide bandwidth transmissions to achieve high data rates, as long as a sufficiently high signal-to-noise ratio can be achieved. One way to relax the requirements on signal-to-noise ratio is the use of less bandwidth-efficient modulation schemes. Since the performance of RF circuits at very high frequencies is limited, constant envelope modulation schemes such as (G)FSK multi-level (G)FSK and constant-envelope phase modulation, become an attractive approach. With constant envelope signals, the linearity requirements

for large parts of the receiver and transmitter can be reduced significantly, thereby reducing cost and implementation risks. However, increasing the symbol rate and channel bandwidth will require higher performance from the data converter and channel equalizer. Current standardization for radio links in this frequency range are not (yet) taking this approach: rather, they are based on approaches that are also used for lower frequency systems. Nevertheless, there investigations into, and proposals for modulation methods that eliminate the need for an equalizer have started to appear [12].

A time domain multiple-access scheme would fit best with simple modulation schemes. It reduces interference from adjacent and alternate channels that would occur in frequency division multiple access schemes, and easily allows flexible and on-demand allocation of total system capacity across multiple sources.

3.2 Receiver architectures

Channel bandwidths for very high frequency systems can be significantly above 1GHz. As a consequence, IF frequencies in a very high frequency receiver are often similar to the RF frequencies of many current cellular and connectivity standards. As a consequence, super-heterodyne receiver architectures are sometimes proposed with as a first stage a down converter from very high frequencies to a first IF in the 1GHz to 10GHz range [10]. Although this might seem like a low-risk architecture, there are several drawbacks to this approach:

- the integration level is lower than for zero-IF or low-IF architectures because of the (usually external) IF filter;
- an image-reject filter is needed at the input to protect the receiver against interferers at image frequencies;
- since the power dissipation of an ADC that operates immediately at this IF frequency is usually prohibitive, a second down conversion (or sub-sampling ADC) is required, adding further power dissipation and chip area.

Especially for systems with constant envelope modulation, a direct-conversion receiver [11] offers a better cost/performance trade-off as well as a higher integration level.

A very high frequency receiver will usually not suffer from very strong interferers, because walls will attenuate signals from unrelated systems significantly, and signals originating in the same room are likely to be part of the same communication system, in which case the higher layers of the communication link can avoid interference by separating such signals in the frequency and/or time domain. Therefore, only limited channel selectivity will be needed, and no image filtering (only band selectivity). After the (limited) channel selectivity, the signal can be processed through (strongly) non-linear

circuits such as limiters to provide the required gain and automatic gain control. This does require a combination of symbol rate and delay spread that achieves low bit-error rates without a channel equalizer for the environment in which the system is intended to work. This requirement is easier to achieve at very high frequencies because the delay spread is inherently lower, and is further reduced when high-gain/narrow-beam antennas (or antenna arrays) are used. In that case, the requirements on the resolution of the ADC converter are now limited to 1 bit only – in effect, the ADC can be replaced by a sampling master-slave flip-flop, allowing low-power conversion of signals with very wide bandwidths.

Circuit requirements for such a receiver will typically emphasize gain at very high frequencies, wide bandwidths, and low noise figures, with moderate requirements for third-order linearity and 1dB compression. For a zero-IF, the second-order non-linearity will of course need attention, but because of the limited interference provided by the higher layers, and with a carefully designed modulation scheme, it is usually possible to achieve the desired performance by careful design of the mixers and AC coupling in the IF path.

Requirements on the phase noise of the VCO are likely to be relaxed as well, since the wide channel bandwidths puts the adjacent channel (for reciprocal mixing) at a large frequency offset. In addition, if the system is completely TDMA based, and therefore effectively single-channel, requirements on the tuning range for the VCO will be relaxed since only process spread, temperature and power supply variations need to be compensated. It will even be possible to clean up the phase noise of the VCO across the full channel with a wide-band synthesizer loop.

3.3 Transmitter architectures

Transmitters for very high frequency systems need to generate wide-band signals. Since in many cases, bandwidth efficiency is not the primary design parameter, and since systems do not have to be dimensioned for minimum interference, requirements on dynamic range and EVM are likely to be relaxed. Therefore, emphasis for most transmitter circuits will be on gain and power at high frequencies and wide bandwidths. For beam steering transmitters, there will be additional requirements on stability and linearity of the power amplifier output stage because the coupling between the antenna elements will cause signals from adjacent elements to feed back into the output of the transmitter.

For constant envelope systems, transmit architectures in which the oscillator itself is modulated with the frequency/phase information are attractive. For very low-cost systems with relaxed specifications direct modulation of a free-running VCO could be considered, similar to the transmitter architectures used in early DECT and BlueTooth systems. Systems with higher performance requirements

could use closed-loop architectures such as offset-loops and 2-point modulated synthesizers. Since very high radio frequency radio systems will usually have wide bandwidth channels, pulling of the oscillator is likely to be manageable, although of course the crosstalk might also be higher. However, such closed-loop modulation schemes will prohibit the sharing of the synthesizer in full-duplex systems.

When beam steering transmitters are used, special attention has to be paid to phase consistency between the branches. In this case a common, unmodulated VCO/synthesizer with a classical up-conversion transmitter might be both more cost-effective and robust, especially for single-chip integration of all transmitters of the array. When transmitters are instead physically integrated with the antenna element, and therefore not on the same die, the potential cross-talk and parasitic coupling between individually modulated VCOs is unlikely to be a serious issue, and closed-loop concepts are the more obvious choice, also in order to avoid distributing high frequency signals between the individual transmitters.

3.4 Beam steering architectures

Beam steering technology is well known from applications at lower frequencies, and it might seem obvious at first to use similar architectures and circuits for implementing this function at high frequencies. However, this is not necessarily the most optimum approach. Figure 5 shows a block diagram of a typical beam steering transceiver for the low GHz range, with phase shifting carried out in the digital domain (for simplicity, only the receive path is shown for a system with just 2 antennas):

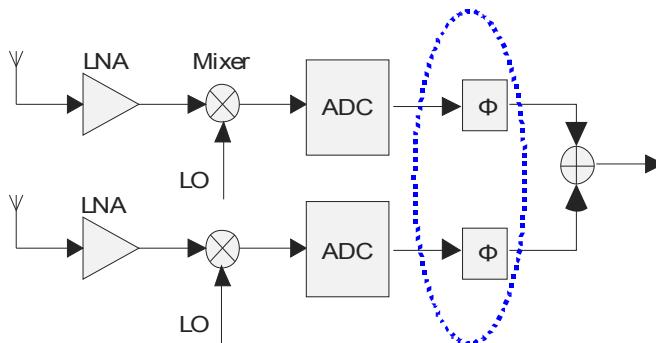


Figure 5: Block diagram of typical beam steering architecture at lower frequencies

Phase shifting in the digital domain is often used because it offers several advantages:

- high flexibility
- high accuracy
- relatively easy to design
- robust against process, temperature and supply voltage variations

However, this architecture has several disadvantages at high frequencies:

- the IF bandwidth of a very high frequency transceiver will usually be much higher than at lower frequencies, making the phase shifting and adding operation non-obvious to design, and potentially power-hungry;
- the RF/IF signal path, including the data converters, has to be implemented multiple times (once for every antenna), typically increasing cost;
- interference cancellation only occurs after the adder in the digital domain. Consequently, all circuits before that adder need to provide sufficient dynamic range to process these interferers without degrading the signal through desensitization, blocking, cross-modulation or detection. This will increase the difficulty of the design of the RF/IF circuits and data converters, as well as increase the power dissipation of these circuits.

Therefore, in most cases it will be attractive to move the signal combining operation to the left towards the antenna. Various architectures can be considered for this, as shown in Figure 6, Figure 7 and Figure 8.

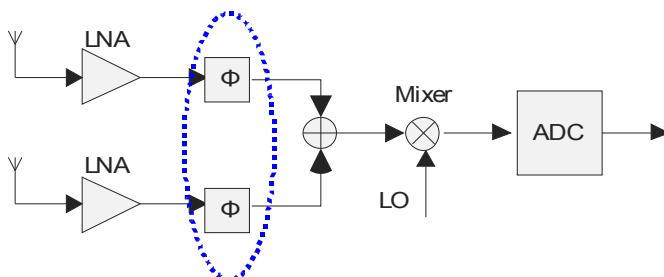


Figure 6: Beam steering combining at RF

In Figure 6, the combining of the antenna signals for beam steering is carried out at RF. This could be done immediately after the antennas, but the programmable phase shifters at these frequencies will typically have significant losses, and

therefore a better compromise is usually to insert them between the LNAs and the mixer.

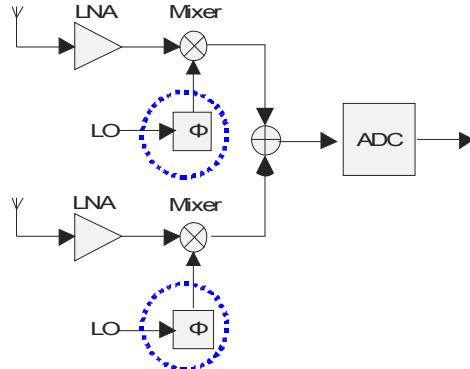


Figure 7: Beam steering
in the LO path

Figure 7 shows an architecture where the phase shifting is accomplished through phase shifting in the LO path in combination with combining at IF. This requires multiplication of more circuits, but has the advantage that combining of signals at IF is easier to implement. Also, the phase shifting in this architecture is not in the signal path, making the total performance less sensitive to the losses of the phase shifters (since they can be compensated for by generating more LO power). Finally, the phase shifter only needs to operate within a relatively narrow bandwidth (compared to the center frequency), making it relatively easy to implement.

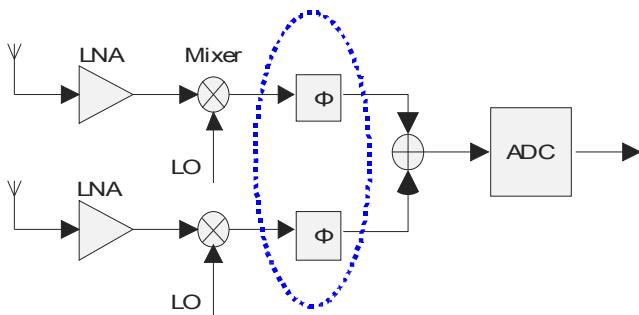


Figure 8: Beam steering
combining at IF

In Figure 8 an architecture with phase shifting and combining at IF is shown, which has the advantage that both operations now occur at lower frequencies (although still in the analog domain). This allows for easier and less critical implementation, but requires a relatively broadband analog phase shifter.

Obviously, the most efficient implementation, both in terms of cost and power dissipation, is the architecture shown in Figure 6. The main challenge is the design of a phase shifter with good performance and reconfigurability at very high frequencies.

The requirements for such a phase shifter depend, among others, on the step size of the phase shift that needs to be achieved. In the remainder of this paper, an 8-path phased-array transceiver employing shaped QPSK modulation ($\alpha=0.5$) is assumed with an RF carrier frequency of 60GHz and a bandwidth of 7.5GHz. Omnidirectional antennas are used with an antenna spacing of 0.5 wavelength.

A first indication of the step size is derived from simulation results of the EVM as a function of the incident angle with various step sizes. The result is shown in Figure 9. Using a continuous phase shifter, the peak EVM is 0.7% at incident angle of 90° due to the approximation of a uniform delay (linear phase) to a constant phase shift, since the delay is the largest here. Using a discrete phase shifter, the phase shift can compensate the carrier phase shift exactly at only a few incident angles. For other angles, the signal constellation at each received path is rotated by a different (and incorrect) phase shift, thus the signals are not added coherently at the output [13]. The peak happens at places where the phase shift errors are the largest. Using a 4 bit phase shift (step size of 22.5 degree) the peak EVM is about 5% at the incident angle of 70 degree.

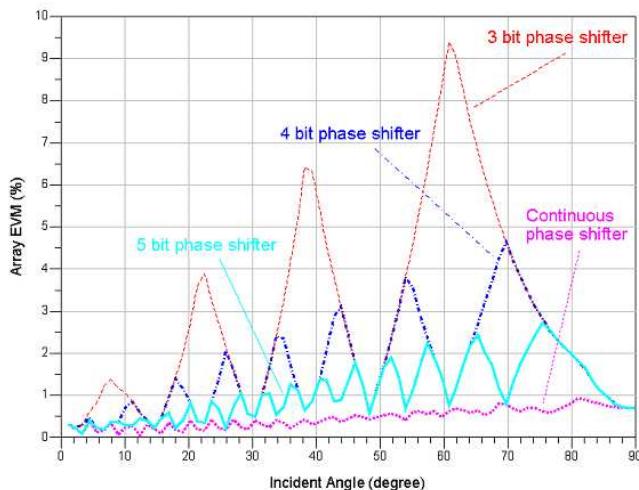


Figure 9: Output EVM ($\sim 1/\text{SNR}$) vs. incident angle

The simulated array pattern with a 4 bit phase shifter is shown in Figure 10. Relative phase shifts of $0, \phi, 2\phi, \dots, 7\phi$ are implemented in each of the eight paths respectively. By increasing the incremental phase shift ϕ from 0 to 180 degree in a 22.5 degree step size, the beam direction can be steered from 0 to 90 degree in nine patterns. From this figure it can be seen that a 4 bit phase-shift resolution is sufficient to radiate at all angles at close to peak array gain. In the worst case, the signal loss is less than 1dB.

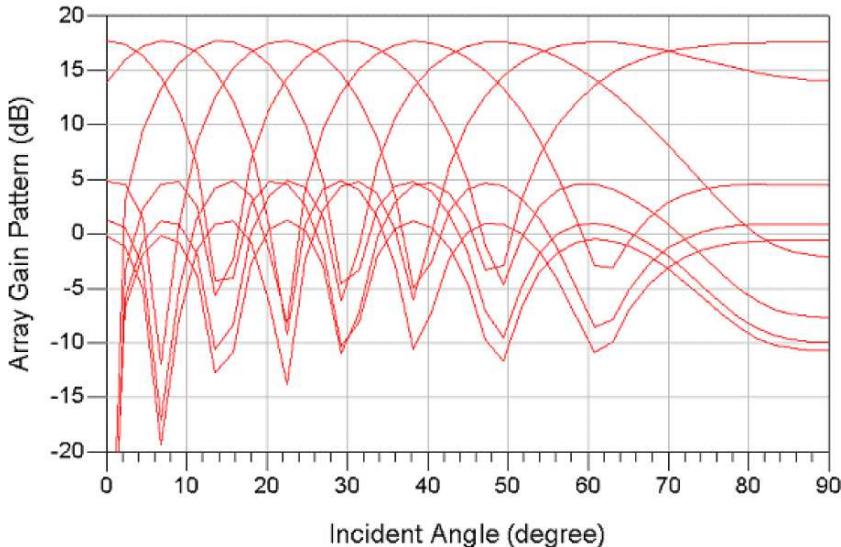


Figure 10: Array pattern with 4-bit phase-shift resolution

3.5 Beam steering implementations

From both results it is clear that a 4 bit phase shifter is very close to the ideal, continuous phase shifter for this application. Such a phase shifter can be readily implemented in modern mainstream silicon CMOS technologies. For relatively wide-band systems, the circuit has to be optimized for constant group delay by tuning the phase shift versus frequency characteristics of the phase shifter. This can be achieved by using a switched line or loaded line structure, although special attention is needed to minimize the impact of the parasitics of the (less than ideal) switches on the performance of the phase shifter.

A better solution is possible by using a series-tuned phase shifter, as shown in Figure 11, consisting of a number of non-ideal switches S0, S1 and S2, and a

number of transmission lines. In this phase shifter, the parasitics of the switches are balanced with transmission line segments that have characteristic impedances different from the source and load impedance of the phase shifter.

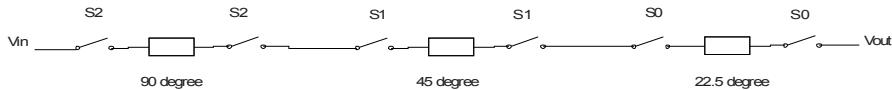


Figure 11: Series-tuned phase shifter

The performance of such a series-tuned phase shifter in a mainstream CMOS technology is shown in Figure 12. The values for the markers m6 and m7 are shown at the bottom right.

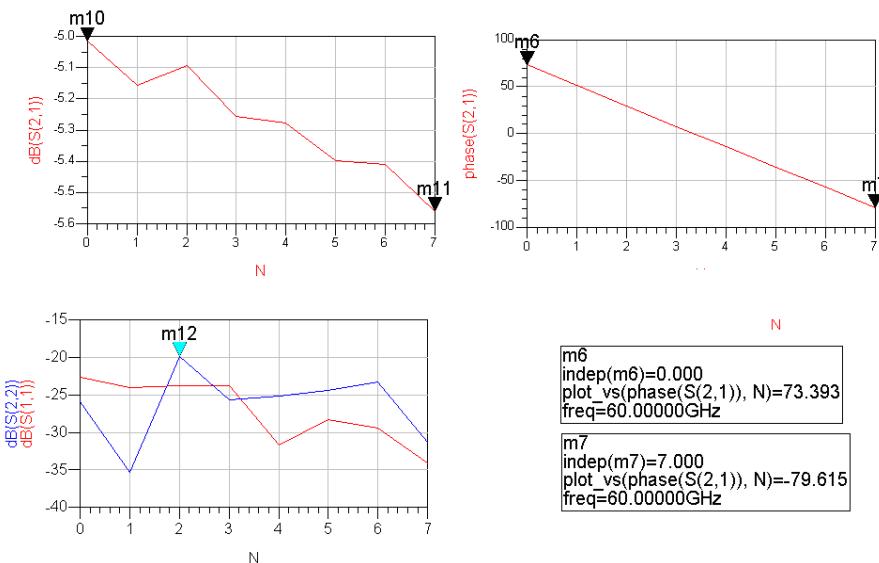


Figure 12: Insertion loss (top-left), phase shift (top-right) and input/output matching (bottom-left) performance of a series-tuned phase shifter

Such a phase shifter meets the requirements for beam steering signal processing at RF for transceivers operating at very high frequencies (in this case 60GHz), and enables the preferred architecture shown in Figure 6.

3.6 Choice of process technology

Traditionally 60 GHz radio frequency (RF) technology has been the domain of expensive chip technologies based on III-V compound materials such as Gallium Arsenide and Indium Phosphide. These technologies were mainly intended for military applications for which the cost-factor is not of much relevance. A relatively new development is the achievement of considerable RF performance with low-cost process technologies based on Silicon. With Silicon Germanium (SiGe) technology the maximum frequency of operation f_{max} amounts to hundreds of GHz and it has the best physical properties for providing sufficient RF performance. The RF performance of standard CMOS is worse but increases more rapidly due to the enormous world-wide effort to scale to lower gate-lengths which implies a higher f_{max} . The speed of analog CMOS circuits increases by roughly one order of magnitude every ten years. High power amplifiers implemented in todays 90 nm RF-CMOS technology can produce an output power level of about 6 dBm with sufficient linearity whereas low noise amplifiers with noise figure of 5 dB can be realised [15]. The CMOS Chip industry already invests massively in 65 nm technology with 45 nm as the next step promising a further increasing performance in future. This makes CMOS the lowest cost option and with its rapid performance improvement due to continuous scaling CMOS is becoming the technology of choice to address the low-cost millimeter-wave market.

4 Conclusions

Radio transceivers at very high frequencies are required to meet the needs of high data rate, high capacity, secure, private, and high spatial resolution applications. This puts special requirements on system design, architecture and circuits. To exploit the advantages of these high frequency transceivers for these applications, beam steering is essential. In contrast to beam steering at lower frequencies, the architecture and circuit requirements for such very high frequency beam steering transceivers can be met with phase shifting and recombining at RF rather than in the digital domain.

References

- [1] C.E. Shannon, "A Mathematical Theory of Communication", Bell System Technical Journal, vol. 27, pp. 379-423, 623-656, July, October, 1948
- [2] H. Hashemi, "The Indoor Radio Propagation Channel", Proceedings of the IEEE, Vol. 81, No. 7, July 1993
- [3] Smulders, P.F.M. (2003). "60 GHz radio: prospects and future directions" In proc. IEEE 10th Symposium on Communications and Vehicular Technology in the Benelux (pp. 1-8). Eindhoven, The Netherlands: IEEE.
- [4] Koichi Tsunekawa, "Recent Antenna System Technologies for Next-generation Wireless Communications", NTT Technical Review Vol. 3 No. 9, Sept. 2005
- [5] Dr. Isaiah M. Blankson, "PASSIVE MILLIMETER WAVE IMAGING WITH SUPER-RESOLUTION: Application to Aviation safety in extremely poor visibility", Presentation at Institute of Mathematics and its Applications, University of Minnesota: May 5, 2001
- [6] M. Chabert et al., "On the use of High Resolution Spectral Analysis methods in Radar Automotive", 1st International Workshop on Intelligent Transportation (WIT 2004), 23-24 March 2004, Hamburg, Germany
- [7] Clark SM et al., "Autonomous land vehicle navigation using MMW radar". 1998 IEEE International Conference on Robotics and Automation 16-20 May 1998, Leuven, Belgium.
- [8] Klohn, K. L. , "Metal walls in close proximity to a dielectric waveguide antenna", IEEE Transactions on Microwave Theory and Techniques, vol. MTT-29, Sept. 1981, pp. 962-966
- [9] Arun Natarajan et al., "A 77-GHz Phased-Array Transceiver With On-Chip Antennas in Silicon: Transmitter and Local LO-Path Phase Shifting", IEEE journal of solid-state circuits, vol. 41, no. 12, December 2006
- [10] S. Reynolds, "A 60-GHz Superheterodyne Downconversion Mixer in Silicon-Germanium Bipolar Technology", IEEE J. Solid-State Circuits, vol. 39, no. 11, pp. 2065-2068, Nov. 2004
- [11] Brian A. Floyd, "SiGe Bipolar Transceiver Circuits Operating at 60 GHz", IEEE Journal of Solid-State Circuits, Vol. 40, No. 1, January 2005
- [12] Thomas H. Williams, "Frequency Domain Reciprocal Modulation (FDRM)", 1999 IEEE Radio and Wireless Conference Proceedings, pp. 129-132

- [13] H. Hashemi et al., "A 24-GHz SiGe Phased-Array Receiver – LO Phase-Shifting Approach", IEEE Transactions on Microwave Theory and Techniques, Vol. 53, No. 2, February 2005
- [14] P. Smulders, " Exploiting the 60 GHz band for local wireless multimedia access: prospects and future directions", IEEE Commun. Mag., vol. 40, pp. 140-147, Jan. 2002.
- [15] T. Yao, M. Gordon, K. Yau, M.T. Yang, and S.P. Voinigescu, "60-GHz PA and LNA in 90-nm RF-CMOS", IEEE RFIC Symp., June 2006.
- [16] FCC, "Code of Federal Regulation, title 47 Telecommunication Chapter 1, part 15.255", Oct. 2004. (US and Canada)
- [17] Regulations for enforcement of the radio law 6-4-2, specified loe power radio (11) 59 – 66 GHz band. (Japan)
- [18] Ministry of Information and Communication of Korea, "Frequency Allocation Comment of 60 GHz Band", April 2006. (Korea)
- [19] ACMA, "Radiocommunications (Low Interference Potential Devices) Class License Variation 2005 (No. 1)", Aug. 2005. (Australia)
- [20] ETSI DTR/ERM-RM-049, "Electromagnetic compatibility and Radio spectrum Matters (ERM); System reference Document; Technical Characteristics of Multiple Gigabit Wireless Systems in the 60 GHz Range", March 2006. (Europe)
- [21] H. Yang, P.F.M. Smulders and M.H.A.J. Herben, "Channel Characteristics and Transmission Performance for Various Channel Configurations at 60 GHz", Eurasip JWCN, 2007.
- [22] K. Maruhashi, S. Kishimoto, M. Ito, K. Ohata, Y. Hamada, T. Morimoto and H. Shimawaki, "Wireless Uncompressed-HDTV-Signal Transmission System Utilizing Compact 60-GHz-band Transmitter and Receiver", Microwave Symp. Digest, IEEE MTT-S Int. June 2005.

KEY BUILDING BLOCKS FOR MILLIMETER-WAVE IC DESIGN IN BASELINE CMOS

Mihai A.T. Sanduleanu¹, Eduardo Alarcon¹, Hammad M. Cheema², Maja

Vidojkovic², Reza Mahmoudi² and Arthur van Roermund²

¹Philips Research Eindhoven, ²Eindhoven University of Technology

Abstract

In this work, new receiver concepts and key building blocks, at circuit level, for future millimeter-wave wireless communications standards are introduced. Starting from passive and active devices, trade-offs between technology, performance and circuit choices of millimeter-wave RF front-end circuits are discussed. In particular, power consumption, noise and linearity trade-offs in low-noise amplifiers, mixers, frequency dividers and oscillators are considered. The concepts derived are applied to a large class of wireless communications standards that are broadband in nature at RF and/or require a broadband IF.

1. Introduction

The rapid growth of wireless communications, for wireless local area networks (WLANs) and wireless personal area networks (WPANs) [1], has sparked interest in using silicon RF integrated circuits for operation in millimeter-wave bands [3-4]. The assignment of the 5GHz large bandwidth around 60GHz has created new opportunities for 60GHz front-end technology. While excessively high path loss in this frequency range due to oxygen absorption and obstructions such as the walls of buildings precludes long-range communications, gigabit/s point-to-point links or short-range wireless local area networks actually benefit from this attenuation; it minimizes interference with other systems, enabling frequency reuse as well as providing more security [1]. Other possibilities in the millimeter-wave frequency range are automotive applications, such as long-range (77/79GHz) radars for collision avoidance, security (94GHz) and extreme wide-band communications in the 120GHz band. We are pushing towards higher frequencies as a demand for wireless capacity. This is a consequence of Moore's law for wireless.

Although operation in these frequency bands was once the exclusive domain of III-V-compound semiconductors [2] due to their superior electron

mobility, higher breakdown voltage and semi-insulating substrate ($10^7\text{-}10^8 \Omega\text{cm}$), silicon CMOS benefits from faster technology evolution and offers cost reduction at higher integration. Currently the f_T and f_{max} demonstrated by the 90 nm, LP generation NMOS transistors, are in the order of 120GHz/280GHz respectively and the active devices are capable of 8dB power gain in the 60 GHz band. Other technology nodes, such as 65nm and 45nm, have even better RF performance with less power consumption. We are currently facing a technology push and the result of it are faster and faster devices. The passive components (inductors, transmission lines and metal-metal capacitors) are subject to scaling too and they can complement the operation of active devices by e.g. generating more gain in a small bandwidth through resonant loads. This unprecedented combination and the presence of millimeter-wave application areas are clear incentives to further research activities [5-11].

The rationale of millimeter-wave transceivers in baseline CMOS relates to integration and cost reduction. Moreover, high throughput (>4Gb/s) common for some burst data transfer requires high-speed digital signal processing. This can be achieved only in the fastest baseline CMOS processes.

There are many challenges for CMOS designers in the millimeter-wave bands. First of all, due to the high path loss of the 60GHz wireless channel, coupled with the limited noise and power-handling capabilities of conventional CMOS technologies at this frequency, renders high-SNR communication in the 60GHz band extremely difficult. Secondly, a CMOS process poses many design challenges for operation at mm-wave frequencies, due to the presence of lossy substrates. Thirdly, the influence of parasitic elements for deep submicron MOS devices in high-frequency region results in inaccurate MOSFET models.

In the following sections we are addressing the design of some key building blocks for millimeter-wave transceivers in a baseline CMOS LP process at 1.2V supply. Measured results substantiate the design procedures and the analytic models throughout the material.

2. Passive and active components

In order to achieve high gain as well as stability and reverse isolation at 60GHz, a cascode stage is chosen as the basic circuit building block for low-noise amplifiers and power amplifiers. For power amplifiers, the cascode configuration can also relax the voltage stress on each individual transistor that can lead to breakdown and failure. The measurements of the cascode stage are based on two different de-embedding techniques, presented in Fig.1. The first method relies on a short-open-load structure with a cascode device as load. The second de-embedding technique is based on ABCD parameters and accurate measurement of the bond-pads and transmission line structures. As a result, the ABCD matrix of the DUT is extracted. The measured results (see Fig.2) are

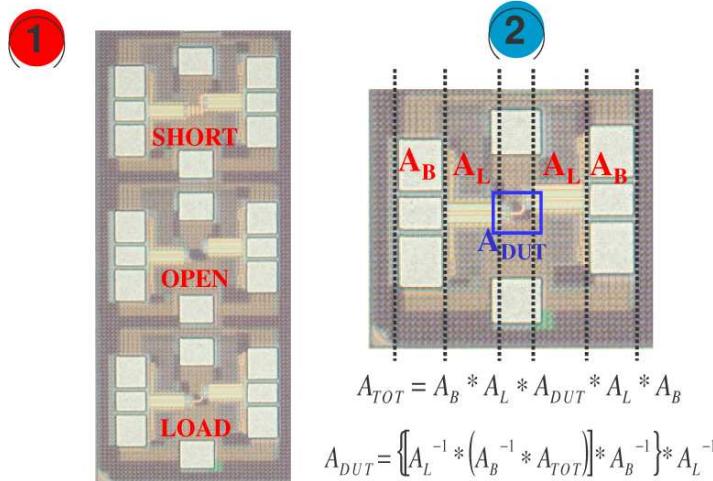


Fig.1. Cascode device characterization

almost similar for S_{11} , S_{12} , S_{22} and differ with 1-dB for S_{21} parameter. We consider the first method more reliable due to lesser number of computations needed to extract the S-parametters of the DUT. Although the silicon substrate is lossy, passive elements with good quality factors are still feasible. Compared to spiral inductors at mm-wave frequencies, which require detailed knowledge of the substrate and analysis of eddy currents, transmission lines are better suitable to realize small inductances in this frequency band.

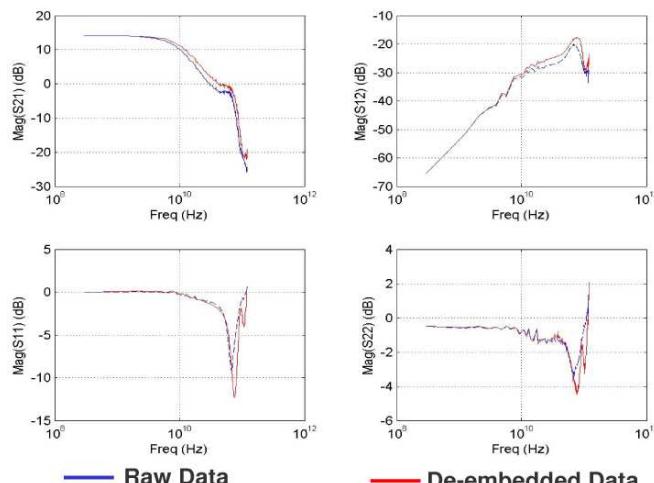


Fig.2. Cascode device: measured results

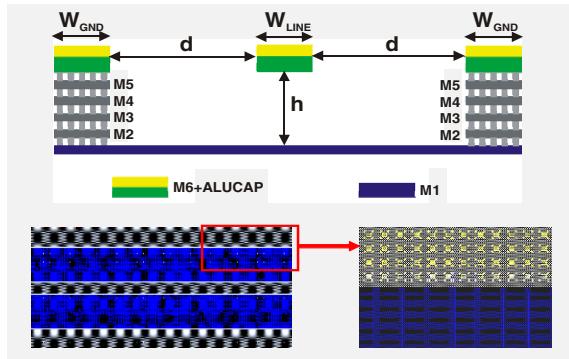


Fig.3. Thin-film, on-chip, microstrip line with side-walls

In the proposed designs, microstrip lines consisting of M1 as the ground plane and M6 as the signal line, are employed in the matching networks and resonant circuits (see Fig.3). The wide gap between signal and M6 ground lines, favors the microstrip behavior of the line instead of its CPW characteristic. For better isolation between the adjacent lines, ground strips (M6), on both sides of the signal line are considered. The distance d from the signal line to the top ground is much larger than the distance h to the bottom ground. In this structure, the top M6 ground lines are connected to M1 by many vias through all the metal stacks. This then, can provide a well-defined ground plane and offers a reasonable metal density that can satisfy the design rules for manufacturing. The process used features damascene copper for the 6 metal layers (5thin+1thick) and low-k (<3.0) inter-metal dielectric between the thin metal layers.

For de-embedding, the L-2L technique was used. Figure 4 shows the measured inductance of the transmission line as a function of length.

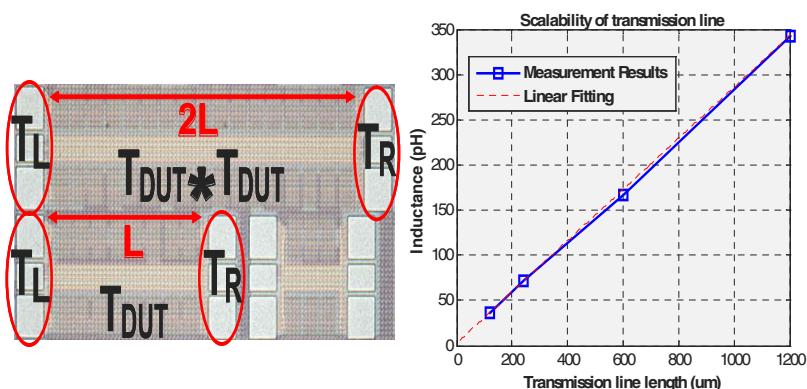


Fig.4. Microstrip-line measured results

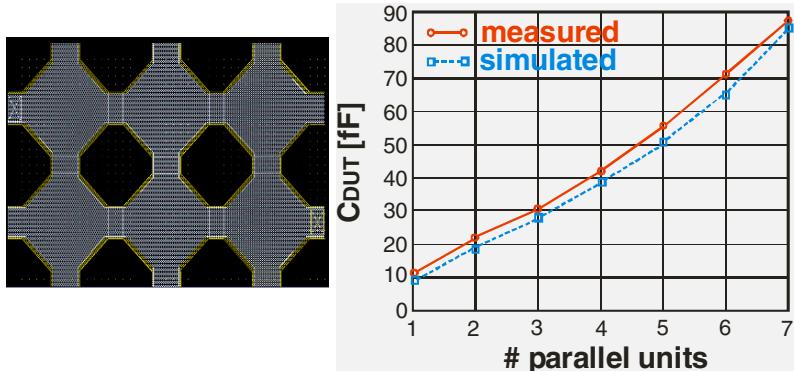


Fig.5. Me6-Me5 capacitors measurements

The measured insertion loss of the line is 0.065 Np/mm (0.52dB/mm) at 60GHz. For decoupling and resonators, Me6-Me5 capacitors are proposed. A unit cell of about 10fF is used for generating larger capacitance values. As expected, the total capacitance and the parasitics, as well, scale fairly linear with the number of cells (see Fig.5). An extra design dimension is the presence of the ALUCAP layer on top of the capacitors.

3. I/Q down-converter

A broadband wireless system capable of providing 10x capacity and operating at 10x the carrier frequency of current wireless radios will require a fundamental shift in the design of CMOS circuits and new approaches to circuit/system design. Arguably, different architectures apply for a millimeter-wave receiver. To illustrate the possible problems, an I/Q down-converter for zero-IF/near zero-IF receivers will be discussed in the next section.

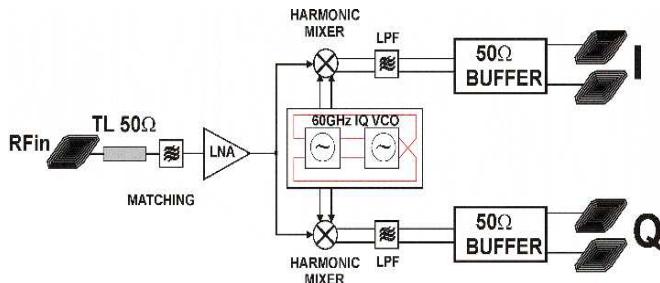


Fig.6. I/Q down-converter

The quadrature down-converter from Fig.6 requires a 60GHz quadrature oscillator and two harmonic mixers. The generation of quadrature signals with polyphase filters or frequency division by 2 with a static frequency divider was excluded after an in-depth analysis. In the first case, the impedance level would be low, to counteract parasitic effects. Therefore, low output impedance buffers are needed, rendering this solution as difficult or too power hungry. A static frequency divider working at 120GHz is again excluded due to low transit times needed in the devices at such operating speed. As we will see, frequency dividers operating at 60GHz are yet in a research phase. Therefore, the whole concept should be devised after solving the LO generation problem. This will be explained in the next chapters.

3.1. Low-noise amplifiers

For large power gains, a multi-stage design is required. A two-stage cascode LNA is presented in Fig.7. The two stages are identical and have 50Ω inputs/outputs. Although the common gate configuration provides a wideband input match, considering the relatively larger noise and lower power gain, this topology is ruled out as the input stage. Inductive degeneration, common in a low-GHz design, decreases the effective transconductance and therefore, the power gain. In our approach, the cascode transistor provides reverse isolation and ensures unconditional stability. The input matching is realized with a transmission line L1 with a length of $155\mu\text{m}$ shunted to ground with a 5pF capacitor. The output resonator consists of a transmission line L3 with a capacitive divider (C_1 and C_2) for 50Ω output match. The bias current I_{BIAS} controls the operating points of the cascode stages. When CMOS dimensions scale down, the gate resistance (R_G) sets the lower bound of the minimum noise figure. Increasing the folding factor is an effective way to reduce the noise. However, when the folding factor exceeds a value of 40, the noise from R_G has insignificant contributions to the total noise. The pole introduced by the cascode transistor depends on the transconductance of M2 (M4) and the parasitic capacitance at the cascode node. As a consequence, at high frequencies, the power gain decreases and the noise contribution of the cascode transistor is enhanced. Thus, the input referred noise of a cascode stage will rise considerably. A large cascode transistor can be tolerated if an inductor tunes out the capacitance at the cascode node. The transmission line of $174\mu\text{m}$ length acts as a coil, improving the inter-stage matching, by tuning away the effect of the parasitic capacitance at the source node. This boosts the gain of the stage by almost 20%. In the traditional cascode stage, the noise contribution of M2 increases from 1% at low-GHz range to 32%, at 60GHz. When the resonant inductor is used, the noise contribution of the cascode device is reduced to 4.6% at 60GHz. The CMOS LNA with inter-stage gain boosting has been realized in

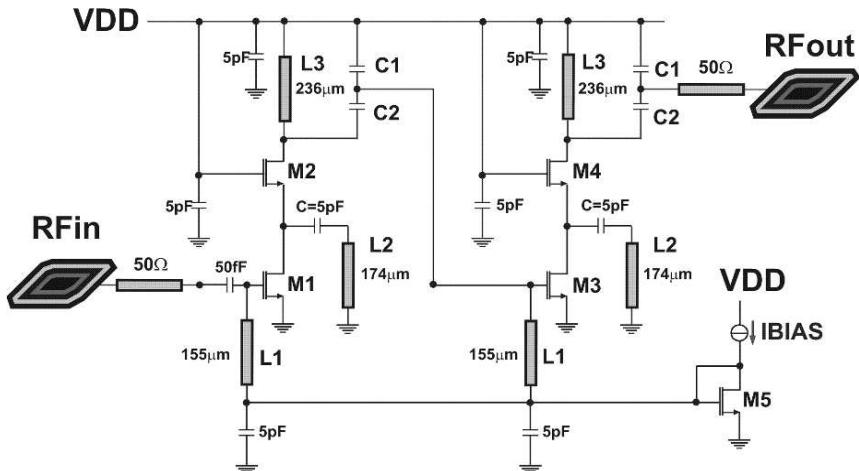


Fig.7. Two-stage cascode LNA with gain-boosting

a digital, 90-nm CMOS process. The power consumption of the two-stage LNA with gain-boosting is 10mW from a 1.2V supply ($\pm 10\%$) and the total area is 0.856mm^2 . Figure 8 shows the chip photomicrograph. The measured power gain and the noise figure of the LNA is presented in Fig.9. The noise figure (NF) is close to 5dB in the 59-61GHz band. In the same frequency band, the power gain is better than 19dB.

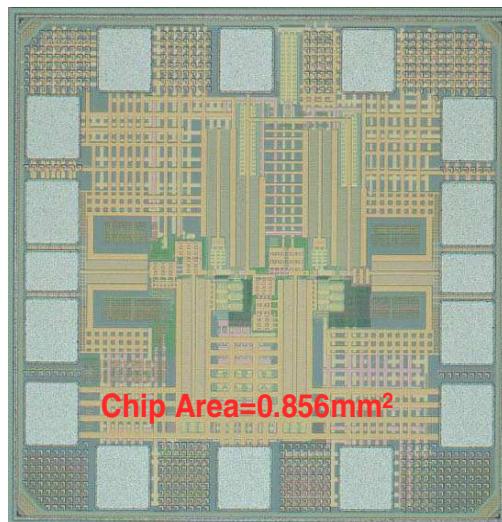


Fig.8. Two-stage cascode LNA: chip photo

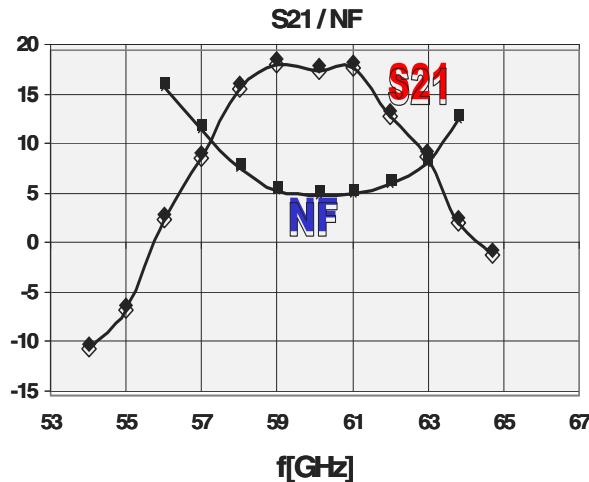


Fig.9. Two-stage cascode LNA: measured results

3.2. Harmonic/subharmonic mixers

The down-conversion I/Q mixer employed in the receiver path translates the incoming 60 GHz RF signal, amplified by the LNA, to zero intermediate frequency. In the overall mixer design, power gain, linearity, single-sideband (SSB) noise figure and power consumption are important design parameters. These parameters are not easy to achieve simultaneously, especially at 60 GHz. Usually, a double-balanced Gilbert mixer is used as a down-converter, for it generates less even-order distortion and has good LO-IF isolation. However, in our single-ended design, it is quite difficult to generate differential input signals for the mixer without phase mismatch, which may lead to even-order distortion. In addition, for a given power dissipation, a single-balanced mixer has less input referred noise than the double balanced one. Therefore, a single-balanced mixer is used in this work.

The harmonic, singly-balanced mixer from Fig.10 can be operated in harmonic or sub-harmonic mode (see Fig.11) depending on the phase of the LO signals present at the gates of the switching pairs. The transistor M_1 represents the input transconductor, converting the input RF voltage signal into a current signal. Resistors are used as output loads for zero-IF/near zero-IF conversion. The two upper transistors operate in the differential configuration and are driven by the LO signals switching the transistors from low to high g_m states. Accordingly, a current from the transconductor is delivered to the differential branches and converted to the desired frequency. The way of connecting the transconductor, the switching core and the load resistors, poses difficulties for keeping all transistors in their saturation region at 1.2V voltage supply.

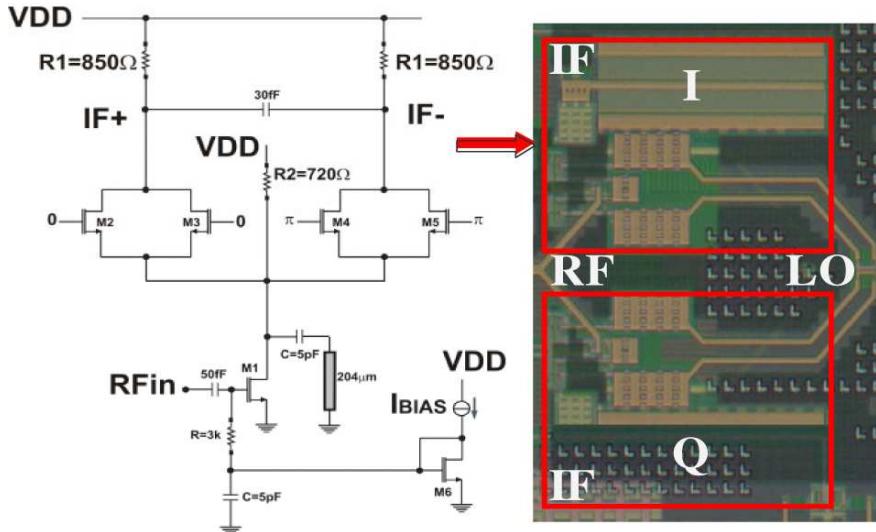


Fig.10. Harmonic single-balanced mixer

Among the intermodulation caused by the nonlinearity, the second and third order intermodulation products are of our concern; they could generate sideproducts in the IF or at baseband. The conversion gain can be roughly estimated from the conversion gain of the transconductor and the output impedance. If the switching LO signal is given by an ideal square wave with amplitude of 1, neglecting higher order terms and the up-converted sideband present at the output, we get the voltage conversion gain:

$$G_{conversion} = \frac{2}{\pi} g_m R_{load} \quad (1)$$

and the *LO* feedthrough signal at the output:

$$V_{LO_leakage} = \frac{4}{\pi} R_{load} I_{tail} \cos(\omega_{LO} t) \quad (2)$$

Therefore, better linearity and higher gain can be achieved by increasing the bias current of the transconductor. The price paid is an increase in power consumption. Furthermore, a large current through the switching pair causes voltage headroom problems and requires a larger LO drive voltage. A possible solution is current bleeding. In order to reduce the noise contribution from the bleeding path, a resistor instead of using active devices generates the bleeding current. Current bleeding allows the independent control of the DC current for the switching pair decreasing the drain current of the switches without transconductor performance degradation. This results in a higher conversion

gain by using higher load resistors. However, this approach decreases the transconductance of the switches, increasing the impedance at the source terminals of the switches. This causes more RF current flow into the parasitic capacitor at the source node. Combining the advantages of these two solutions, an inductor at the source node is used to resonate with the parasitic capacitance, allowing most of the current produced by M1 to flow to the switches. A conversion power gain of 4dB and a noise figure of 8.6dB are possible with this configuration. The same circuit can be operated in a sub-harmonic mode as suggested by Fig.11. For better understanding, the four LO phases at 30GHz ($\omega_{RF}/2$) are sketched. Let us assume that switching transistors M₂... M₅ are switched on/off completely by LO signals. The currents flowing in the switching pairs are given by:

$$i_2(t) + i_3(t) = \frac{I_{BIAS}(\pi - 2)}{\pi} + 4 \frac{I_{BIAS}}{\pi} \sum_{k=1}^{\infty} \frac{1}{4k^2 - 1} * \cos(2k\omega_{LO}t) \quad (3)$$

$$i_4(t) + i_5(t) = \frac{I_{BIAS}(\pi - 2)}{\pi} - 4 \frac{I_{BIAS}}{\pi} \sum_{k=1}^{\infty} \frac{1}{4k^2 - 1} * \cos(2k\omega_{LO}t) \quad (4)$$

As expected, these two currents are in anti-phase and obviously, the second LO harmonic will mix with the RF signal to generate beat-tones at IF. Compared to harmonic mixer a sub-harmonic mixer has little self-mixing problem, but it has lower conversion gain and poorer noise figure.

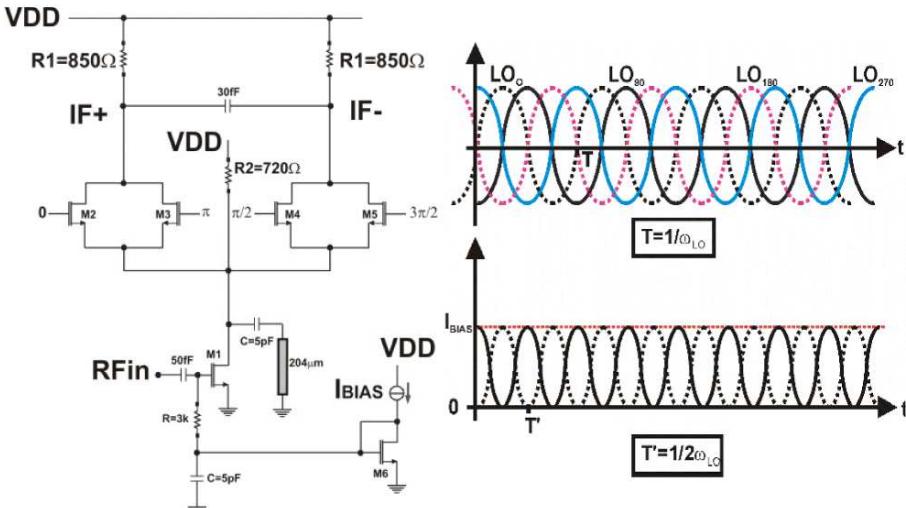


Fig.11. Sub-harmonic single-balanced mixer

3.3. Frequency dividers

In a frequency tuning system the first stage of the frequency divider is a critical component due to requirements of high frequency, wide bandwidth and high input sensitivity. The commonly used high frequency divider architectures are static, Miller and injection locked. Although, the latter two configurations can achieve high frequencies with low power consumption, they are inherently narrow band. Thus, they are not entirely suitable for the 7GHz bandwidth required worldwide for the 60GHz band. In comparison, static frequency dividers are broadband but consume more power. In this section we present an improved static frequency divider with broadband operation. As a key enabling component, a compact, stack inductor, was used in the latch and output buffers. The output buffers, based on f_T doublers are included for measurement purposes.

The block diagram of a conventional 2:1 static frequency divider is shown in Fig.12. It consists of two cascaded (master and slave) D-latches. Anti-phase clock pulses drive the D-latches and the dividing operation is achieved by connecting the inverted slave outputs to master D-latch inputs. Each D-latch is based on MOS current mode logic (MCML) and consists of data transistors: M1, M2 and latch transistors: M3, M4 as shown in Fig.13. The MCML logic is characterized by small voltage swings, thus high switching speed, constant power consumption and noise immunity due to complete differential architecture. The maximum operating frequency of the divider depends on the speed of CML latches, which in turn is limited by technology, parasitic capacitances of devices and the layout parasitics. In this design, the effect of the latter two has been minimized by circuit optimization of the D-latch and optimal layout of the divider. In comparison to earlier published designs, which employ similar dimensions for the data and latch transistors, this design has different dimensions for the same, as sketched in Fig.13.

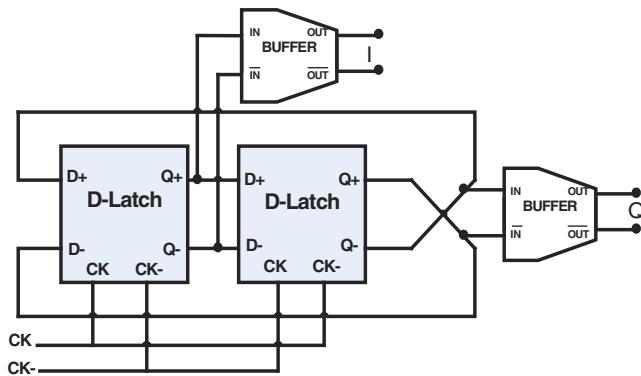


Fig.12. Static frequency divider with output buffers

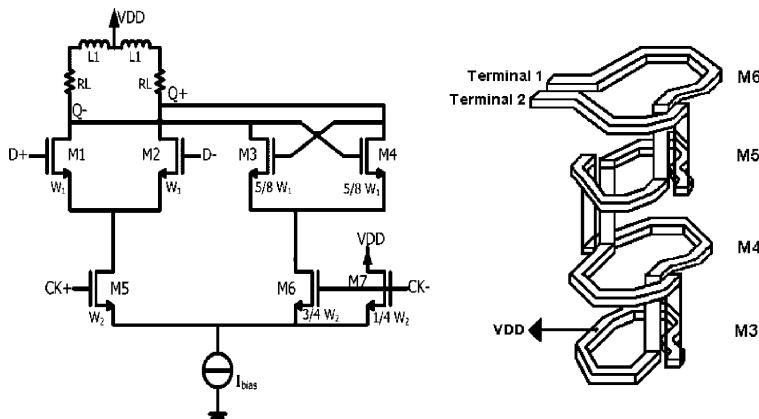


Fig.13. MCML D-latch for static divider

The positive feedback between M3 and M4 implies that only a small current is required by the latch part to take a hard decision between high and low level of the output. Parametric simulations reveal that the optimum width of transistors M3 and M4 is 5/8 times the width of M1 (or M2). In addition, smaller latch transistors offer less parasitic capacitances at the output nodes of M1 and M2 and thus enhance the switching speed of the charge transfer between them. As the latch transistors do not require the same current as the data transistors, selecting different dimensions for transistors M6 and M7 controls the current distribution. The latter transistor is added to maintain the differential structure for the clock inputs and avoid any imbalance. The current consumption in the data and latch transistors is 12mA and 9mA respectively. The optimal dimensions for transistor M6 and M7 are shown in Fig.13. The shunt inductors L₁, in series with R_L resistors, are included for broadband operation.

The shunt peaking inductor (L₁) is implemented in a differential structure with the center point connected to VDD and its value is 125pH. A high quality factor Q is not required as the load resistor connected in series with the inductor determines the effective Q. By realizing it as a stacked inductor using the top four metal layers, the area of the inductor is minimized. Metal 1 is used as a fish bone structure below the inductor for better isolation.

Compared to planar inductors, this structure offers less (up to 45%) parasitic capacitance. Two consecutive layers (e.g. M6 and M5) are displaced by one metal width by which, the parasitic capacitance is present only between even layers (M6, M4) and odd layers (M5, M3). The metal to substrate capacitance is also limited to M3. Assuming the metal lengths of each half turn of the inductor as l_1, l_2, \dots, l_{2n} , the metal width as W , metal to metal overlap capacitance as $C_{m-m}(k)$ and metal to substrate capacitance as $C_{m-s}(k)$ the total parasitic capacitance is estimated by the distributed capacitance model using (5) and (6):

$$d_k = \frac{\sum_{i=1}^k l_i}{\sum_{i=1}^{2n} l_i} \quad (5)$$

$$\begin{aligned} C_{total} \cong & \frac{1}{4} \sum_{k=n}^{n+1} C_{m-s}(k) \cdot l_k \cdot W \cdot (2 - d_k - d_{k-1})^2 \\ & + \frac{1}{4} \sum_{k=1}^{n-2} C_{m-m}(k) \cdot l_k \cdot W \cdot (d_{k+2} - d_k + d_{k+1} - d_{k-1})^2 \\ & + \frac{1}{4} \sum_{k=n+1}^{2n-2} C_{m-m}(k) \cdot l_k \cdot W \cdot (d_{k+2} - d_k + d_{k+1} - d_{k-1})^2 \end{aligned} \quad (6)$$

In many cases, output buffers excessively load the critical circuits, degrading their performance. This is true in case of frequency dividers, thus, the design of the output buffer attempts to minimize capacitive loading at the output nodes of the latch. It consists of a differential stage, modified into an f_T doubler stage by adding transistors M2 and M3 as presented in Fig.14. This setup approximately halves the C_{gs} capacitance as compared to a standard differential stage (keeping transconductance constant) and also increases the unity-gain bandwidth of the buffer. The compact shunt peaking inductor designed for the D-latch is re-used in the buffer without significant increase in area.

A differential design is adopted both in the D-latch and output buffers; the layout is kept as symmetric as possible. The quadrature outputs use 50Ω transmission lines to the bond pads. Four RF output and two RF input pads are included for measurements. Due to bond pad limitations the chip area is $0.9 \times 0.7 \text{ mm}^2$. However, the active area is less than half of the above value. The chip micrograph is shown in Fig.15. The divider was measured on wafer with high frequency differential probes (GSGSG). A 180° hybrid coupler provides the required anti-phase clock inputs. The measured input sensitivity as a function of frequency is shown in Fig.16.

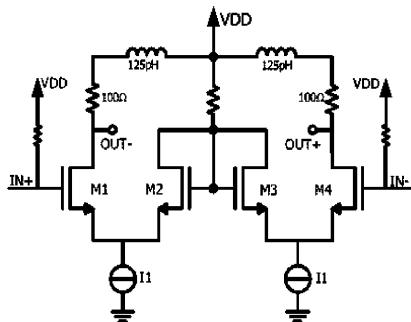


Fig.14. Output buffer

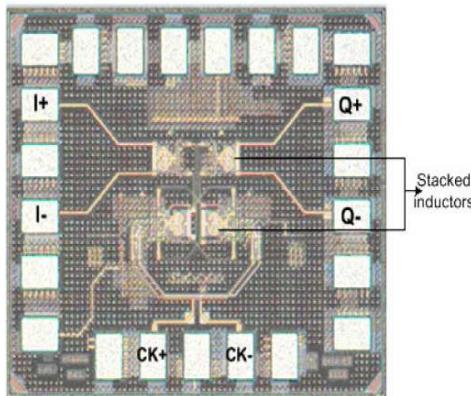


Fig.15. Output buffer

The highest input sensitivity is measured at 22GHz and the maximum operating frequency of the divider is 35.5GHz. The maximum frequency division is achieved at a power consumption of 14.4mW per D-latch from a 1.2V power supply. Each output buffer consumes 9.6mW and the total power consumption is 48mW. Fig.17 shows the measured output spectrum at maximum frequency of operation; the output power of -40dBm includes the losses from the cables and the 180° hybrid coupler (almost 30dB). The measured phase noise of the divider is -107.7dBc/Hz at 1MHz offset from the carrier whereas the input phase noise of the generator is -102.5dBc/Hz at double frequency and the same carrier power.

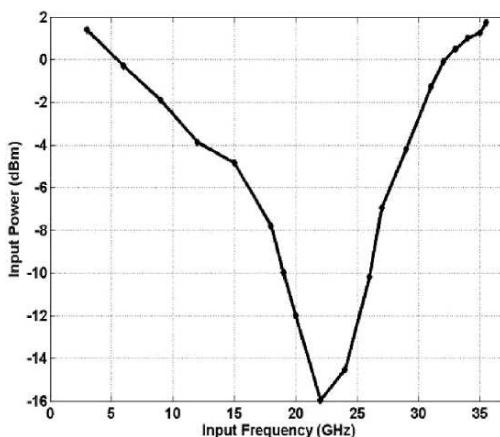


Fig.16. Static divider measurements: sensitivity curve

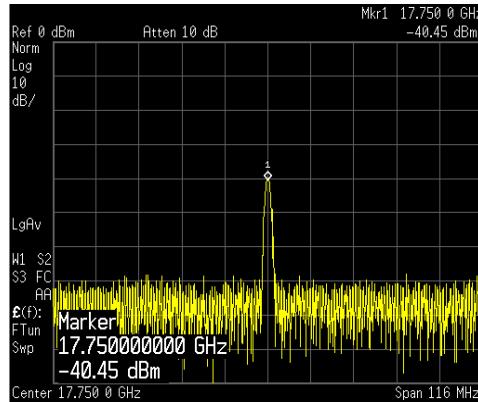


Fig.17. Static divider measurements: output spectrum

A static frequency divider operating at 60GHz is yet far of reach. Other possible solutions are injection-locked dividers. The problems related to the inherent small synchronization range makes injection-locked dividers less attractive for a VCO tunable in a 5-7GHz range around 60GHz as required by standardization instances. Given those factoids, together with sub-harmonic operation of the mixer of Fig.11 we conclude that architectures based on a lower frequency VCO are a better solution.

3.4. 60GHz quadrature VCO

The principle of the quadrature oscillator is based on coupling two LC Colpitts sections. According to Barkhaussen criteria the phase-shift on the loop must equal 2π or a multiple of it. The cross-coupling introduces a phase shift of π , thus, the phase shift of the two sections must equal π . Hence, if the two sections are identical, they should oscillate in quadrature.

When the transfer function of one section $H(j\omega)$ is symmetrical around its maximum, the theoretical model yields two equi-probable solutions for the oscillation frequency. However, this is not true in a practical implementation. The circuit diagram of the VCO is sketched in Fig.18. A differential Colpitts LC section with cross-coupled output buffers is the basic stage of the total VCO solution. An extra cross-coupling enhances the positive feedback around the Colpitts section providing a buffering mechanism for tapping energy from the VCO core. The oscillation frequency can be changed from I_{TUNE} and/or V_{BULK} . Denote ω_C the Colpitts frequency given by:

$$\omega_C \cong \frac{1}{\sqrt{L * \frac{C_1 * C_2}{(C_1 + C_2)}}} \quad (7)$$

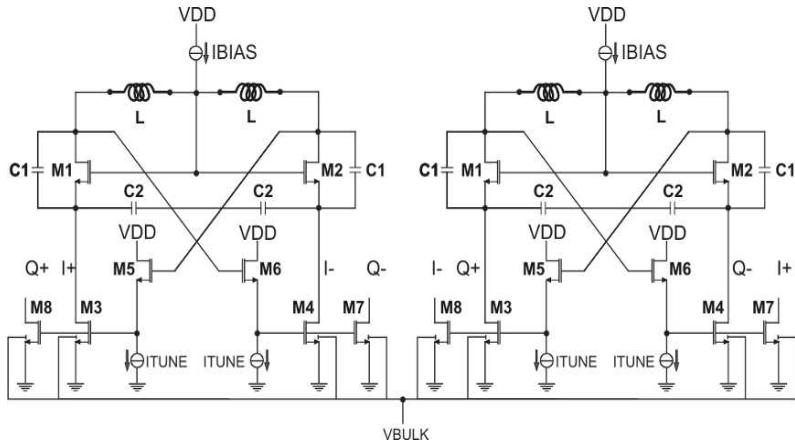


Fig.18. Quadrature VCO: circuit diagram

Denote ω_0 the frequency for the condition $k=0$ and $\omega_{1,2}$ the two equidistant frequencies around ω_0 . When the tank is lossy ($G_p \neq 0$), the real oscillation frequency for $k=0$ differs from the Colpitts frequency as:

$$\omega_0 \equiv \omega_C \sqrt{1 - \frac{g_m G_p L}{(C_1 + C_2)}} \quad (8)$$

The two frequencies ω_1 and ω_2 can be found from:

$$\omega_{1,2} \equiv \omega_0 * \left[1 \pm \frac{G_p}{2} \sqrt{\frac{L(C_1 + C_2)}{C_1 C_2}} \right] \quad (9)$$

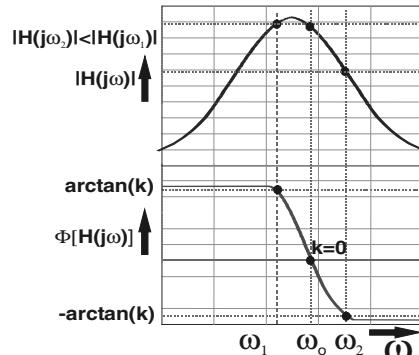


Fig.19. Quadrature VCO: operation principle

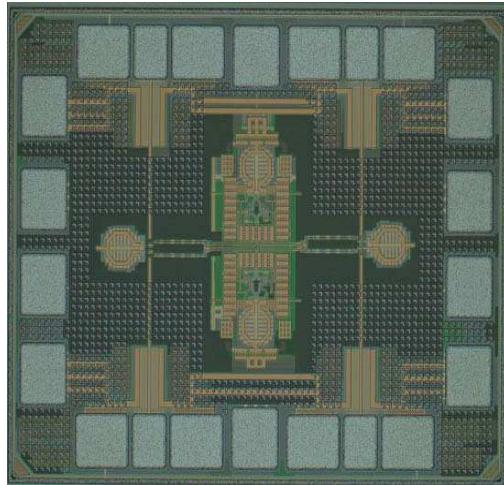


Fig.20. Quadrature VCO: chip photo

This case is depicted in Fig.19. Clearly, ω_2 is more probable to occur due to the larger loop-gain at ω_2 . By changing the coupling factor k , the oscillation frequency of the oscillator can be tuned as $\omega_{\text{LO}} \in (\omega_0, \omega_2)$. This VCO was implemented in CMOS090 LP and consumes 44mA from 1.2V supply. Figure 25 shows an active area of $495 \times 216\mu\text{m}^2$. Apparently, the chip is pad-limited; the total chip area (with bonding pads and including the output buffers) is $1100 \times 904\mu\text{m}^2$. The tuning characteristic of the VCO (different bulk voltages V_{BULK}) is presented in Fig.21; the total tuning range is about 2GHz.

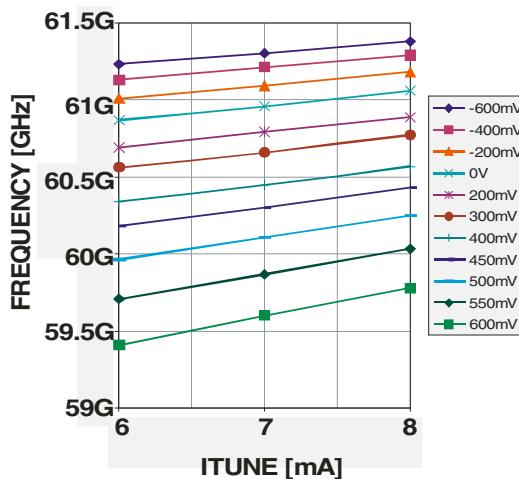


Fig.21. Quadrature VCO: fine-coarse tuning characteristic

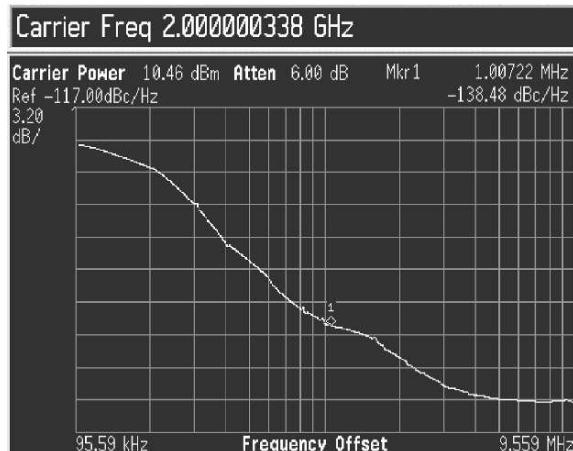


Fig.22. Quadrature VCO: phase noise @ 2GHz (after down-mixing)

The tuning mechanism works as a “gear-box” by stepping coarsely with V_T control through fine intervals generated by I_{TUNE} . For measurement purposes, the VCO output was down-mixed at 2GHz with an external mixer and the phase noise was measured using the spectrum analyzer phase-noise measurement setup. The result is presented in Fig.22; it shows a phase noise of -138.5dBc/Hz at 2GHz.

3.5. Complete I/Q down-converter

For the I/Q down-converter from Fig.6 we need an IF buffer. The buffer should not degrade the linearity of the total chain. The IF buffer from Fig.23 employs a super source-follower to linearly track the input terminal voltages and a 20Ω series resistance for a total 50Ω output impedance.

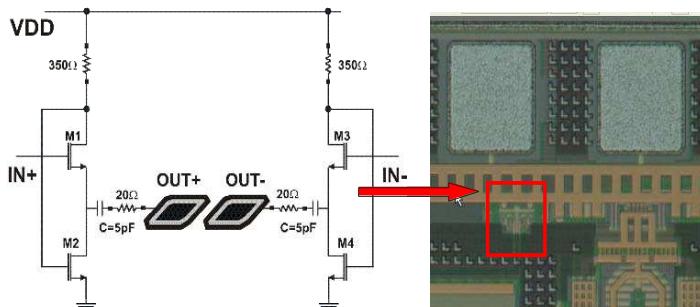


Fig.23. IF output buffer (50Ω)

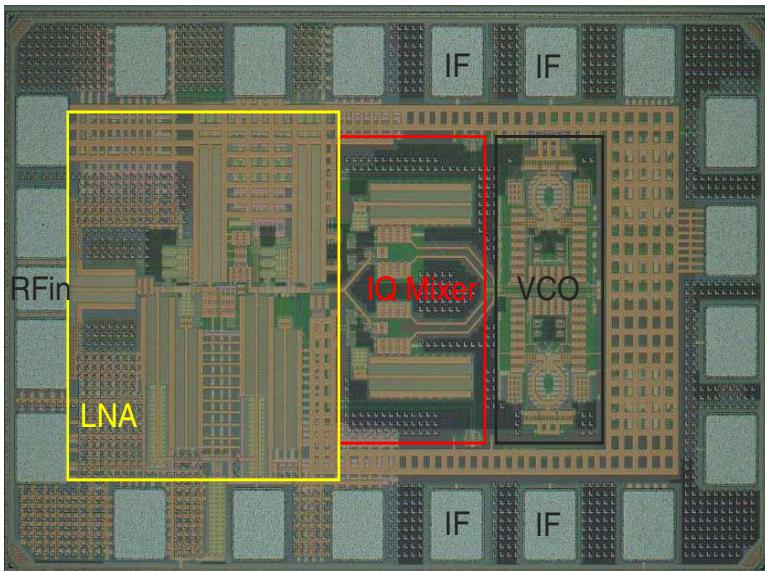


Fig.24. Complete I/Q down-converter

The complete I/Q down-converter is presented in Fig.24. For the reasons explained in section 3.3 the frequency divider was not included in this design. The measured results of the I/Q down-converter are presented in Table 1.

Frequency	58GHz—61.3GHz
Noise Figure	9.5dB
Gain	23dB
Supply voltage	1.2V(+/-10%)
Power dissipation	54mW
S11, S22	<-10dB
IIP3	-21.5dBm

Table 1:Measured results of the I/Q down-converter

4. Broadband IF amplifier

A broadband IF amplifier is an important building block in the total millimeter-wave radio concept. As the bandwidth of IF signals is in the order of few GHz a low noise amplifier with a large bandwidth is required. Size and power consumption are important design aspects. An inductor-less LNA with resistive feedback is sketched in Fig.25.

The first stage in the feed-forward path of the LNA is a common source, cascode amplifier. It consists of the common source amplifier M_{n1} , the cascode transistor M_{n2} and the load resistor R_{d1} . The cascode transistor M_{n2} increases the output impedance and the reverse isolation. At low voltage supply (<1.2V) the voltage drop across the load resistor and transistors becomes critical. The PMOS transistor M_{p1} takes a part of the DC current. For that, the AC current flows through the transistor M_{n2} as the output impedance of M_{p1} is larger relative to the input impedance of M_{n2} .

The second stage of the LNA provides a voltage-to-current conversion. The transistors M_{n6} and M_{n7} provide a DC voltage in series to the resistor R_m . In this way no DC current will flow through R_m . In Fig.26, the follower circuit, M_{n3} and M_{n4} , provides a voltage-to-voltage conversion. The voltage in node C tracks the voltage in node B. The current variation of M_{n3} is sensed on the output resistance of M_{p3} . A local feedback loop, consisting of M_{n4} , M_{p2} and R_{p2} follows these fluctuations by modulating the current source M_{n4} . In this way, the current of transistor M_{n3} is constant and equal to I_{bias} . As a consequence, the voltage signal V_C is converted in a current on the R_m and M_{n4} combination. The capacitor C_{p2} controls peaking at higher frequencies.

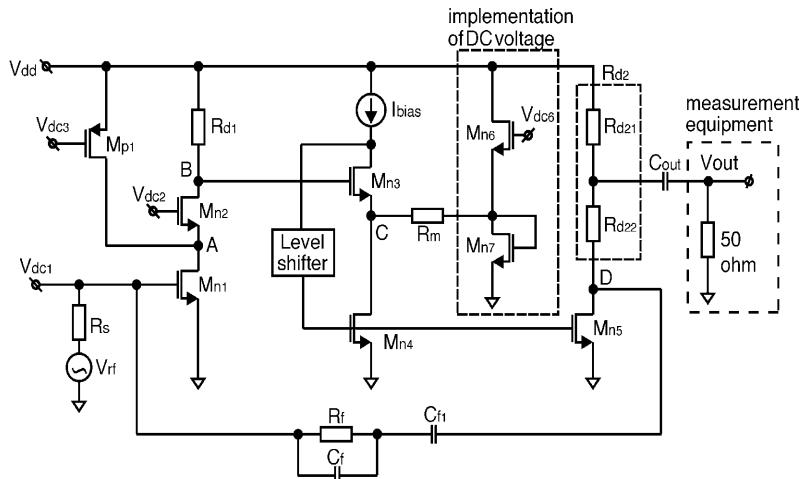


Fig.25. Broadband, low-noise IF amplifier

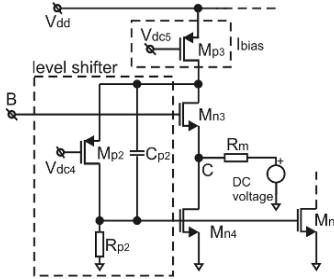


Fig.26. Second stage of the LNA

The gate voltages of M_{n4} and M_{n5} are equal. Therefore, the drain current of M_{n5} is a copy of the current in M_{n4} . The output current is converted in a voltage on the load resistor $R_{d2} = R_{d21} + R_{d22}$. For measurement purposes, the resistor R_{d2} of the output stage is divided in two resistors R_{d21} (50Ω) and R_{d22} .

The feedback resistor R_f is DC blocked by C_{f1} , while C_f is used to control peaking at higher frequencies and input matching (S_{11}). Assuming a perfect source follower ($V_C = V_B$), the voltage gain of the broadband LNA is:

$$G \approx -\frac{\frac{R_f}{R_s + R_f} g_{m1} R_{d1} A_c \frac{R_f \parallel r_{05} \parallel R_{d2}}{R_m}}{1 + \frac{R_s}{R_s + R_f} g_{m1} R_{d1} A_c \frac{R_f \parallel r_{05} \parallel R_{d2}}{R_m}} = -\frac{A_{V0}}{1 + (R_s A_{V0}) / R_f} \quad (10)$$

where R_s is the source resistance ($R_s = 50\Omega$) and g_{m1} is the transconductance of M_{n1} . The current gain A_c is equal to g_{m5}/g_{m4} , where g_{m4} and g_{m5} are the transconductances of M_{n4} and M_{n5} , respectively. The variable r_{05} is the output resistance of M_{n5} . The variable A_{V0} is the voltage gain of the feedforward path of the LNA. The input matching of the LNA is set by:

$$R_{in} \approx \frac{R_f}{1 - A_{V0}} \approx -\frac{R_f}{A_{V0}} \quad (11)$$

For sufficiently low values of resistor R_m and a low ratio R_m/R_{d1} , the NF of the broadband LNA can be approximated by:

$$NF \approx 1 + \frac{\gamma_n g_{d01}}{(g_{m1})^2 R_s} + \frac{\gamma_p g_{pd01}}{(g_{m1})^2 R_s} + \frac{1}{(g_{m1})^2 R_s R_{d1}} + \frac{R_s}{R_f} \quad (12)$$

where γ_n and γ_p are larger than 1 in the case of short channel transistors. The variables g_{d01} and g_{pd01} are the drain conductances of the transistors M_{n1} and M_{p1} ,

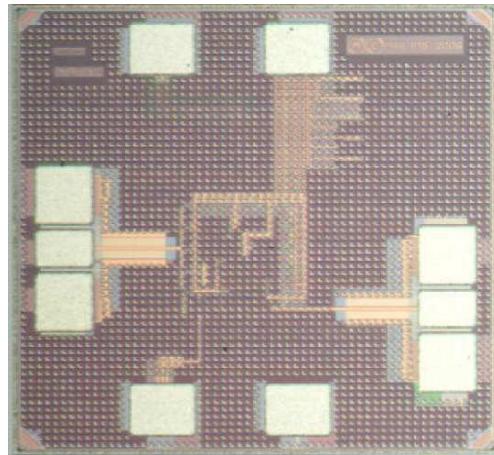


Fig.27. Inductor-less LNA: chip photomicrograph

respectively. The NF of the LNA will be mainly determined by the noise of the first stage and the feedback resistor R_f . For a high value of R_m and a high R_m/R_{d1} ratio, the noise of the second and the third stage will contribute to the total noise. Reducing the voltage swing at the node C, increasing R_m and/or increasing the bias currents can improve the linearity of the second stage. We can conclude that NF of the circuit can be traded for a good linearity.

The circuit is designed and implemented in a baseline CMOS 90nm LP process. Figure 27 shows the micrograph of the realized LNA. The active chip area is $400 \times 300 \mu\text{m}^2$. Total dissipation of the IC is 16.8mW at a supply voltage of 1.2V. In Fig.28 the power gain at 1GHz is 16dB, with a 3-dB bandwidth of 2GHz. A noise figure of 3.5dB is measured at 1GHz, while at 2.7GHz the noise figure is less than 3dB.

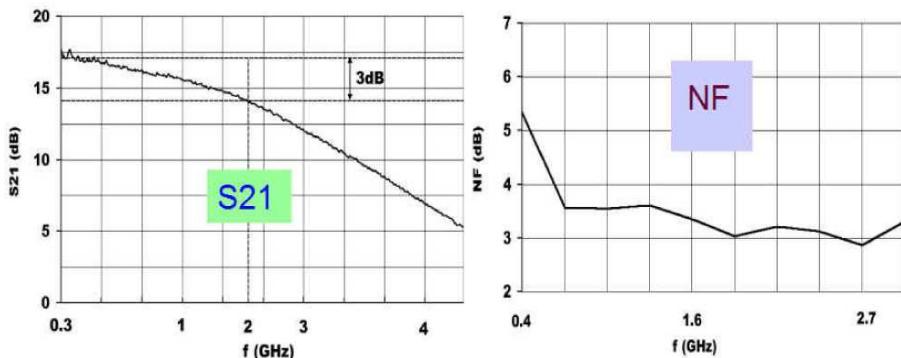


Fig.28. Inductor-less LNA: Power Gain and Noise Figure

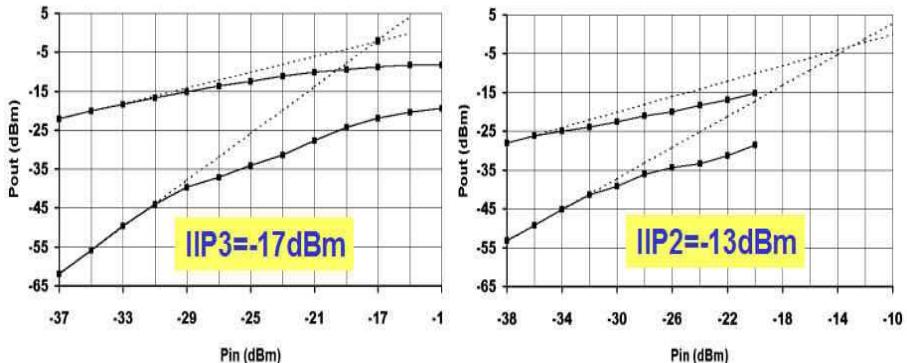


Fig.29. Inductorless LNA: IIP3 and IIP2

For linearity test, the frequencies of the fundamentals are chosen 800MHz and 900MHz. Figure 29 shows a measured IIP3 of -17dBm and a measured IIP2 of -13dBm.

5. Conclusions

We have presented some key CMOS building blocks for millimeter-wave wireless transceivers with emphasis on WPAN standards (802.15.3.c) at 60GHz. At the beginning, an in-depth analysis of active and passive devices brings the technology dimension into discussion. In the proposed designs, microstrip lines consisting of M1 as the ground plane and M6 as the signal line, are employed in the matching networks and resonant circuits together with high-Q metal-metal capacitors. Thereafter, a complete I/Q down-converter is presented. Trade-offs between technology, performance and circuit choices of millimeter-wave RF front-end circuits are discussed with circuit examples and measured results where applicable. The first example is a 60GHz two-stage cascode LNA with inter-stage matching. The harmonic, single-balanced mixer from the next section can be operated in harmonic or sub-harmonic modes depending on the phase of the LO signals present at the gates of the switching pairs. This design flexibility opens new ways for different mm-wave architectures with voltage-controlled oscillators at 60GHz or 30GHz. A 35.5GHz frequency divider illustrates the design hurdles inherent to static dividers beyond Ka bands. As a key enabling component, a compact, stack inductor, was used in the latch and output buffers. Compared to planar inductors, the proposed passive structure offers less (up to 45%) parasitic capacitance. This exercise substantiates the need for millimeter-wave radio architectures working in sub-harmonic modes with voltage-controlled oscillators at frequencies lower than f_{RF} . Thereafter, a novel 60GHz differential-Colpitts quadrature VCO is presented with a detailed analysis on its

operation. A complete I/Q down-converter with all the presented building blocks is discussed. Finally, a new broadband IF amplifier is presented. It points out the challenges of a broadband IF chain that requires low-noise, low-power and small footprint. All presented circuits function at 1.2V and are realized in the same CMOS090 LP technology. The process used, features damascene copper for the 6 metal layers (5thin+1thick) and low-k (<3.0) inter-metal dielectric between the thin metal layers. No extra process options (MIM capacitors or laser trimmed polysilicon resistors) were included in the presented designs.

References

- [1] Peter Smulders, "Exploiting the 60 GHz band for Local Wireless Multimedia Access: Prospects and Future Directions", *IEEE Communications Magazine*, pp. 140-147, January 2002.
- [2] K. Ohata et al., "1.25 Gb/s wireless gigabit Ethernet link at 60 GHz band", *IEEE int. Microwave Symp. Dig.*, pp. 373-376, June 2003.
- [3] Brian Floyd et al., "SiGe Bipolar Transceiver Circuits Operating at 60 GHz", *IEEE JSSC*, Vol.40, No.1, January 2005.
- [4] Scott K. Reynolds et al., "A Silicon 60-GHz Receiver and Transmitter Chipset for Broadband Communications", *IEEE JSSC*, Vol.41, No.12, pp. 2820-2831, December 2006.
- [5] C. Wang, et al., "A 60GHz Transmitter with Integrated Antenna in 0.18 μ m SiGe BiCMOS Technology", *ISSCC 2004 Digest of Tech. Papers*, pp. 186-187, February 2006.
- [6] B. Razavi, "A 60GHz CMOS Receiver Front-end", *IEEE JSSC*, Vol.41, No. 1, January 2006, pp.17-22.
- [7] T. Yao et al., "60-GHz PA and LNA in 90nm RF-CMOS", *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, San Francisco, CA, June 2006.
- [8] C.H. Doan, et al., "Design Considerations for 60 GHz CMOS Radios", *IEEE Communications Magazine*, pp. 132-140, December 2004.
- [9] C.H. Doan, et al., "Design of CMOS for 60GHz Applications," *ISSCC 2004 Digest of Tech. Papers*, pp. 440-538, February 2005.
- [10] D. Huang et al., "A 60GHz CMOS Differential Receiver Front-End Using On-Chip Transformer for 1.2 Volt Operation with Enhanced Gain and Linearity", *Symposium on VLSI Circuits Digest of Technical Papers*, Honolulu, Hawaii, June 2006.
- [11] M.A.T. Sanduleanu et al., "31-34GHz Low Noise Amplifier with On-chip Microstrip Lines and Inter-stage Matching in 90-nm Baseline CMOS", *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, San Francisco, CA, June 2006.

ANALOG/RF DESIGN CONCEPTS FOR HIGH-POWER SILICON BASED mmWAVE and THz APPLICATIONS

Ullrich R. Pfeiffer

Institute of High-Frequency and Quantum Electronics
University of Siegen, Hoelderlinstrasse 3, D-57068 Siegen, Germany
Email: ullrich@ieee.org

Abstract

An overview of analog and RF design concepts in silicon process technologies is presented, with the focus on high-power applications at mmWave frequencies and above. Key power amplifier design tradeoffs and high-frequency trends will be given. This includes power amplifier circuit design and packaging examples at 60 GHz.

1. Introduction

Recent advancements in silicon process technologies have made it possible to integrate active and passive devices with cut-off frequencies penetrating the mmWave and near Terahertz region. Active SiGe hetero-junction bipolar transistors (HBTs), for example, have demonstrated cut-off frequencies as high as $f_{max}/f_T = 350/300$ GHz [1] while silicon integrated passive Schottky Barrier Diodes (SBD) have shown cutoff frequencies above 1 THz [2]–[4]. The strength of silicon germanium (SiGe), for example, which exploits the same manufacturing steps as silicon-based CMOS, is that it allows analog, digital, and RF functionality to be integrated on a single chip. Because of their high integration level and cost-effectiveness, silicon process technologies have the potential to transform today’s mmWave and THz systems [5].

The following sections summarize analog and RF design concepts in silicon process technologies with the focus on high-power applications at very high frequencies, e.g. mmWave frequencies and above. The primary focus will be placed on SiGe BiCMOS technologies such as IBM’s advanced bipolar technology SiGe8HP. The key design tradeoffs for power amplifiers (PAs) will be discussed in detail in Sec. 2. Circuit design examples are given in Sec. 3 for single PAs and recent advancements in power combining techniques. A summary of analog and RF design concepts is given in Sec. 4 and a technology outlook for applications at THz frequencies is given in Sec. 5.

2. Power Amplifier Performance Tradeoffs

Silicon device performance improvements have been accomplished through steady progress in vertical and lateral scaling for parasitic resistance and capacitance reduction [6]. Various technology nodes are typically compared at their peak cutoff frequencies f_{max}/f_T . Higher cutoff frequencies can be leveraged in two ways. On one hand, a faster technology node can be used to bias devices at lower currents to achieve a given f_{max} performance. This leads to a reduced power dissipation. On the other hand, a faster technology can drive higher operating (e.g. carrier) frequencies with typically about 1/3 of peak f_{max} in practical applications. One of the drawbacks of the vertical profile scaling, however, is the reduction in achievable breakdown voltages (BV) by pushing out the Kirk effect to higher current densities. Power amplifiers are therefore regarded as one of the most challenging circuits in high-frequency communication systems.

Power amplifier performance tradeoffs can be expressed by several key parameters. This includes the output power P_{out} , power gain G , operating frequency f , linearity in terms of the third order input intercept point (IIP3), and power-added-efficiency (PAE). The following Eq. 1 shows a Figure of Merit for power amplifiers (FoM_{PA}) derived from the System Drivers Section of the 2005 ITRS document [7]. The FoM_{PA} is limited to linear PAs for ease of comparison. To the first order, the one-pole system response of a transistor current or voltage gain exhibits a 20-dB per decade roll-off over frequency. The FoM_{PA} therefore includes a factor $(f/f_{max})^2$ to compensate the power gain roll-off. The operation frequency f is normalized with the unity power gain f_{max} to make the FoM_{PA} independent of the technology capabilities.

$$FoM_{PA} = P_{out} \times G \times PAE \times (f/f_{max})^2 \quad (1)$$

The optimal technology choice for high-power applications in the mmWave regime is driven by two factors: (1) the technology should be fast enough to provide sufficient power gain at the desired operation frequency, and (2), the output power level should be high enough to meet the system requirements. The following describes various design tradeoffs between the key parameters of Eq. 1.

2.1. Output power limitations in silicon process technologies

The maximum power P_{out} of Eq. 1 that an amplifier can deliver to an external load R_{load} (e.g. 50 Ω) depends to the first order on three technology factors: (1) the breakdown voltage (BV) of the device, (2) the maximum collector current densities (j_c), and (3) the achievable impedance transformation ratio (r). This

simply reflects $P_{out} = (BV)^2/rR_{load}$, where $r = R_{load}/R_{in}$ is the impedance transformation ratio of the amplifiers output matching network. The impedance R_{in} is provided by the matching network and is seen at the device level, e.g. at the collector in a common emitter transistor configuration. R_{in} defines the load-line of the amplifier. In other words, the achievable output power of a single device is independent of the operating (e.g. carrier) frequency. Although, at higher frequencies the amplifier's gain may be too low to provide sufficient amplification. High-performance silicon technologies therefore provide various device performance options to tradeoff output power (e.g. breakdown voltages) versus cutoff frequencies.

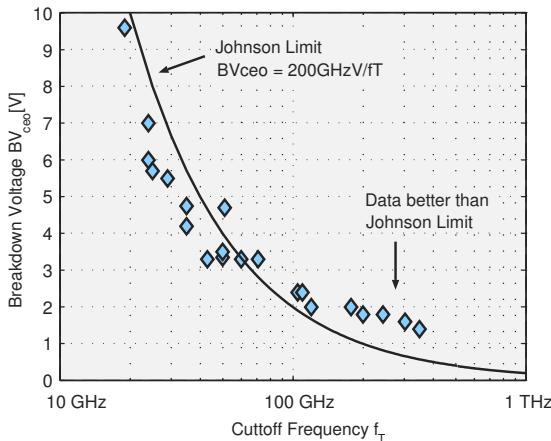


Fig. 1. Breakdown voltage vs. peak cutoff frequency (f_T) for various SiGe process technologies nodes. Data courtesy of IBM [6].

Fig. 1 shows the scaling of the breakdown voltage (BV_{ceo}) for high-performance SiGe HBT technologies. The data by courtesy of IBM [6] shows the breakdown voltage versus the peak cutoff frequency f_T for various technology nodes. As the SiGe technologies advance to faster cutoff frequencies their breakdown voltages initially drop rather fast and then level off to about 1 V. This BV_{ceo} roll-off is a fundamental device limit first explained by Johnson in 1965 [8]. The Johnson limit assumes that, to the first order, BV_{ceo} can be calculated from the product of a critical electrical field E_{crit} at which carrier multiplication by impact ionization runs away, times the width of the collector-base space-charge region W_{CB} . If one reduces W_{CB} to increase f_T (shorter transit time), one automatically reduces BV_{ceo} . According to the Johnson limit the following rule-of-thumb applies:

$$f_T \cdot BV_{ceo} \leq 200\text{GHzV}. \quad (2)$$

The data in Fig. 1 initially follows the Johnson limit very nicely, advanced technologies however, show some deviation from the simple Johnson limit. Note, the BV data is shown versus f_T , the short circuit current gain, instead of the more relevant unity power gain f_{max} because f_T is generally available in literature and less prone to extraction and measurement errors. In general, f_T and f_{max} are correlated.

The maximum available power in a given technology node can be estimated from the maximum voltage swing across the transistor's load-line impedance R_{in} . The voltage swing is limited by impact ionization which can be overcome to a certain extent if the bias circuit provides a low enough external base resistance [9], [10]. A low external resistance provides an escape path for hot carriers and hence works against the avalanche process taking place in a HBT device. As a result the base current will change its polarity for the duration of the ac-swing that is within the avalanche region. In practice BV_{ce} can be about 50% higher than BV_{ceo} as long as BV_{cbo} is not exceeded. For a class-A operation the maximum available power can be estimated as

$$P_{out} = \left(\frac{xBV_{ceo} - V_{knee}}{2\sqrt{2}} \right)^2 / R_{in}, \quad (3)$$

where $V_{knee} \approx 0.3V$ is the knee voltage at the transistor's saturation region, BV_{ceo} is the open base breakdown voltage, $R_{in} = R_{load}/r$ is the load-line impedance provided by the output matching network, and x represent a 50% BV_{ceo} increase due to the bias network impedance. Based on Eq. 3 one can estimate the maximum available output power for the BV_{ceo} data shown in Fig. 1. The output power, though, strongly depends on the achievable load-line impedance. At high frequencies a low R_{in} is difficult to achieve due to design limitations in the output matching network. For instance, high quality factors for resonant matching networks in silicon are only in the order of 10-20 which make it difficult to achieve an impedance transformation r much larger than 5 in realistic applications. Under such conditions one can calculate the single device output power trend of Eq. 3, which is shown in Fig. 2. The data suggests that at 60 GHz, e.g. 1/3 of $f_T = 180$ GHz, single transistor amplifiers should be able to reach output powers as high as 20 dBm. These observations are supported by recent publications in this frequency range. See also Fig. 6 for a comparison.

The data in Fig. 1 initially follows the Johnson limit very nicely with a 20dB/decade roll-off, however, as one gets closer to about 1 THz the output power exhibits a faster than $1/f_T$ roll-off. This is known as the THz-gap at which xBV_{ceo} in Eq. 3 equals the knee voltage V_{knee} and creates zero output power. At that point the breakdown voltage is too small to make the transistor switch – a fundamental limit for electronics systems trying to bridge the THz gap.

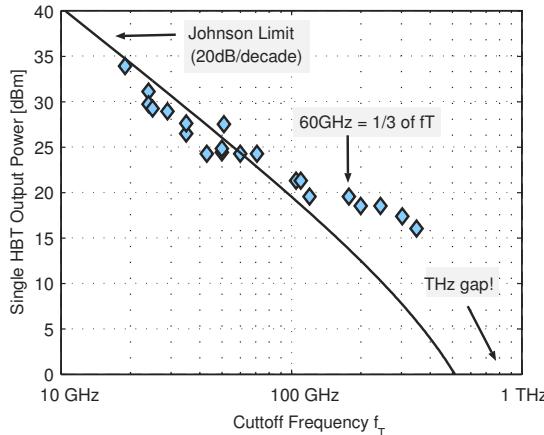


Fig. 2. Maximum single device output power trend versus cutoff frequency (f_T). The data is shown for single transistors operating in a class-A mode without power combining. An external base resistance that supports a 50% extension of BV_{ceo} , a load-line impedance (R_{in}) of 10Ω with $r = 5$, and $R_{load} = 50\Omega$. Note, other values in Eq. 3 may shift this data within 3 dB or more but its trend remains the same.

Power combining techniques are often used to enhance the total output power of an amplifier by combining the power of several in parallel operating transistors. Such techniques are often combined with an impedance transformation. Practical limitations for such power combiners are set by three parameters: (1) the power loss in the combiner, (2) the total power dissipation capabilities of the final amplifier and package assembly, and (3) by the ability to model the parasitics of large power combining networks accurately. Because of this, power combiners have limited gain and efficiency which in turn affect their output power as follows:

$$P_{out} = \frac{P_{dc} \cdot PAE}{(1 - 1/G)}, \quad (4)$$

where P_{dc} is the thermal power handling capability of the package assembly and PAE and G have their usual meaning. For instance, a packaged power-combined amplifier with a thermal power handling capability of 4W and a total power gain of 6 dB at a peak PAE of 10% can only generate up to 530 mW per chip. These considerations show that higher output powers can only be achieved if the individual transistor amplifiers are able to operate at high efficiency and high power gain.

2.2. Power-gain tradeoffs

The favorite design technique for high-frequency power amplifiers is to cascade many gain stages to form one circuit with high overall power gain. Because each amplifying device also contributes its own high frequency roll-off the last stage in the amplifier chain should compress first for optimum linearity. The design of the last amplifier stage is most critical in terms of overall amplifier performance and shall be our focus here. The gain of the last stage depends on the circuit architecture, the intrinsic device gain, and the electrical losses in the input and output matching networks. A low overall gain mainly affects the amplifier efficiency, whereas a high gain stage exhibit a larger negative feedback and tends to be more unstable.

The loss of impedance matching networks is most critical at the output. A 3-dB loss for instance affects the efficiency different at the input or output. With α and β being the loss of the input and output matching network, respectively, the peak power added efficiency (PAE_{peak}) is

$$PAE_{peak} = \left(\frac{1}{\beta} - \frac{\alpha}{G} \right) \eta, \quad (5)$$

where $\eta = P_{out}/P_{dc}$ is the amplifier's drain efficiency. Eq. 5 shows that a higher power gain G can reduce the absolute power lost at the input, whereas, loss accrued at the output may not be compensated. For instance, the PAE_{peak} roll-off for a $G = 16$ -dB power amplifier with a drain efficiency of $\eta = 10\%$ is shown in Fig. 3 a). In case of loss-less matching the PAE_{peak} versus amplifier gain is shown in Fig. 3 b).

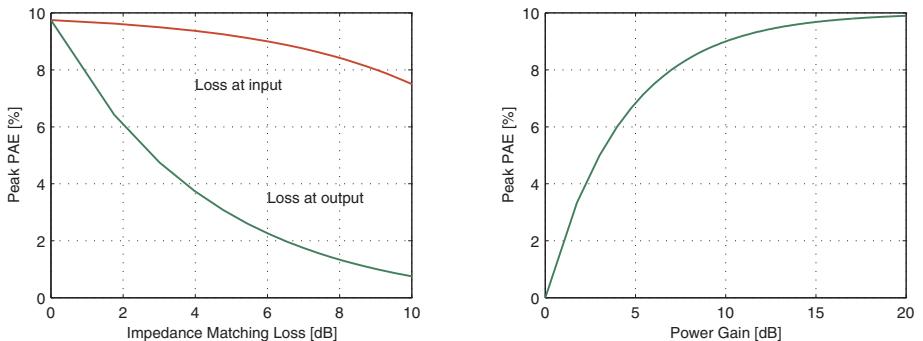


Fig. 3. Peak power-added-efficiency (PAE_{peak}) roll-off for a 16-dB power amplifier with a $\eta = 10\%$ drain efficiency. b) shows the PAE_{peak} versus G for loss-less matching.

As a result, high-gain stages like cascode amplifiers are preferred over common-emitter stages. Specific circuit design examples are shown in Sec. 3. Intrinsic gain improvements are discussed in more details in section Sec. 2-4.

2.3. Power added efficiency tradeoffs

Applications in the mmWave frequency range may require a transistor bias point at about 1/3 of the technologies peak f_T/f_{max} . This imposes a limit on their achievable power added efficiencies because high-performance SiGe devices inherently have a steep f_T roll-off at high injection. The roll-off is due to the high germanium content and a large germanium gradient in the neutral base which enables the high beta and f_T of the device [11]. At peak efficiency, the dc bias current through the device is increased up to a point beyond the peak of the f_T curve and the device happens to be biased at a collector current density which heavily compresses the amplifier gain and therefore limits the available efficiency for class-A operation to only a few percent [12]. Higher efficiency modes of operation are difficult to achieve. Early results have shown peak efficiencies up to about 20.9% [13], although, their low power gain makes these amplifiers impractical for realistic applications. This suggests that high power added efficiencies can only be achievable if the device is operated at sufficient back-off, e.g. about 1/10 of its peak f_T/f_{max} .

2.4. Technology roadmap for future high-power applications

The trends in SiGe HBT performance are expected to continue over the next years with vertical and lateral device scaling being the main driver for improved f_{max} performance. Reducing the base resistance R_{bb} and collector base capacitance C_{cb} without impacting both avalanche and self-heating mechanisms are the key device challenges. Clearly, the output power roll-off at high frequencies as shown previously in Fig. 2 is likely to continue. New approaches will have to combine devices in parallel either within amplifiers directly or on the system and package level. The integration of passive devices may open up alternative approaches. Schottky barrier diodes with cutoff frequencies up to 1.5 THz [2], [4] are possible candidates today. Such diodes can be integrated in standard SiGe or CMOS foundry processes paving the way to high-density system architectures. A heterodyne approach may combine active and passive devices to overcome headroom and electro-migration issues in the future.

3. mmWave power amplifiers in silicon

Early SiGe mmWave power amplifiers at 60 GHz and 77 GHz have been based on common emitter topologies for simplicity in layout and design, see Fig. 4 a). Such circuits have demonstrated that SiGe pre-production technologies are well suited to generate high output powers in excess of 16 dBm. To mitigate the impact of process, voltage and temperature variations (PVT-variations), however,

circuits should be operated at back-off to implement a dynamic bias control for constant gain or output power. Recently, silicon mmWave amplifiers are turning to other solutions like cascode topologies [14].

In a cascode topology, see Fig. 4 b), the gain device (T3) can be operated in common-emitter (ce) with a second transistor (T1) in common-base (cb) mode whose emitter is connected to the collector of the gain transistor. The common-base device shields the common-emitter gain transistor from voltage changes in the circuit. The collector-base voltage of T3 is held constant with minimal charging of the collector-base junction capacitance, eliminating the effects of this internal lag capacitance and allowing higher frequency responses. A cascode circuitry, therefore, is a good candidate for efficient high-gain power amplifiers and is preferred over a common-emitter stage. The improvement in peak *PAE* for a high gain amplifier can be estimated from Fig. 3. For instance, a 20-dB amplifier can be about 2% points more efficient than a 6-dB amplifier can be. The high gain, however, may exacerbate stability problems.

3.1. Power amplifiers without power combining

Both amplifiers shown in Fig. 4 drive a differential antenna to gain an extra 3-dB power pick-up at the antenna port. The common-emitter circuit in Fig. 4 a) uses a balanced operation (half-circuit shown), whereas Fig. 4 b) uses a push-pull topology with two cascode gain stages. Unlike balanced circuits which are independent halves driven differentially, the push-pull circuit in Fig. 4 b) can create ac-grounds to form low-impedance circuit nodes. The ac-grounds are created in the center of single inductors (transmission lines), which connect directly across devices. The length of the inductor may be smaller than a quarter-wave RF-choke because its residual reactance can be absorbed by the input and output matching networks. The ac-grounds can be used to supplement bypass capacitors. Capacitors alone may operate beyond self-resonance at very high frequencies and can cause stability problems.

For instance, an ac-ground in Fig. 4 b) is located at the base of the common-base output pair T1 and T2 to provide the required low external base resistance for operation above the BV_{ceo} limit. To the first order, the base current circulates back and forth between the two devices and only some bypass capacitance is needed to account for base current asymmetries. This allows the output voltage to swing ± 2.5 V around the 4-V DC supply voltage without causing the device to break down. The two devices are laid out in close proximity to support the shortest possible base connection. The amplifier achieves a peak power gain of 18 dB with a 13.1 dBm output referred 1-dB compression point, a 12.7% peak PAE, and a saturated output power of 20 dBm [14], [16]. The amplifier achieves a compact layout with a core area of 0.075 mm^2 . The highest peak PAE is 13% at 59 GHz. It includes an adaptive bias control that is programmable through a

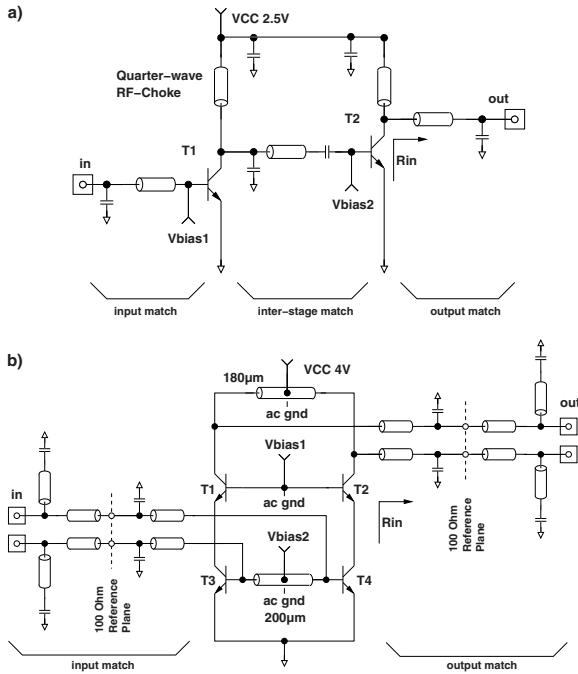


Fig. 4. Figure a) shows a schematic of a common emitter balanced circuit (half-side) [15] and figure b) shows a differential cascode circuit [14].

three-wire serial digital interface. The chip can be fully molded in a low-cost plastic packaging technology as described in [17].

3.2. Power amplifiers with power combining

On-chip power combining and balanced device operation has been exploited to enhance the maximum available output power per chip (20 dBm [14], 18.5 dBm [18], 17.5 dBm [19] and 21 dBm [20]) in the past. At very high frequencies, power combining techniques seem to be a last resort to overcome the low transistor breakdown voltage trend shown previously in Fig. 1. While on-chip power combining techniques have been used for some time [21], [22], distributed active transformers (DATs) have recently created some excitement. The DAT topology provides power combining and efficient impedance transformation simultaneously and promises highly efficient, fully integrated amplifiers in standard low-voltage process technologies. First DAT implementations have been shown at frequencies around 2.4 GHz for example [23], [24], and recently, a 23 dBm DAT has been demonstrated at 60 GHz [25], [26].

Distributed Active Transformers as described in [23] use two single-turn planar slab inductors at 2.4 GHz to form a transformer where the primary

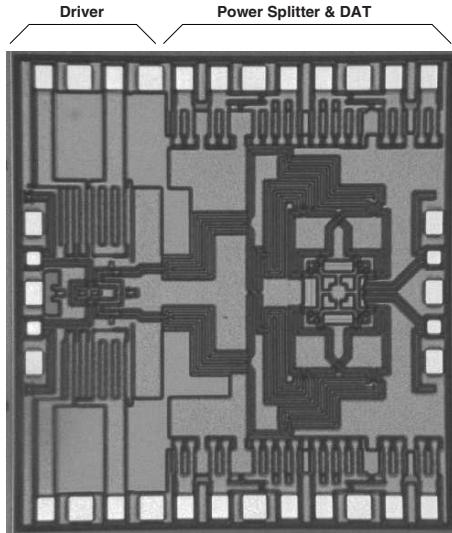


Fig. 5. Chip micrograph. The overall size of the chip is $1.9 \times 1.8 \text{ mm}^2$ while the size of the transformer is only $160 \times 160 \mu\text{m}^2$. Most of the area between the driver amplifier and the DAT is due to a Wilkinson power divider network. Figure courtesy of IBM [25], [26].

inductor is broken up into 4 quarter-sections to facilitate the connection of four synchronized push-pull amplifiers. Each synchronized push-pull amplifier couples magnetically to the same single turn secondary inductor in such a way, that their alternating magnetic fluxes add constructively to form a uniform circular current in the secondary winding. Since each amplifier on the primary side utilizes only one quarter of the primary inductor length, and not its full length, the impedance transformation ratio is $1:n$ (1:4) instead of $1:n^2$ known for a regular four-turn transformer. Scaling the DAT topology from 2.4 GHz to mmWave frequencies imposes a series of challenges. Coplanar transformers are typically used only at lower frequencies where low coupling factors, substrate and skin effect losses, and inaccuracies caused by model to hardware discrepancies can be tolerated [24]. Commonly used on-chip transformers are either made of interwound spiral inductors or coplanar coupled wires (slab inductors) to promote mutual magnetic coupling. In order to operate any transformer in the mmWave frequency range its primary inductance has to be reduced substantially, which in turn, requires the values of additional tuning capacitors to be extremely small. Therefore, it is crucial to have a transformer or DAT structure which allows accurate modeling and the prediction of parasitic effects. The most important design challenges for mmWave DATs are as follows: (1) the DAT requires well synchronized push-pull amplifiers under all operating conditions to maintain the

correct load line impedance for each amplifier, (2) tuning of the DAT for low loss and high efficiency requires accurate compact electromagnetic modeling as well as accurate parasitic extraction techniques, (3) non-idealities of the transformer such as its inter-winding capacitance limit the scaling to higher frequencies and require optimized 1:1 transformer structures.

Fig. 5 shows a chip micrograph of a 60 GHz DAT published in [25], [26]. The overall size of the chip is $1.9 \times 1.8 \text{ mm}^2$ while the size of the transformer is only $160 \times 160 \mu\text{m}^2$. Most of the area between the driver amplifier and the DAT is due to a Wilkinson power divider network used to distribute the input power to the four corners of the DAT. The loss of the dividers is less significant at the input of the DAT as has been previously shown in Fig 3.

Fig. 6 shows a comparison of recent SiGe power amplifier designs operating in the mmWave frequency range, see [13], [15], [16], [18]–[20], [25], [27], [28]. The data includes single amplifiers as well as power-combining techniques in silicon.

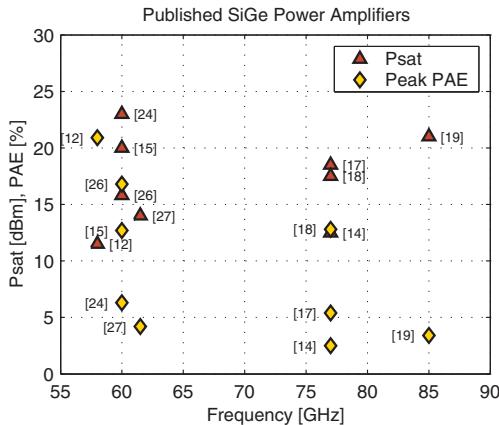


Fig. 6. Comparison of recent SiGe power amplifier designs operating in the mmWave frequency range. The data corresponds to a f_T in the range of 165–270 GHz. Compare also to Fig. 2.

4. Design considerations for mmWave frequencies and above

There is a series of design challenges for circuits in silicon beside low breakdown voltages and output power limitations mentioned previously. They are related to modeling, simulation, design and test of complex circuits. Historically, high-frequency design tools like ADS [29] have been used to simulate the performance of basic circuit elements. Such circuit designs have been rather interconnect centric with a limited number of active devices. Circuits in silicon, however, are traditionally larger and use lower frequency design environments

like Cadence [30]. The push in frequency imposes a number of challenges for the design of mmWave circuits in silicon [12] that will be addressed next.

4.1. Interconnect modeling

Transmission lines have been widely used for the design of mmWave amplifiers in silicon. They are primarily used for device interconnects, RF-chokes and for input/output impedance match purposes. Unlike in GaAs technologies, for example, where the substrate is enclosed by the signal conductor on the top surface and the ground metalization on the backside, transmission lines in silicon technologies are entirely formed within the back-end of the process technology. As such, the spacing of the signal to the ground is about an order of magnitude smaller than in GaAs technologies, e.g. only $10\ \mu\text{m}$. As such they have to be designed to comply with current density and electro migration design rules. Scalable 7-segment lumped ladder networks have been used for simulation purposes up to 110 GHz [31]. Side shields can be used on both sides of the signal conductor to minimize coupling to adjacent structures. Small inductors may be an alternative to transmission lines if electro-migration rules can be met [32], [33].

Because of the design size and complexity electromagnetic simulations of larger areas are rarely used. Interconnects are rather broken up into simple circuit elements that have a known electrical characteristic. Special care, however, is needed for microstrip transitions. Design libraries with predefined bends, T-sections, steps complement other passive library elements. Vertical interconnects (vias) can be modeled with lumped inductors because of the small vertical separation between interconnect layers. Short interconnects may be modeled based on parasitic extractions, e.g. parallel instances with parasitic capacitance, resistance, and inductance associated with their wiring may be extracted to form a distributed interconnect model.

The physical design of RF contact pads has a large effect on the circuit's performance and the input/output impedance matching technique. At mmWave frequencies contact pads are electrically large and their parasitic capacitance and losses are significant. Common practice is to simply scale the contact area down to reduce the parasitic capacitance and to limit the amount of lossy eddy currents in the substrate. There are scaling limits, however, since the minimum pad size is restricted by wire or flip-chip bonding constraints. Additional DT (deep-trench) isolation or NS (npn sub-collector) shields are commonly used to reduce losses without increasing the pads parasitic capacitance unnecessarily. Metal shields directly below the pad structure, however, have not been widely adopted due to their high capacitance.

The circuits shown in Fig 4 b) for instance uses a solid metal ground shield right below the contact pad. The low loss weighs out the increase in capacitance

since the increased pad capacitance can be tuned out by a shunt transmission line stub that adds little loss to the overall structure. Larger pad structures are possible this way with sufficient contact area ($70 \times 70 \mu\text{m}^2$) to comply with packaging constraints. The shunt transmission line is ac coupled to ground with a series metal-insulator-metal (MIM) capacitor. The resonant structure is wide-band and has low insertion and return-loss. It is therefore nearly electrically transparent to the off-chip load impedance. Note, the pad is not part of the impedance transformation network and the length of the on-chip transmission line that connects to the pads can have an arbitrary length while only affecting the loss but not the output match of the PA – a very convenient layout feature to accommodate various chip layouts.

4.2. Transistor device modeling

Active devices are based on the Vertical Bipolar Inter-Company (VBIC) transistor model which has been fit to DC and small-signal S-parameter measurements. Such models have shown reasonable good results but have known shortcomings specifically if accurate large-signal responses are required [34]. Other more physics based models like the High Current Model (HICUM) may be more appropriate for large-signal applications like power amplifiers [12].

4.3. Chip package design at mmWave frequencies

Packaging of millimeter-wave (mmWave) components is particularly challenging because of the associated complexity both in design and fabrication. The small wavelength involved demands high-precision machining, accurate alignment, or high-resolution photo-lithography. Furthermore, mmWave circuits usually exhibit low integration levels and are often assembled using expensive and bulky waveguides [35], [36]. Common are monolithic microwave integrated circuit (MMIC) packages [37], [38] which are primarily an outgrowth of the microwave hybrid integrated circuit (MIC) technology and the discrete device packaging technology. The use of conventional low-cost packaging technologies is limited and has only been reported at lower frequencies [39]. This is in part because the typical 1 nH/mm lead- and wirebond-inductances are prohibitively high at mmWave frequencies [40] and plastic packaging materials are quite lossy. Moreover, standard mold materials have not generally been characterized at mmWave frequencies [41].

Recently, 60 GHz transmitter and receiver chip-sets have been direct-chip-attached (DCA) on low-cost printed circuit boards together with 7-dBi cavity-backed folded dipole antennas [17]. The DCA package has a size of $7 \times 11 \text{ mm}^2$ and is shown in Fig. 7 before encapsulation. The package build-up provides a well-controlled EM environment, which makes the antenna performance less

sensitive to the surrounding package- and PCB-level dielectric and metal structures. The package represents an individual component with a formal package for environmental protection around it.

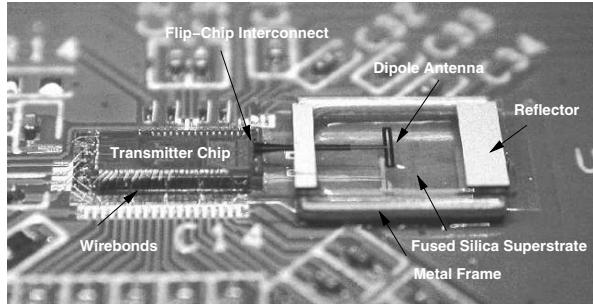


Fig. 7. Photo of the 60 GHz package before encapsulation. The fused silica substrate is transparent. The PA output is flip-chip connected to the antenna feed and radiation is orthogonal to the board surface. Figure courtesy of IBM [17].

5. Technology outlook for future THz applications

Today's advanced SiGe semiconductor process technologies are capable of processing signals at frequencies as high as 120 GHz [2], [26], [42]. Historically, this mmWave frequency range was limited to III/V semiconductor technologies which exhibit low integration levels and higher cost. Such mmWave systems were based on MMIC packaging technologies to combine discrete devices together in a single package or module which are bulky and expensive in larger quantities. Unlike III/V technologies, SiGe for example, can enable high-density applications like multiple antennas arrays or imaging array systems at a fraction of the cost of competing technologies [43]. This mixture of high-integration and yet high-frequencies make this technology so appealing for applications in the terahertz (THz) region.

The terahertz region of the electromagnetic spectrum, spanning from 100 GHz through 10 THz, is of increasing importance for a wide range of scientific, and commercial applications. This interest is spurred by the unique properties of this spectral band and the recent development of terahertz sources and detectors. For instance, the great promise of terahertz imaging systems is the ability to see through clothing with the required spatial resolution to detect concealed weapons, whether made of metal, ceramics, plastics or other materials. Terahertz radiation is nonionizing so that health risks are minimal (as opposed to X-rays). Unfortunately, as of today, THz imaging capabilities are extremely limited due to their low level of integration, limited operation speed, and small pixel count.

A detector which provides amplitude and phase information and not only an average measure of the wave intensity will contain three dimensional information about objects. This represents an important difference to other types of detectors, especially bolometers that only measure the wave power intensity. By keeping spectral and phase information about the detected wave, imaging systems enable spectral analysis in the THz spectrum. Amplitude and phase information is also being used in data communication systems, THz phased arrays, and radar systems. Although SiGe technologies are new in this field, their cutoff frequencies are well within the lower-end of the THz frequency spectrum. In the light of scaling trends and novel analog and RF design techniques it is expected that SiGe will have a strong influence on future THz circuits and systems. In light of the output power limitations being discussed in Sec. 2-1, circuit design at the border of the THz-gap remains to be an exciting task.

6. Summary and Conclusion

Silicon circuit design concepts for high-power applications at very high frequencies have been discussed, with the primary focus being on SiGe BiCMOS technologies. In light of scaling trends SiGe is the most promising technology for future highly integrated mmWave and THz systems [44]. Today's applications include high-speed communications systems at 60 GHz [17], [45]–[47] or automotive radar systems at 77 GHz [48]. Although SiGe has been first in this application space, with product-like performance, alternative state-of-the-art CMOS technologies may soon have adequate performance [49]. Either technology will face a similar output power roll-off at high frequencies and power amplifiers will be key in future applications. For a summary of state-of-the-art SiGe power amplifiers see Table I.

TABLE I
COMPARISON OF SiGe MMWAVE POWER AMPLIFIERS

Freq. [GHz]	Tech. [μ m]	Mode ¹	Power ² Comb.	P _{sat} [dBm]	GT _{sat} ³ [dB]	CP1dB [dBm]	GT _{max} ⁴ [dB]	PAE _{peak} [%]	Reference
58	SiGe 0.13	singl. 50 Ω	no	11.5	—	—	4.2	20.9	Valdes-Garcia [13]
60	SiGe 0.13	diff. 100 Ω	4x	23	13	—	20	6.3	U. Pfeiffer [25]
60	SiGe 0.13	diff. 100 Ω	no	20	4.5	13.1	18	12.7	U. Pfeiffer [16]
60	SiGe 0.18	bal. 100 Ω	no	15.8	—	11.2	11.5	16.8	Wang et al. [27]
61.5	SiGe 0.13	diff. 100 Ω	no	14	6	8.5	12	4.2	Pfeiffer et al. [28]
77	SiGe 0.13	diff. 100 Ω	no	18.5	—	—	—	5.4	Li et al. [18]
77	SiGe 0.13	singl. 50 Ω	2x	17.5	12	14.5	17	12.8	Komijani et al. [19]
77	SiGe 0.13	diff. 100 Ω	no	12.5	4.5	11.6	6.1	2.5	Pfeiffer et al. [15]
85	SiGe 0.13	singl. 50 Ω	4x	21	5	—	8	3.4	Afshari et al. [20]

¹Single-ended, differential, or balanced mode of operation

²On-chip power combining used

³Gain at saturation

⁴Maximum transducer power gain, or small-signal gain

Acknowledgment

The author would like to thank the IBM T.J. Watson Research Center, Yorktown Heights NY, USA, for their support of this work. In particular the 60 GHz

wireless design team, e.g. S. Reynolds, B. Floyd, A. Valdes-Garcia, T. Beukema, and D. Liu. Special thanks goes to B. Gaucher, S. Gowda, and M. Soyuer from the Communications Technology Department for their support. Thanks to A. Joseph from the Analog & Mixed-Signal Technology Development, IBM Burlington, Essex Junction VT, USA for f_T -BV data.

References

- [1] M. Khater, J.-S. Rieh, T. Adams, A. Chinthakindi, J. Johnson, R. Krishnasamy, M. Meghelli, F. Pagette, D. Sanderson, C. Schnabel, K. Schonenberg, P. Smith, K. Stein, A. Stricker, S.-J. Jeng, D. Ahlgren, and D. Freeman, "SiGe HBT technology with $f_{max}/f_t = 350/300$ GHz and gate delay below 3.3 ps," *IEEE Int. Electron Devices Meeting*, pp. 247–250, Dec. 2004.
- [2] C. Mishra, U. Pfeiffer, R. Rassel, and S. Reynolds, "Silicon schottky diode power converters beyond 100 GHz," *Radio Freq. Integr. Circuits Symp.*, to be published in June 2007.
- [3] R. Rassel, J. Johnson, B. Orner, S. Reynolds, M. Dahlstrom, J. Rascoe, A. Joseph, B. Gaucher, J. Dunn, and S. S. Onge, "Schottky barrier diodes for millimeter wave SiGe BiCMOS applications," *IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 255–258, Oct. 2006.
- [4] S. Sankaran and K. K.O, "Schottky barrier diodes for millimeter wave detection in a foundry CMOS process," *IEEE Electron Device Lett.*, vol. 26, no. 7, pp. 492–494, July 2005.
- [5] T. Crowe, W. Bishop, D. Porterfield, J. Hesler, and R. Weikle, "Opening the terahertz window with integrated diode circuits," *IEEE J. Solid-State Circuits*, vol. 40, no. 10, pp. 2104–2110, Oct. 2005.
- [6] A. Joseph, D. Harame, B. Jagannathan, D. Coolbaugh, D. Ahlgren, J. Magerlein, L. Lanzerotti, N. Feilchenfeld, S. Onge, J. Dunn, and E. Nowak, "Status and direction of communication technologies – SiGe BiCMOS and RFCMOS," *Proceedings of the IEEE*, vol. 93, pp. 1539–1558, Sept. 2005.
- [7] "Section system drivers," *International Technology Roadmap for Semiconductors (ITRS)*, 2005.
- [8] E. Johnson, "Physical limitations on frequency and power parameters of transistors," *RCA Rev.*, vol. 26, pp. 163–177, 1965.
- [9] M. Rickelt and H.-M. Rein, "A novel transistor model for simulating avalanche-breakdown effects in Si bipolar circuits," *IEEE J. Solid-State Circuits*, vol. 37, no. 9, pp. 1184–1197, Sept. 2002.
- [10] R. Singh, D. L. Harame, and M. M. Oprysko, *Silicon Germanium: Technology, Modeling, and Design*. IEEE Press, 2003.
- [11] J. Pan, G. Niu, A. Joseph, and D. L. Harame, "Impact of profile design and scaling on large signal performance of SiGe HBTs," in *IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, Sept. 2004, pp. 209–212.
- [12] U. Pfeiffer and A. Valdes-Garcia, "Millimeter-wave design considerations for power amplifiers in a SiGe process technology," *IEEE Trans. Microw. Theory and Tech.*, vol. 54, no. 1, pp. 57–64, Jan. 2006.

- [13] A. Valdes-Garcia, S. Reynolds, and U. R. Pfeiffer, "A 60GHz Class-E power amplifier in SiGe," *Asian Solid-State Circuits Conf.*, pp. 199–202, Nov. 2006.
- [14] U. R. Pfeiffer, "A 20dBm fully-integrated 60GHz SiGe power amplifier with automatic level control," *European Solid-State Circuits Conf.*, pp. 356–359, Sept. 2006.
- [15] U. Pfeiffer, S. Reynolds, and B. Floyd, "A 77 GHz SiGe power amplifier for potential applications in automotive radar systems," *Radio Freq. Integr. Circuits Symp.*, pp. 91–94, June 2004.
- [16] U. Pfeiffer and D. Goren, "A 20dbm fully-integrated 60GHz SiGe power amplifier with automatic level control," *IEEE J. Solid-State Circuits*, to be published 2007.
- [17] U. Pfeiffer, J. Grzyb, D. Liu, B. Gaucher, T. Beukema, B. Floyd, and S. Reynolds, "A chip-scale packaging technology for 60-GHz wireless chipsets," *IEEE Trans. Microw. Theory and Tech.*, vol. 54, no. 8, pp. 3387–3397, Aug. 2006.
- [18] H. Li, H.-M. Rein, T. Suttorp, and J. Boeck, "Fully integrated SiGe VCOs with powerful output buffer for 77-GHz automotive radar systems and applications around 100 GHz," *IEEE J. Solid-State Circuits*, vol. 39, no. 10, pp. 1650–1658, Oct. 2004.
- [19] A. Komijani and A. Hajimiri, "A wideband 77GHz, 17.5dBm power amplifier in silicon," *IEEE Custom Integr. Circuits Conf.*, pp. 566–569, Sept. 2005.
- [20] E. Afshari, H. Bhat, X. Li, and A. Hajimiri, "Electrical funnel: A broadband signal combining method," *IEEE Int. Solid-State Circuits Conf.*, pp. 206–207, Feb. 2006.
- [21] K. Russell, "Microwave power combining techniques," *IEEE Trans. Microw. Theory and Tech.*, vol. 27, no. 5, pp. 472–478, May 1979.
- [22] K. Chang and C. Sun, "Millimeter-wave power-combining techniques," *IEEE Trans. Microw. Theory and Tech.*, vol. 83, no. 2, pp. 91–107, Feb. 1983.
- [23] I. Aoki, S. Kee, D. Rutledge, and A. Hajimiri, "Distributed active transformer – a new power-combining and impedance-transformation technique," *IEEE Trans. Microw. Theory and Tech.*, vol. 50, no. 1, pp. 316–331, Jan. 2002.
- [24] ———, "A fully-integrated 1.8-V, 2.8-W, 1.9-GHz, CMOS power amplifier," *Radio Freq. Integr. Circuits Symp.*, pp. 199–202, June 2003.
- [25] U. Pfeiffer and D. Goren, "A 23dBm 60GHz distributed active transformer in a silicon process technology," *IEEE Trans. Microw. Theory and Tech.*, to be published 2007.
- [26] B. Floyd, U. Pfeiffer, S. Reynolds, A. Valdes-Garcia, C. Haymes, Y. Katayama, D. Nakano, T. Beukema, and B. Gaucher, "Silicon millimeter-wave radio circuits at 60-100GHz," *Silicon Monolithic Integrated Circuits Conference*, pp. 213–218, Jan. 2007.
- [27] C. Wang, Y. Cho, C. Lin, H. Wang, C. Chen, D. Niu, J. Yeh, C. Lee, and J. Chern, "A 60GHz transmitter with integrated antenna in 0.18um SiGe BiCMOS technology," *IEEE Int. Solid-State Circuits Conf.*, pp. 186–187, Feb. 2006.

- [28] U. R. Pfeiffer, D. Goren, B. A. Floyd, and S. K. Reynolds, "SiGe transformer matched power amplifier for operation at millimeter-wave frequencies," *European Solid-State Circuits Conf.*, pp. 141–144, Sept. 2005.
- [29] Agilent Advanced Design System, "eesof.tm.agilent.com."
- [30] Cadence Design Systems, "www.cadence.com."
- [31] T. Zwick, Y. Tretiakov, and D. Goren, "On-chip SiGe transmission line measurements and model verification up to 110GHz," *IEEE Microwave and Wireless Components Letters*, vol. 15, no. 2, pp. 65–67, Feb. 2005.
- [32] T. Dickson, M.-A. LaCroix, S. Boret, D. Gloria, R. Beerkens, and S. Voinigescu, "30-100-GHz inductors and transformers for millimeter-wave (Bi)CMOS integrated circuits," *IEEE Trans. Microw. Theory and Tech.*, vol. 53, no. 1, pp. 123–133, Jan. 2005.
- [33] M.Q. Gordon and T. Yao and S.P. Voinigescu, "65GHz receiver in SiGe BiCMOS using monolithic inductors and transformers," *IEEE SiRF Tech. Dig.*, pp. 265–268, Jan. 2006.
- [34] C.-J. Wei, J. Gering, and Y. Tkachenko, "Enhanced high-current vbiC model," *IEEE Trans. Microw. Theory and Tech.*, vol. 53, no. 4, pp. 1235–1243, Apr. 2005.
- [35] D. Parker, "Microwave industry outlook – defense applications," *IEEE Trans. Microw. Theory and Tech.*, vol. 50, no. 3, pp. 1039–1041, Mar. 2002.
- [36] J. Schepps and A. Rosen, "Microwave industry outlook – wireless communications in healthcare," *IEEE Trans. Microw. Theory and Tech.*, vol. 50, no. 3, pp. 1044–1045, Mar. 2002.
- [37] T. Midford, J. Wooldridge, and R. Sturdivant, "The evolution of packages for monolithic microwave and millimeter-wave circuits," *IEEE Trans. Antennas Propag.*, vol. 43, no. 9, pp. 983–991, Sept. 1995.
- [38] M. Hauhe and J. Wooldridge, "High density packaging of X-band active array modules," *IEEE Transactions on Manufacturing Technology*, vol. 20, no. 3, pp. 279–291, Aug. 1997.
- [39] A. Bessemoulin, M. Parisot, and M. Camiade, "1-Watt Ku-band power amplifier MMICs using low-cost quad-flat plastic package," *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 2, pp. 473–476, June 2004.
- [40] J.-Y. Kim, H.-Y. Lee, J.-H. Lee, and D.-P. Chang, "Wideband characterization of multiple bondwires for millimeter-wave applications," *Asia-Pacific Microwave Conference*, pp. 1265–1268, Dec. 2000.
- [41] T. Zwick, A. Chandrasekhar, C. Baks, U. R. Pfeiffer, S. Brebels, and B. P. Gaucher, "Determination of the complex permittivity of packaging materials at millimeter wave frequencies," *IEEE Trans. Microw. Theory and Tech.*, Jan. 2006.
- [42] E. Laskin, S. Nicolson, P. Chevalier, A. Chantre, B. Sautreuil, and S. Voinigescu, "Low-power, low-phase noise SiGe HBT static frequency divider topologies up to 100GHz," *IEEE BCTM Digest*, pp. 235–238, Oct. 2006.
- [43] A. Hajimiri, H. Hashemi, A. Natarajan, X. Guan, and A. Komijani, "Integrated phased array systems in silicon," *Proceedings of the IEEE*, vol. 93, no. 9, pp. 1637–1654, 2005.

- [44] J.-S. Rieh, B. Jagannathan, D. Greenberg, M. Meghelli, A. Rylyakov, F. Guarin, Z. Yang, D. Ahlgren, G. F. P. Cottrell, and D. Harame, "SiGe heterojunction bipolar transistors and circuits toward terahertz communication applications," *IEEE Trans. Microw. Theory and Tech.*, vol. 52, no. 10, pp. 2390–2408, Oct. 2004.
- [45] S. Reynolds, B. Floyd, U. Pfeiffer, T. Beukema, J. Grzyb, C. Haymes, B. Gaucher, and M. Soyuer, "Silicon 60-GHz receiver and transmitter chipset for broadband communications," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2820–2830, Dec. 2006.
- [46] B. Floyd, S. Reynolds, U. R. Pfeiffer, T. Beukema, J. Grzyb, and C. Haymes, "A silicon 60GHz receiver and transmitter chipset for broadband communications," *IEEE Int. Solid-State Circuits Conf.*, pp. 184–185, Feb. 2006.
- [47] Y. Sun, S. Glisic, and F. Herzl, "A fully differential 60ghz receiver front-end with integrated PLL in SiGe:C BiCMOS," *European Microw. Conf.*, pp. 198–201, Sept. 2006.
- [48] A. Natarajan, A. Komijani, X. Guan, A. Babakhani, Y. Wang, and A. Hajimiri, "A 77GHz phased array transmitter with local LO-path phase-shifting in silicon," *IEEE Int. Solid-State Circuits Conf.*, pp. 182–183, Feb. 2006.
- [49] T. Yao, M. Gordon, K. Yau, M. Yang, and S. Voinigescu, "60-GHz PA and LNA in 90-nm RF-CMOS," *IEEE RFIC Symposium Digest*, pp. 147–150, June 2006.

SiGe BiCMOS AND CMOS TRANSCEIVER BLOCKS FOR AUTOMOTIVE RADAR AND IMAGING APPLICATIONS IN THE 80-160 GHz RANGE

S.P. Voinigescu¹, S. Nicolson¹, E. Laskin¹, K. Tang¹ and P. Chevalier²

1) ECE Dept., University of Toronto, Toronto, ON M5S 3G4, Canada

2) STMicroelectronics, 850 rue Jean Monnet, F-38926, Crolles, France

Abstract

This paper examines the suitability of advanced SiGe BiCMOS and sub 65nm CMOS technologies for applications beyond 80GHz. System architectures are discussed along with the detailed comparison of VCOs, LNAs, PAs and static frequency dividers fabricated in CMOS and SiGe BiCMOS, as required for automotive cruise-control radar, high data-rate radio, and active and passive imaging in the 80GHz to 160GHz range. It is demonstrated experimentally that prototype SiGe HBT and BiCMOS technologies have adequate performance for all critical 80GHz building blocks, even at temperatures as high as 125 C. Although showing promise, existing 90nm GP CMOS and 65nm LP CMOS circuits at these frequencies remain significantly inferior to their SiGe counterparts.

1. Introduction

Potential applications of silicon ICs in the 80-160 GHz range include automotive cruise control (ACC) radar [1], millimeter-wave passive [2],[3] and active [4] imaging, and 10Gb/s short-range wireless links [5]. Over the last 4 years, several publications have explored the implementation of 77GHz IC building blocks in SiGe HBT technology [7]-[18]. Although mm-wave CMOS oscillators have been reported at frequencies as high as 194 GHz [19], only recently the phase noise and tuning range of 77GHz CMOS VCOs have become competitive with those of SiGe BiCMOS implementations [20]. Several 90nm and 65nm CMOS amplifiers operating in the 80-100 GHz range with less than 10dB gain have recently been announced [21] or are in press [4],[22]. The interest in SiGe BiCMOS and CMOS for mm-wave SOCs has been kindled by the favorable impact that transistor scaling has on practically all transistor high-frequency figures of merit (FoMs), and by the hope that the expected lower wafer cost will unravel a wide range of new applications and consumer

products. Integration beyond the basic building blocks, at the receiver, transmitter and even transceiver level, has already been demonstrated in SiGe HBT technology at 77 GHz [23]-[27] and at 160 GHz [28]. An amplifier with over 15 dB gain at 140 GHz, the highest in silicon, has also been fabricated [28]. This paper compares transistor and basic building block performance in SiGe HBT, SiGe BiCMOS and nanoscale CMOS technologies for mm-wave SOCs and discusses the most suitable system architectures that lead to the lowest power dissipation, smallest die area and die cost.

2. SiGe HBT vs. 65nm n-MOSFET performance comparison

Benefiting from the clear guidelines set forth by the International Roadmap for Semiconductors (ITRS), CMOS technology scaling has continued unabated to nanometre dimensions. Power dissipation, noise figure, and phase noise performance of mm-wave ICs all improve with scaling. At the same time, Fig. 1 illustrates that SiGe BiCMOS technology now retains a three-generation lithography advantage over CMOS in terms of f_T and f_{MAX} [29] and therefore results in significantly lower product development cost.

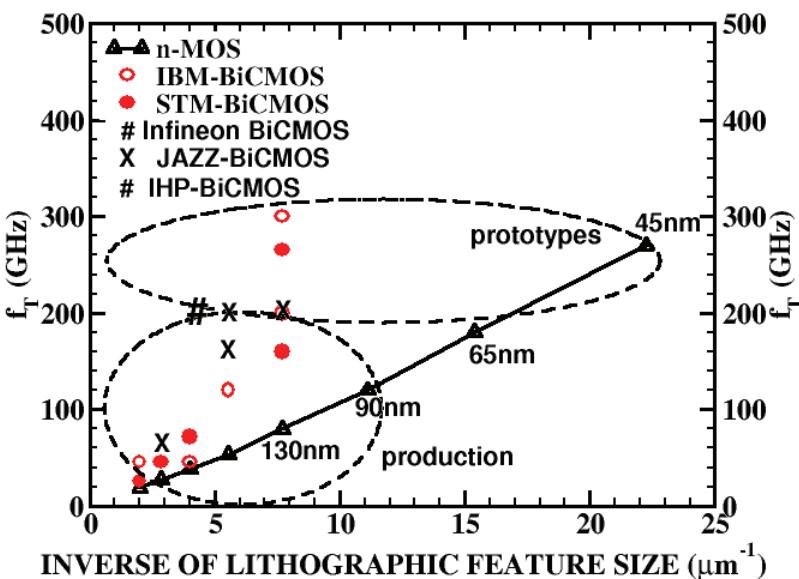


Fig.1. f_T scaling in CMOS and SiGe BiCMOS technology nodes [29].

Fig.2 compiles the measured f_B , f_{MAX} and NF_{MIN} characteristics of a $65\text{nm} \times 90 \times 1\mu\text{m}$ low-power (LP) n-MOSFET, and of a $3 \times 0.13\mu\text{m} \times 2.5\mu\text{m}$ SiGe HBT, as a function of drain current, and collector current, per unit gate width and emitter length, respectively [30]. In both devices f_{MAX} reaches 300 GHz and

NF_{MIN} , measured at 40 GHz, is about 1 dB, comparable to that of InP HEMTs. The HBT has 40% higher f_T and its optimal bias current densities for minimum noise or maximum gain are 5-6 times larger than in the 65nm MOSFET. Both devices are biased at a drain-source (collector-emitter) voltage of 1.2V, but the HBT can also operate safely with collector-emitter voltages exceeding 1.6V in common emitter CE, and beyond 3V in common base CB configurations [31]. At comparable f_{MAX} , the higher current densities and voltage swing, lower collector-substrate capacitance, along with the higher transconductance, give the HBT a significant advantage over MOSFETs in power amplifiers [32] and high-speed output drivers [33]. Furthermore, as illustrated in Fig. 3, even though the MOSFET has lower noise figure below 15 GHz, because of the higher f_T , the HBT noise figure increases at a slower rate at mm-wave frequencies, making it more suitable for LNAs above 60 GHz. Note that in Fig. 3 the MOSFET optimum noise bias does not change with frequency, whereas the optimum noise current density and, therefore the f_T , increase with frequency for HBTs.

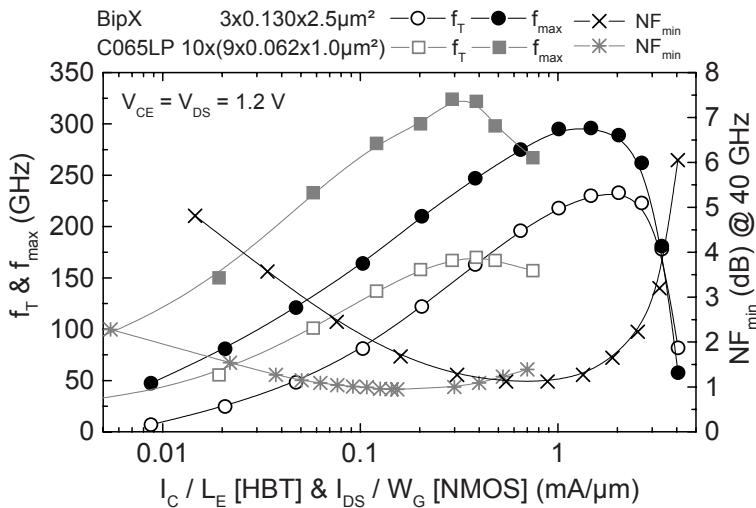


Fig.2. Measured 65nm LP n-MOSFET vs. SiGe HBT f_T , f_{MAX} and NF_{MIN} vs. collector current characteristics [30].

Fig. 4a, shows that GP bulk and SOI MOSFETs from different foundries exhibit remarkably similar f_T - I_D characteristics which scale almost ideally from one technology node to another [4],[33]. Note that, for the first time, there is no improvement in the peak f_T value between 90nm GP and 65nm LP n-MOSFETs because the physical gate lengths are practically identical. On the contrary, as illustrated in Fig. 4b, the peak f_T current density of SiGe HBTs increases in every new generation [34], and the optimal biasing conditions for HBT-circuits must be revisited, typically increased, in new nodes or at higher frequencies.

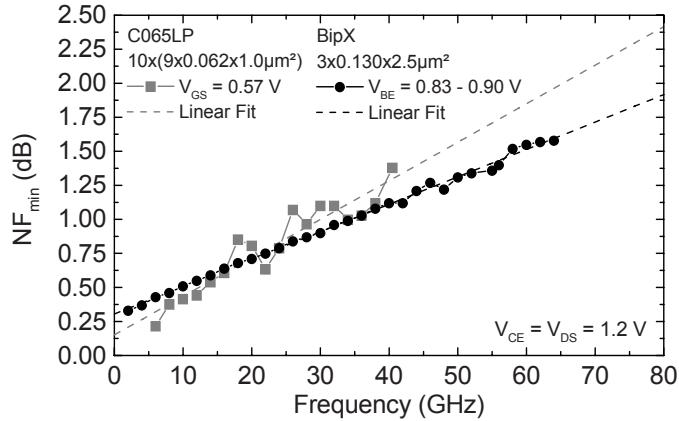


Fig.3. Measured N_{FMIN} as a function of frequency for a SiGe HBT and a 65nm LP n-MOSFET [30].

Finally, the measured intrinsic voltage gain is plotted in Fig. 5 vs. current density - rather than versus effective gate voltage - for n-MOSFETs across technology nodes and for different gate lengths in the 65nm LP node. These results show that 90nm GP MOSFETs have higher voltage gain than 130nm MOSFETs for all gate lengths, and that high threshold voltage (HVT) 65nm LP devices have less gain than low threshold voltage (LVT) ones. Furthermore, a 130nm MOSFET fabricated in the 65nm LP node has higher gain than a 130nm device fabricated in the 130nm node. Increasing gate length beyond $2 \times L_{MIN}$ brings no improvement in analog performance with severe degradation of HF performance [35]. Ironically, the GP LVT 90nm MOSFETs have better analog performance and dissipate less power than the LP 65nm MOSFETs.

3. Inductors, transformers and antennas

Similar to MOSFETs and HBTs, passive components such as antennas, inductors and transformers also follow Moore's law. For example, (1) shows that when the inductor diameter d , average diameter d_{avg} , metal width W , and inter-winding spacing [36] are all reduced by the scaling factor S , the inductance decreases proportionally. It can also be shown that the parasitic capacitance to ground decreases by S^2 and that the self-resonant frequency (SRF) and the peak Q frequency (PQF) increase S times while the peak Q remains largely unchanged. This suggests that one can continue to employ lumped inductors and transformers at mm-wave frequencies and thus take advantage of the most natural and most economical way to shrink the size of mm-wave silicon ICs far

beyond what has been accomplished with transmission lines, distributed baluns and power splitters [1],[9],[15].

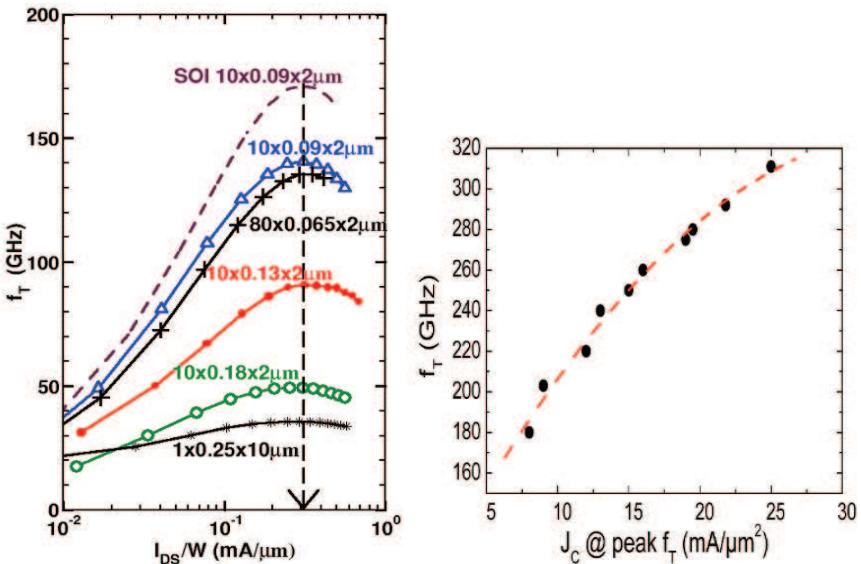


Fig.4. a) Measured f_T vs. drain current density per unit gate width for a) n-MOSFETs in different technology nodes [4] and b) measured peak f_T value of SiGe HBTs as a function of the peak f_T current density per emitter area [34].

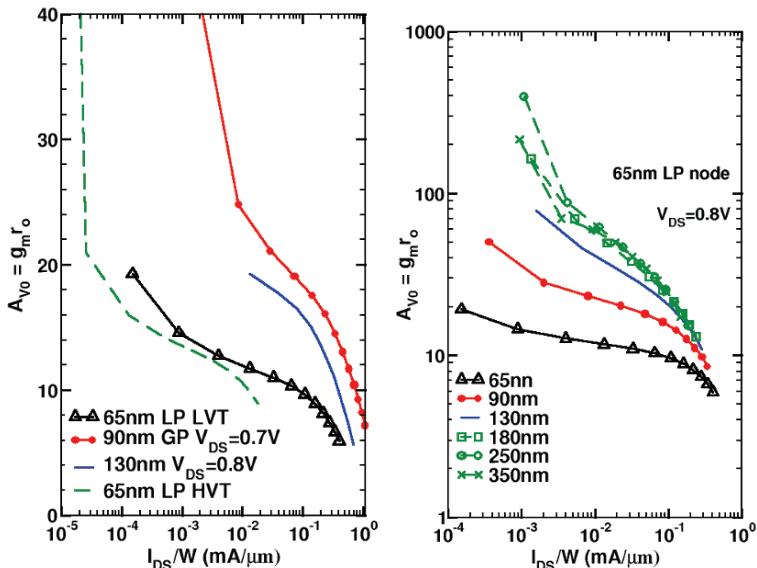


Fig.5. Intrinsic voltage gain a) across technologies and b) for different gate lengths in a 65nm LP CMOS technology as a function of drain current density.

$$L \approx \frac{6\mu_0 n^2 d_{avg}^2}{11d - 7d_{avg}} \quad \Rightarrow \quad \frac{L}{S} \approx \frac{6\mu_0 n^2 \left[\frac{d_{avg}}{S} \right]^2}{11 \frac{d}{S} - 7 \frac{d_{avg}}{S}} \quad (1)$$

Fig. 6 reproduces the die photo of a differential dipole antenna designed for 160GHz operation which occupies less than $200\mu\text{m} \times 200\mu\text{m}$ and is driven by a differential-to-single-ended converter realized with a vertically stacked transformer. The simulated gain and return loss of the antenna are plotted in Fig.7, while the structure and equivalent circuit of the transformer, extracted from ASITIC y-parameter simulations are shown in Fig.8.

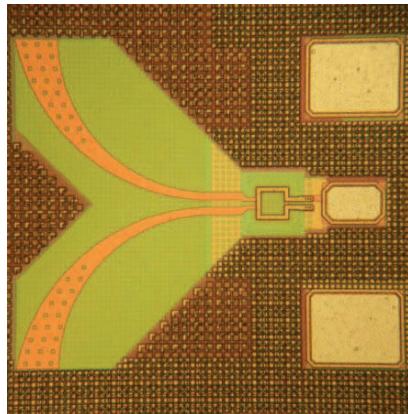


Fig.6. Die photo of 160-GHz dipole antenna with vertically-stacked transformer as single-ended to differential converter.

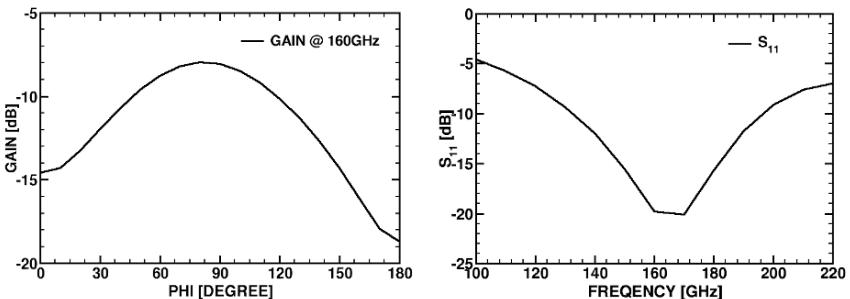


Fig.7. Simulated gain at 160 GHz and simulated return loss of antenna using ANSOFT's HFSS.

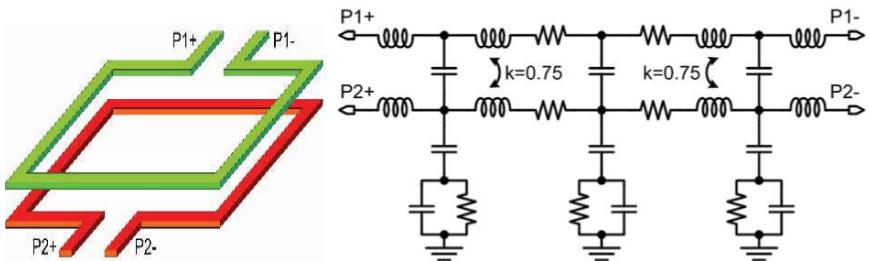


Fig.8. Pictorial view of vertically-stacked transformer and multi-section equivalent circuit model extracted using ASITIC.

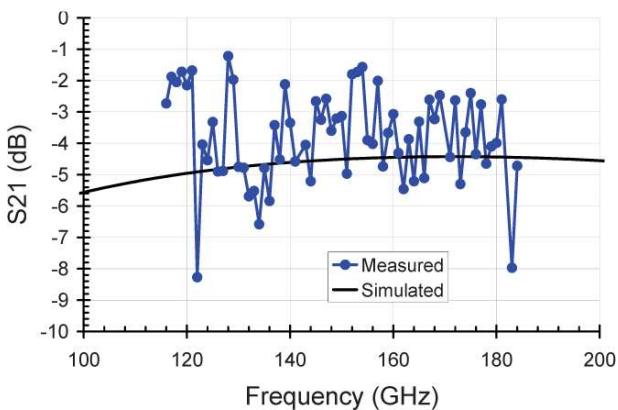


Fig.9. Measured vs. simulated S₂₁ of a vertically-stacked transformer.

The transformer was fabricated as a separate test structure in a standard digital back-end with 6 copper layers [28]. Its transmission loss is about 4 dB and was measured on wafer in the 110 to 170 GHz range. Fig. 9 compares measurements with simulations showing good agreement, well within the measurement scatter. While thick and wide metal lines are useful to reduce loss in t-lines and baluns [15],[24],[37],[38], to increase coupling and to reduce the footprint of transformers and vertically-stacked inductors, it is critical that the vertical and lateral spacing between windings is shrunk below 1 μ m. This is difficult to accomplish in a process with a thick aluminum top metal.

Finally, should t-lines or inductors be used as matching elements at mm-waves? Are transformers [37] or classical quarter-wavelength couplers and baluns the most effective components for single-ended-to-differential conversion in mm-wave circuits above 60 GHz? The wealth of experimental evidence regarding inductance and Q per layout area, circuit size, and circuit performance [1],[15]-[18],[19]-[21],[24],[27]-[28],[37]-[38], all point to the fact that, just as at lower

frequencies, lumped inductors and transformers lead to lower die size with comparable or better overall circuit performance.

4. Design flow for mm-wave silicon ICs

Compared to analog and RF design flows, the design flow for mm-wave ICs is complicated by the need to model every piece of interconnect longer than $15..20\mu\text{m}$ as a distributed transmission line. An effective way to contain the modelling effort is to include all interconnect leading to and from an inductor in the inductor itself and to extract the 2π equivalent circuit of the ensemble using ASITIC, as in [39]. At the cell level, the main goal is to minimize footprint by merging the transistor layouts of differential pairs and mixing quads and thus shrink the length and parasitic capacitance of local interconnect. The accurate extraction of RC parasitics at the layout-cell level (i.e. interdigitated transistor or varactor cell, cascode cell, differential pair cell, switching quad cell, cross-coupled pair cell, etc.) is critical for the accurate modelling of the significant gain and noise figure degradation in circuits with nanoscale MOSFETs. The MOSFET series parasitics are notoriously degraded by layout contact and via resistance. This is illustrated in Fig. 10, where the gain of a 90GHz 3-stage cascode amplifier implemented in 65nm LP CMOS [4] is reduced from 15 dB to 8 dB, and its noise figure increases from 5 dB to 7 dB when the parasitics of the transistor layout are included in simulation. All other components are unchanged. Note that there is hardly any shift in $S_{11}(f)$ and $S_{22}(f)$, or in the centre frequency of the $S_{21}(f)$ and $NF(f)$ characteristics, suggesting that the transistor layout parasitics are mostly resistive and not capacitive. Because of the larger R_E and R_B [30] and smaller C_{bc}/C_{be} ratio (i.e. reduced Miller effect) for the same current, circuits realized with HBTs are less sensitive to layout parasitics than those with MOSFETs.

Based on these general observations, a design flow that has been found to work well up to 160 GHz is summarized below:

- Optimize the transistor/varactor emitter length l_E or gate finger width W_f to balance the degradation of f_{MAX} and NF_{MIN} due to R_E/R_S , R_B/R_G and minimize C_{bc}/C_{gd} . In circuits with MOSFETs and AMOS varactors, fix W_f and vary N_f to contain the impact of channel strain variation with W_f .
- Design the circuit at schematic level with R_G added to the MOSFET digital model. The latter is sufficient to turn a “digital” into a good “RF” model. R_S and R_D are normally already included in the digital model.
- Optimize the transistor, cascode, or CMOS inverter cell layout through proper choice of metal stack on drain/collector and source/emitter, by monitoring f_{MAX} and NF_{MIN} . The optimal transistor layout depends on the stage topology: CE/CS, CB(CG, CC/CD, cascode, CMOS inv., etc.
- Include RC-extracted transistor (cascode) layout in schematic.

- Design and model inductors and interconnect in ASITIC based on the desired inductance obtained from schematic-level design with extracted transistors and pad capacitance.
- Add the ground-plane and power-plane metal mesh and the metal fill patterns to the cell and extract the layout of the cell, excluding inductors.
- Add inductor and interconnect models to schematic of RC-extracted cell.
- Add interconnect between cells and model it in ASITIC, ADS or HFSS.

With this approach, the number of iterations between layout and schematic simulations is minimized and first-pass success with at least 10% accuracy is assured, even in the absence of RF foundry models for MOSFETs and varactors.

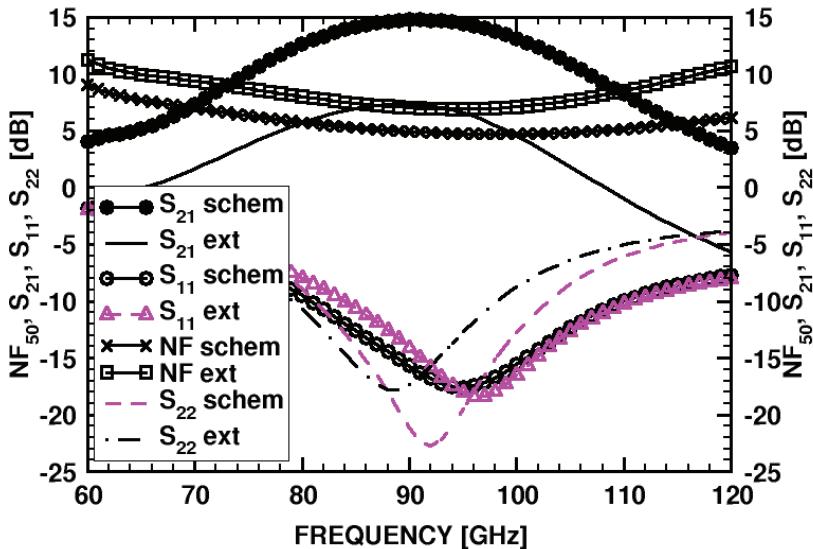


Fig. 10. Impact of transistor layout RC-parasitics on 90GHz 65nm LP-CMOS amplifier gain and noise figure degradation.

5. Doppler radar and active imaging transceivers

Fig. 11 illustrates a generic mm-wave transceiver block diagram suitable for multi-gigabit radio, ACC radar, and active imaging applications. Using lumped inductors and transformers as tuning and matching elements, such a system can be realized in a silicon area smaller than 2 mm^2 [4],[18],[28]. Large receiver arrays sharing a fundamental or second harmonic VCO and PLL, as in Fig. 12, are needed for remote sensing. For robust operation over process, temperature and power supply variation, the PLL should be implemented with a static frequency divider chain. To be practical, these SOCs must first overcome the

cross-talk between adjacent transceivers, the leakage from the transmitter to the receiver, large 1/f noise at sub-MHz offsets from the carrier, and large power dissipation, particularly in the VCO and PLL blocks. To contain the power dissipation at acceptable levels, particularly in imagers, all mm-wave building blocks should be powered from 2.5V or lower supplies.

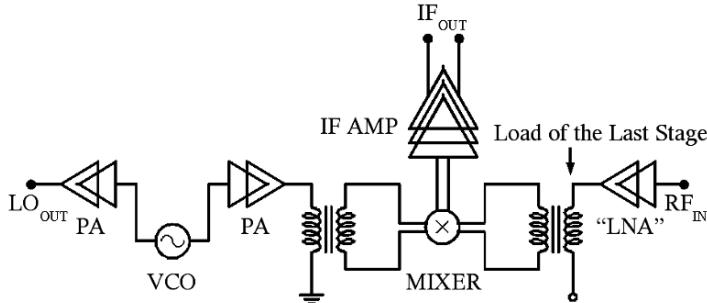


Fig.11. Block diagram of a generic SiGe BiCMOS or 65nm CMOS 80/160GHz transceiver for automotive radar and active imaging applications.

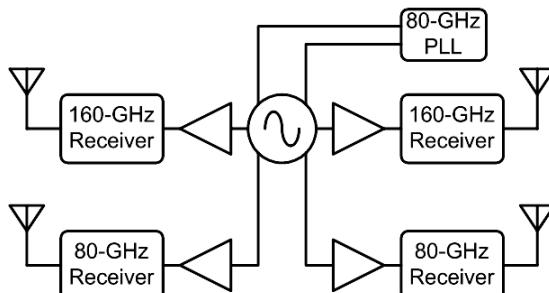


Fig.12. Block diagram of a generic SiGe BiCMOS or CMOS 80/160GHz receiver array for passive imaging applications.

The ACC radar has been the first mm-wave application to draw the attention of SiGe technology foundries due to its potentially large volume and relatively stringent requirements for output power and phase noise, which cannot be easily satisfied in CMOS. A system breakout with separate transmitter and receiver dies has been preferred [1], with the antenna placed on the board or in the package. Fig. 13 illustrates a 5V, 77GHz transmitter implemented in 225GHz SiGe HBT technology which consumes 2.8W and features a VCO, a variable-gain amplifier, a 16dBm power amplifier, an auxiliary power amplifier, and a dynamic frequency divider [1]. A companion receiver chip consists of a high-linearity doubly-balanced Gilbert-cell mixer with common-base RF input stage

and t-line baluns at the RF and LO ports for single-ended to differential conversion. Single-chip transceiver arrays with on-die antennas, not applicable to the ACC radar, were also reported [24]. They require sophisticated packaging to increase antenna gain [24], thus offsetting the cost advantage and the rationale of having on-chip antennas.

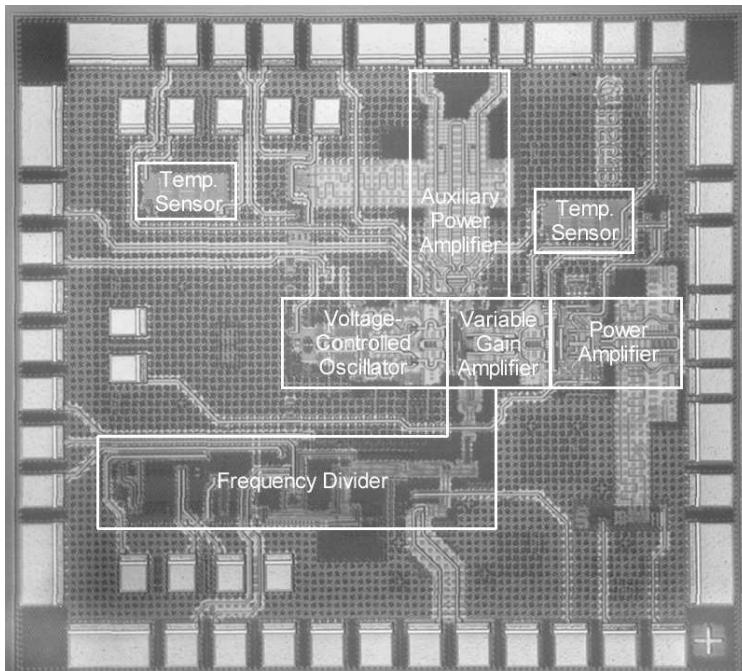


Fig.13. Die photograph of a 77GHz transmitter implemented in SiGe HBT technology courtesy of Infineon Technologies [1].

6. Comparison of SiGe HBT, SiGe BiCMOS and CMOS mm-wave IC building blocks

HBTs and MOSFETs have similar small signal and noise equivalent circuits at mm-wave frequencies. Therefore, the same circuit topologies and circuit design methodologies, relying on constant current density biasing schemes at the characteristic current densities (minimum NF_{MIN} bias, J_{OPT} , peak f_{MAX} , or peak f_T bias) apply to LNAs, PAs, VCOs and CML logic gates implemented with MOSFETs or HBTs [32],[35]. At frequencies above 60 GHz, the input impedance and the noise impedance of MOSFETs and HBTs, or of cascode topologies with HBTs and MOSFETs, described by (2) and (3), become more

resistive due to the parasitic resistances associated with the base/gate and emitter/source regions, and due to the decreasing reactance.

$$Z_{IN}(MOS) = R_S + R_G - j \frac{\omega_T}{\omega g_m} ; \quad Z_{IN}(HBT) = R_E + R_B - j \frac{\omega_T}{\omega g_m} \quad (2),$$

$$R_{SOPT}(MOS) \cong R_S + R_G + k \frac{\omega_T}{\omega g_m} ; \quad R_{SOPT}(HBT) \cong R_E + R_B + k \frac{\omega_T}{\omega g_m} \quad (3),$$

In (2) and (3), ω_T and g_m already include the impact of R_S or R_E , and are easily obtained from high frequency measurements or simulations, while k is a function of the degree of correlation between the input and output noise currents of the transistor, and is typically close to 0.5. The Miller effect is at least partly accounted for in (2) and (3) through ω_T , especially for cascode stages.

For example, for the 65nm MOSFET and SiGe HBT in Fig. 2, R_S+R_G and R_E+R_B are 3.5 Ω and 14 Ω , respectively. Suppose that these devices were to be sized for noise matching at 77 GHz to 40 Ω [21] and 33 Ω [18], respectively, to account for pad capacitances of 20 fF and 30 fF, respectively. A 16×65nm×1 μm MOSFET would be needed, biased for minimum noise at 2.5 mA, with the total series parasitics of 19 Ω , practically half of the optimum noise impedance. Similarly, the corresponding 2×0.13 μm ×3.75 μm HBT will be biased at 8 mA, with total series parasitics of 15 Ω , also about 50% of the optimum noise impedance. These two examples illustrate that transistor parasitics play a primary role at mm-wave frequencies, and that foundries must be able to control them tightly, which appears to be the case in both CMOS and SiGe BiCMOS technologies. In Colpitts VCOs, the increasingly resistive impedance of the transistor is compensated by connecting a high-Q MIM capacitor across the base-emitter or gate-source junction, improving the negative resistance and reducing the phase noise contribution from the resistive parasitics [16],[20].

Next, the experimental performance of 77 GHz LNAs, PAs, frequency dividers and VCOs implemented with HBT-only, MOS-HBT BiCMOS cascodes, and 90nm GP and 65nm LP CMOS transistors will be compared. All these circuits have state-of-the-art performance. The SiGe-HBT circuits were fabricated in a 0.13 μm SiGe BiCMOS production process, as well as in variants of this process with several HBT collector profile splits. This allowed drawing a direct correlation between the circuit performance and the HBT f_T and f_{MAX} . The SiGe HBT f_T/f_{MAX} for the technology splits are listed in Table 1. Measurement results are reported for wafer 5, except where indicated. The comparison with CMOS is only carried out for LNAs [4],[21] and VCOs [20] because at the time of

writing, there are no reported CMOS PAs and static frequency dividers operating at 80 GHz or above.

TABLE 1. TECHNOLOGY SPLIT PROCESS PARAMETERS.

Wafer #	f _T (GHz)	f _{MAX} (GHz)	Collector doping	Emitter width
5	250	290	Reference=C	0.13 μ m
3	245	280	C+	0.13 μ m
7	265	255	C++	0.13 μ m
2	260	240	C+++	0.13 μ m
6	170	210	Production BiCMOS9	0.13 μ m
7	150	160	Production BiCMOS9	0.17 μ m

6.1. Amplifiers

In the design of the SiGe-HBT LNA shown in Fig. 14, a 3-stage topology was chosen, which consists of two CE stages followed by a cascode stage [18]. The CE stages allow for 1.2-1.8V operation and minimize the overall noise figure of the LNA, while the cascode stage provides higher gain and is biased from a 1.8-2.5V supply. The input is simultaneously noise and impedance matched using the techniques described in [21],[32]. The LNA consumes a total of 40(60)mW from 1.5(1.8)V and 1.8(2.5)V supplies. The simulated noise figure, gain and input return loss are 5.3dB, 20dB, and -40 dB, respectively. Fig. 15 compares the measured and simulated gain and input return loss for 1.8V and 2.5V supplies at 25 C and 125 C, showing excellent performance, with less than 3dB gain degradation at 77 GHz and 125C, and the input return loss better than -12 dB from 78 GHz to 95 GHz. The 3-dB bandwidth extends from 77GHz to 90GHz with the highest gain of 19dB centered at 86GHz while S₁₂ is better than -50 dB. Because a standalone down-converter was not available for W-band noise measurements, only the noise figure of a mixer test structure was measured at this time. This was 12.5 dB at 73 GHz, close to simulations [18], indicating that the simulated 5.3 dB noise figure value of the LNA is also realistic.

The schematics of the 65nm LP CMOS LNA is shown in Fig. 16 and consists of 3-cascode stages with inductive broadbanding [21]. As in the SiGe HBT LNA, the input stage is simultaneously noise and impedance matched. The measured and simulated S parameters, shown in Fig. 17, demonstrate a peak gain of 9 dB at 80 GHz when the amplifier is powered from a 2.2V supply and consumes 40mW. The simulated noise figure is 7dB. The large V_{DD} is imposed by the fact that the LVT, 65nm LP n-MOSFET requires a V_{GS} of 0.9 V (similar to the V_{BE} of a SiGe HBT) at peak f_T bias. Finally, the measured gain of the SiGe HBT and CMOS LNAs are compared in Fig. 18. Even the production SiGe HBT with an

f_T of 170 GHz provides more gain than the 65nm LP CMOS one, while dissipating similar power.

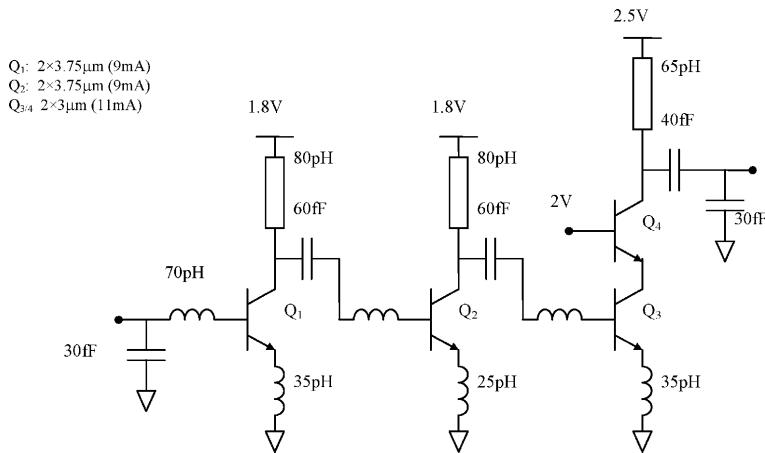


Fig.14. Schematics of 80GHz, SiGe HBT LNA [18].

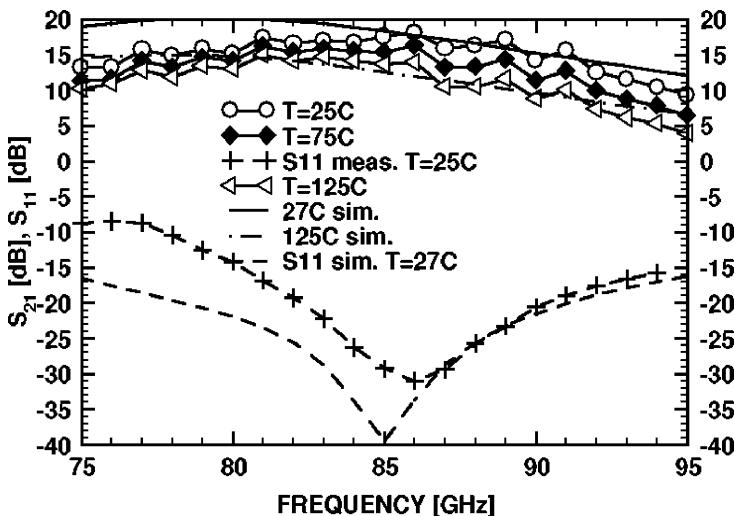


Fig.15. Measured (symbols) vs. simulated (lines) S parameters of 80GHz, SiGe HBT LNA at different temperatures.

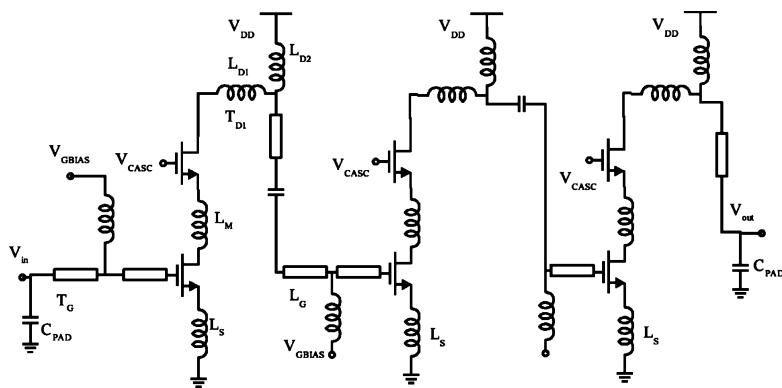


Fig.16. Schematics of 80GHz, 65nm CMOS LNA [4].

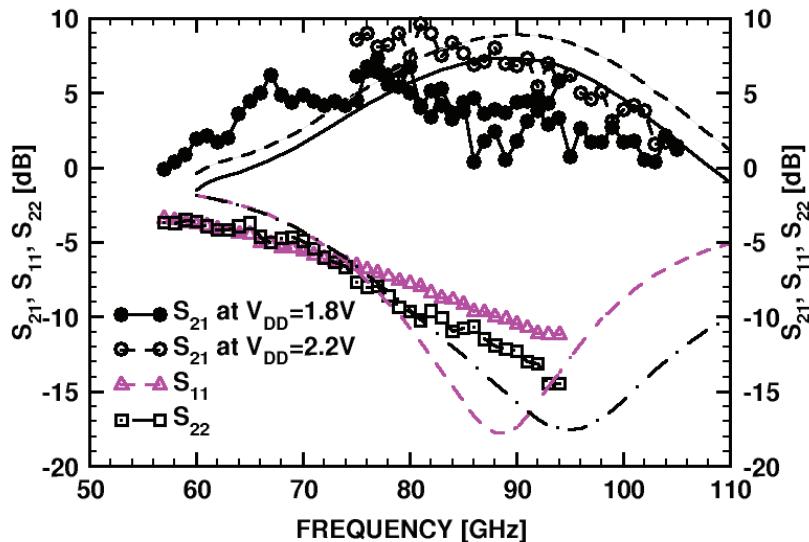


Fig.17. Measured (symbols) vs. simulated (lines) S parameters of 80GHz, 65nm CMOS LNA for different supply voltages.

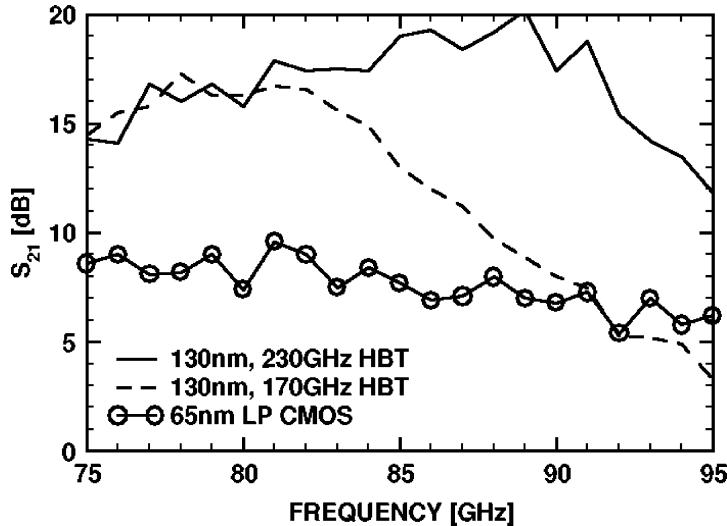


Fig. 18. Measured gain of 80GHz SiGe HBT and 65nm LP CMOS 3-stage LNAs.

Fig. 19.a. reproduces the schematics of a single-ended, 3-stage SiGe-HBT power amplifier consisting of a cascode stage - for large gain and powered from 2.5V - followed by two CE stages, for maximum power-added efficiency (PAE) and 1.8V supply [18]. Transistor sizes and currents increase by a factor of 2 from stage to stage toward the output. The CE stages have no inductive degeneration and are biased in class AB mode to maximize the saturated output power. For comparison, Fig. 19.b. describes a single-stage differential output buffer employing 130nm MOS-HBT cascodes and operating in the 85-90 GHz range. The latter draws 80mA from 2.5-3.3V supplies.

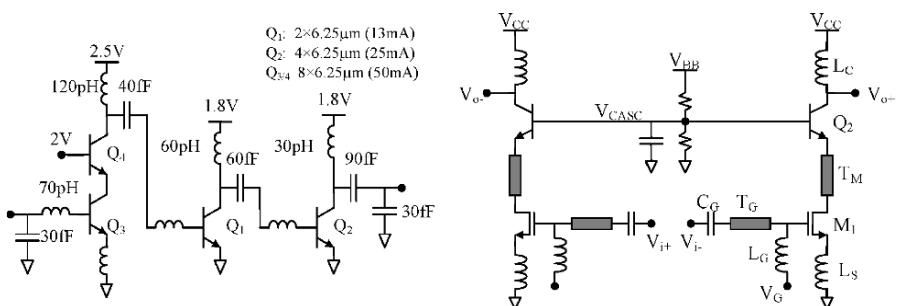


Fig. 19. a) Schematic of a 3-stage SiGe HBT PA. It shows three stages: a cascode stage (Q1, Q3) with 2.5V supply and 120pH/40fF load; a common-emitter stage (Q1, Q2) with 1.8V supply and 60pH/60fF load; and another common-emitter stage (Q2, Q4) with 1.8V supply and 90pH/30fF load. b) Schematic of a single-stage MOS-HBT cascode output buffer. It features a cascode pair (T_M, M_1) with transistors Q_2 and Q_3, and a driver stage (T_G, L_G, L_S) with transistors Q_1 and Q_2. Biasing and feedback components like V_BB, V_CASC, V_i+, V_i-, V_G, and V_CC are also shown.

The gain and output power of the first PA were measured from 75GHz to 95GHz over temperature up to 125 C, and over the 5 wafer splits [18]. At 77GHz, the PA achieves S_{21} of 19 dB, saturated output power of +14.6 dBm, and PAE of 15.5% (based on 161mW P_{DC}). The measured gain of the single-stage differential output buffer which employs 130nm MOS-HBT cascodes, is plotted in Fig. 20 along with that of the 3-stage PA measured for two wafer splits. The single-stage output buffer exhibits higher gain in the 85-89 GHz range than the 3-stage PA realized with 170GHz HBTs, and delivers +10.5 dBm differentially at 87 GHz. This is the first power amplifier operating above 60 GHz that uses MOSFETs, and it demonstrates once again [29],[35] that, by combining MOSFETs and HBTs at high frequencies, one can obtain better performance than that of the corresponding MOS-only or HBT-only circuits.

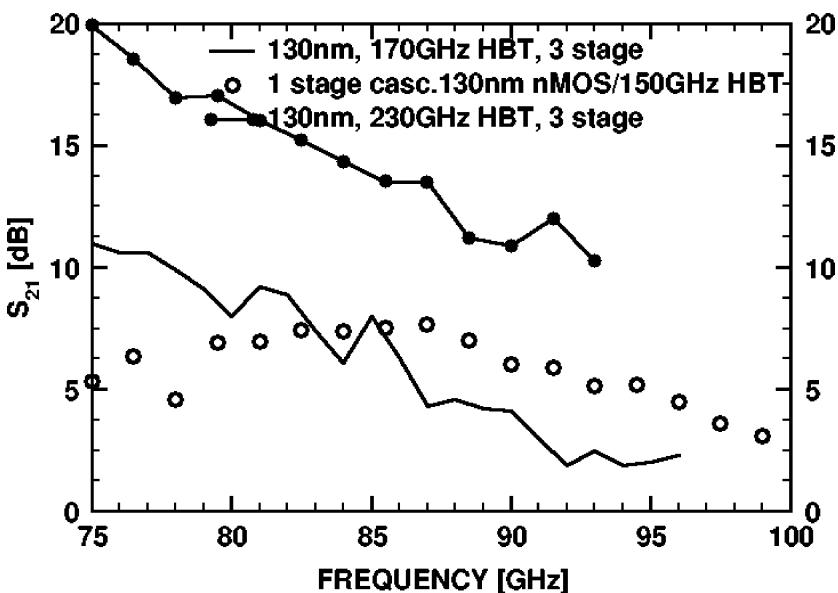


Fig.20. Measured gain of 80GHz SiGe HBT 3-stage PAs and of a 1-stage BiCMOS cascode differential PA.

6.2. Static Frequency Dividers

To verify the viability of a robust, fundamental frequency PLL at 80 GHz, a divide-by-64 static frequency divider chain, based on the low-power, 3.3V SiGe HBT topology in [17] was fabricated. The die photo is reproduced in Fig. 21. The divider was tested over temperature from 25 C to 125 C. The output spectrum, measured for a 77GHz input at 125 C is shown in Fig. 22. The self-

oscillation frequency (SOF) was measured on each of the 5 wafer splits and is plotted in Fig. 23 along with the gains of the 89GHz LNA and PA [18], 140GHz amplifier [28], and with the downconversion gain of an 80GHz Gilbert-cell, mixer [18], as a function of the SiGe HBT f_{MAX} (a different f_{MAX} for each wafer). Remarkably, in all cases, the best circuit performance is obtained for the wafer split with the highest f_{MAX} . Since only the SIC implant was changed in these wafer splits, the f_{MAX} on each wafer degrades as the f_T is improved. There is thus no ambiguity that f_{MAX} rather than f_T is the more important transistor figure of merit for mm-wave ICs. In a separate experiment to be described elsewhere, the noise figure of the 77GHz mixer also improves with the HBT f_{MAX} .

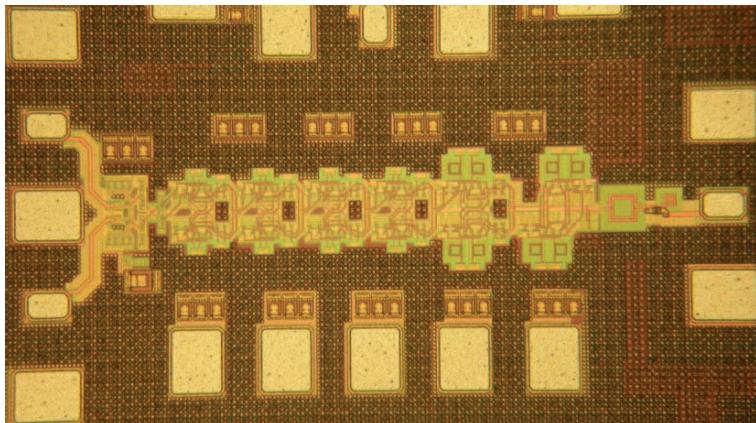


Fig.21. Die photo of 100GHz divide-by-64 chain.

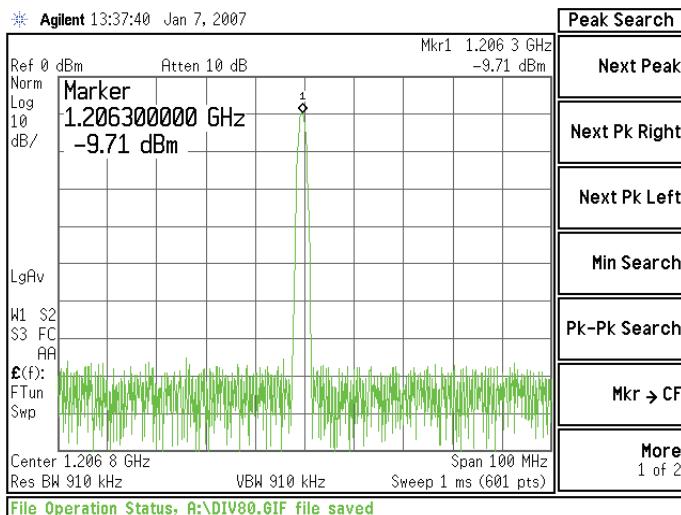


Fig.22. Measured output of divide-by-64 chain for a 77GHz input at 125 C.

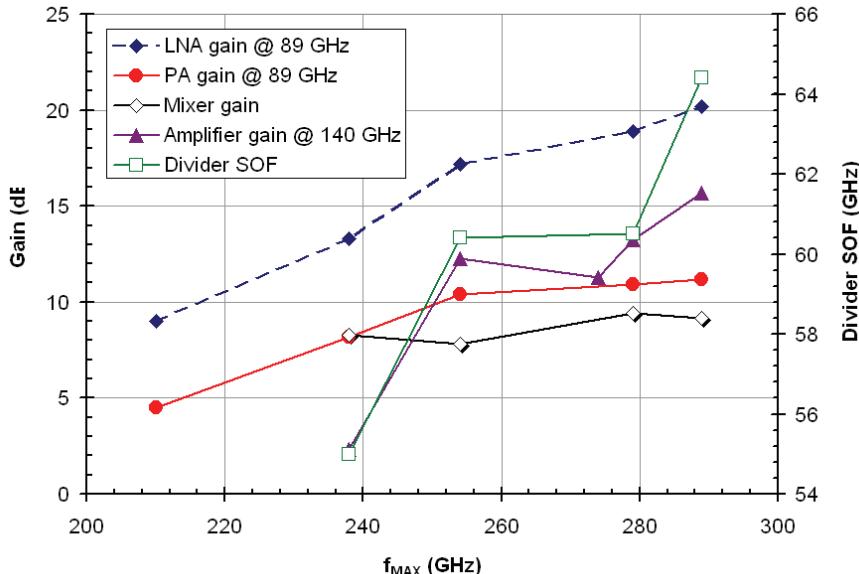


Fig. 23. Measured 80GHz SiGe HBT LNA, PA and static frequency divider SOF across wafer splits as a function of the f_{MAX} of SiGe HBT on each wafer.

6.3. VCOs

Record low-phase-noise Colpitts VCOs (Fig. 24) were implemented with SiGe HBTs [16] (Fig. 24) and 90nm GP MOSFETs [20] (Fig. 25) for operation in the 77GHz to 105GHz range. The oscillation frequency (f_{osc}) of the VCO is given by (4). In line with the VCO design methodology outlined in [32], the tank inductance (L_B) is chosen as the smallest realizable inductance with hi-Q, or about 25pH for the HBT, and 50pH for the CMOS technologies. Thus, C_{EFF} is fixed by the desired oscillation frequency. In reality, C_π (or C_{GS}) is much greater than C_{VAR} , and consequently, $C_{EFF} \approx C_{VAR}$ for design purposes.

$$f_{osc} = \frac{1}{2\pi\sqrt{L_B C_{EFF}}} \quad C_{EFF} = \frac{C_{VAR}(C_1 + C_\pi)}{C_{VAR} + C_1 + C_\pi} \quad (4)$$

The negative resistance provided by Q_1 , given by (5), must be large enough to overcome losses in the tank and the base/gate and emitter/source resistance. In the W-band, the finite Q of the varactor (C_{VAR}) and of the base/gate inductance (L_B) adds substantial losses to the tank.

$$R_{NEG} \cong R_B + R_E + \frac{\omega_{osc} L_B}{Q_{LB}} + \frac{1}{\omega_{osc} Q_{CVAR} C_{VAR}} - \frac{g_m}{\omega_{osc}^2 (C_\pi + C_1) C_{VAR}} \quad (5)$$

Capacitor C_1 is important in minimizing the oscillator phase noise, vital in radar applications.

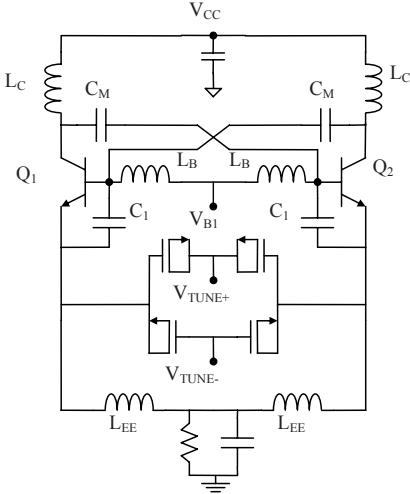


Fig.24. Schematics of 100GHz SiGe BiCMOS VCOs [16].

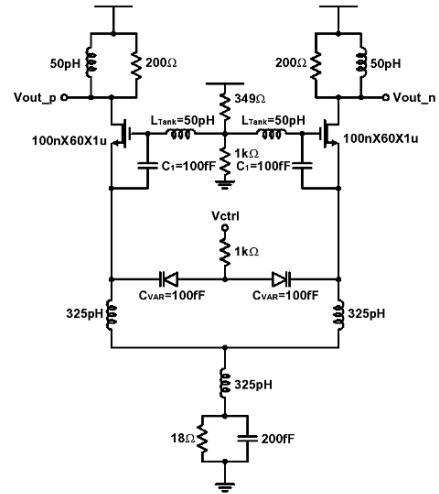


Fig.25. Schematics of 79GHz 90nm GP CMOS VCO [20].

Record phase noise values of -101.3 and -100.2 dBc/Hz, respectively, were measured at 1MHz offset from the 105 GHz SiGe HBT VCO carrier (Fig. 26) and from the 79GHz carrier of the CMOS VCO, as needed in imaging and ACC radar applications. However, the output power is at least 18 dB higher for the SiGe HBT VCO while its power consumption is only 4 times larger: 120mW vs. 30mW. The measured tuning characteristics of the CMOS VCO are very linear, spanning 73 to 79 GHz. The fact that the ITRS FoM for VCOs excludes output power explains why CMOS VCOs rate very highly using this figure of merit. However, in many applications, a mm-wave VCO with low output power would require amplification before becoming useful. A better design strategy is to dissipate greater power in the VCO core, which reduces the overall VCO complexity by eliminating amplifier stages. Furthermore, increased core power dissipation can ultimately improve phase noise, whereas amplifying stages do nothing to improve phase noise.

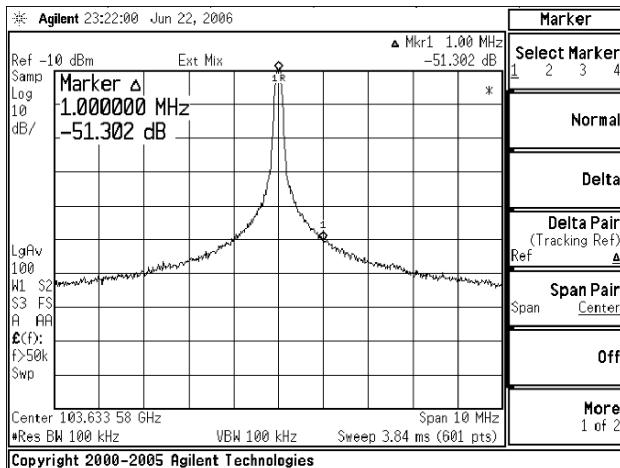


Fig.26. Measured Phase Noise of 105GHz SiGe HBT VCO [16].

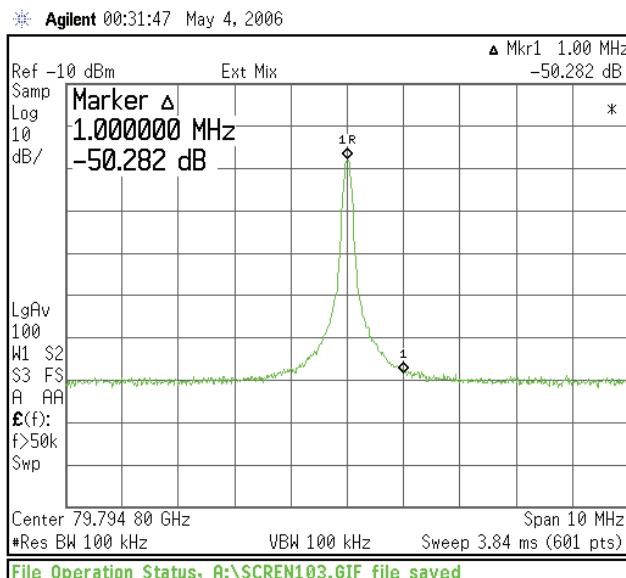


Fig.27. Measured Phase Noise of 79GHz 90nm GP CMOS VCO [20].

7. Conclusions

As a result of the larger breakdown voltage, transconductance, and f_T , and because of their reduced sensitivity to layout parasitics and temperature variation when compared to 65nm CMOS, SiGe HBTs and SiGe BiCMOS

technologies have established a clear advantage and a strong foothold at mm-wave frequencies. They are set to seriously challenge the supremacy of III-V transistors in all but the lowest noise astronomy applications. While showing good promise for WLAN applications at 60 GHz, 90nm GP and 65 nm LP CMOS technologies do not have adequate performance for most IC building blocks required in 77 GHz ACC systems. However, this situation may change in the 45nm node. By applying constant-field scaling rules to inductors and transformers, and design methodologies that have proven successful at GHz-frequencies, it is now possible to integrate 80GHz and 160GHz transceiver arrays on a silicon die and thus bring economies of scales, typical of silicon, to a variety of sensors for security, remote sensing, imaging and automotive radar applications at and beyond 80 GHz.

Acknowledgments

This work was funded by CITO and STMicroelectronics. We would also like Bernard Sautreuil for his support, CMC and Jaro Pristupa for CAD tools. Test equipment was provided by OIT, CFI and ECTI.

References

- [1] H. Knapp et al., "SiGe Circuits for Automotive Radar," *IEEE SiRF Digest*, pp.231-236, Jan. 2007.
- [2] J.J. Lynch, "Low Noise Direct Detection Sensors for Millimeter Wave Imaging," *IEEE CSICS Digest*, pp. 215-218, Nov. 2006.
- [3] H. Kim et al., "SiGe IC-based mm-wave imager," in press, *IEEE ISCAS*, May 2007.
- [4] S.P. Voinigescu *et al.*, "CMOS SOCs at 100 GHz: System Architectures, Device Characterization, and IC Design Examples," in press, *IEEE ISCAS*, May 2007.
- [5] T. Kosugi, et al. "120-GHz Tx/Rx Waveguide Modules for 10-Gb/s Wireless Link System," *IEEE CSICS Digest*, pp. 25-29, Nov. 2006.
- [6] H. Li and H.-M. Rein, "Millimeter-wave VCOs with wide tuning range and low phase noise, fully integrated in a SiGe bipolar production technology," *IEEE JSSC*, vol. 38. pp.184-189, Feb. 2003.
- [7] W. Prendl, et al., "A low-noise and high-gain double-balanced mixer for 77GHz automotive radar front-ends in SiGe bipolar technology," *IEEE RFIC Symp. Digest*, pp.47-50, June 2004.
- [8] U.R. Pfeiffer et al., "A 77 GHz SiGe power amplifier for potential applications in automotive radar," *IEEE RFIC Symp. Digest*, pp.91-94, June 2004.
- [9] B. Floyd, "V-band and W-band SiGe bipolar low-noise amplifiers and voltage-controlled oscillators," *IEEE RFIC Symp. Digest*, pp.295-298, June 2004.

- [10] J-J. Hung, et al., "A 77GHz SiGe Sub-Harmonic Balanced Mixer," *IEEE JSSC*, vol. 40, pp.2167-2173, Nov. 2005.
- [11] Dehlink B. et al., Low-noise Amplifier at 77 GHz in SiGe:C bipolar technology, *IEEE CSICS Digest*, pp. 287, 2005.
- [12] P. Roux *et al.*, "A monolithic integrated 180 GHz SiGe HBT Push-Push Oscillator," *IEEE EGAAS Digest*, pp. 341-343, 2005.
- [13] L. Wang, et al., "An Improved Highly-Linear Low-Power Down-Conversion Micromixer for 77GHz Automotive Radar in SiGe Technology," *IEEE MTTs (IMS)*, San Francisco, June 2006.
- [14] R. Reuter and Y.A. Yin., "A 77GHz (W-Band) SiGe LNA with a 6.2 dB Noise figure and Gain Adjustable to 33 dB," *IEEE BCTM Digest*, pp.134-137, Oct. 2006.
- [15] A. Komijani and A Hajimiri., "A Wideband 77-GHz, 17.5-dBm Fully Integrated Power Amplifier in Silicon," *IEEE JSSC*, vol. 41, pp. 1749-1756, Aug. 2006.
- [16] S.T. Nicolson, et al., "Design and Scaling of SiGe BiCMOS VCOs Operating near 100 GHz," *IEEE BCTM Digest*, pp.142-145, Oct. 2006.
- [17] E. Laskin, et al., "Low-Power, Low-Phase Noise SiGe HBT Static Frequency Divider Topologies up to 100 GHz," *IEEE BCTM*, pp.235-238, Oct. 2006.
- [18] S.T. Nicolson et al., "A 2.5V 77GHz Automotive Radar Chipset," in press, *IEEE IMS*, June 2007.
- [19] K.K. O, *et al.*, "CMOS Millimere-Wave Signal Sources and Detectors," in press, *IEEE ISCAS*, May 2007.
- [20] K.W. Tang, et al., "Frequency Scaling and Topology Comparison of mm-wave CMOS VCOs," *IEEE CSICS Digest*, pp.55-58, Nov. 2006.
- [21] S.T. Nicolson and S.P. Voinigescu, "Methodology for Simultaneous Noise and Impedance Matching in W-Band LNAs," *IEEE CSICS Digest*, pp. 279-282, Nov. 2006.
- [22] B. Heydari et al., "Low-Power mm-Wave Components up to 104GHz in 90nm CMOS," , *IEEE ISSCC*, Digest, pp.200-2001, Feb. 2007.
- [23] W. Mayer, M. Meilchen, W. Grabherr, P. Nuchter, and R. Guhl, "Eight-channel 77 GHz front-end module with high-performance synthesized signal generator for FM-CW sensor applications," *IEEE Trans. MTT*, vol. 52, no. 3, pp. 993-1000, Mar. 2004.
- [24] A. Babakhani, et al., "A 77-GHz Phased-Array Transceiver With On-Chip Antennas in Silicon: Receiver and Antennas" *IEEE JSSC*, vol. 41, no. 12, pp. 2795-2806, Dec. 2006.
- [25] L. Wang, J. Borngraeber, and W. Winkler, "77 GHz Automotive Radar Receiver Front-end in SiGe:C BiCMOS Technology," *ESSCIRC 2006*, pp. 388-391.

- [26] C. Wagner, A. Stelzer, and H. Jager, "PLL architecture for 77-GHz FMCW Radar Systems with Highly-Linear Ultra-Wideband Frequency Sweeps," *IEEE IMS Digest*, pp. 399-402, June 2006.
- [27] B. Dehlink et al., "An 80 GHz SiGe Quadrature Receiver Frontend," *IEEE CSICS*, pp. 197-200, Nov. 2006.
- [28] E. Laskin et al. "80/160GHz Transceiver and 140GHz Amplifier in SiGe Technology," in press *IEEE RFIC Symp.* June 2007.
- [29] S.P. Voinigescu et al., "SiGe BiCMOS for Analog, High-Speed Digital and Millimetre-Wave Applications Beyond 50 GHz," *IEEE BCTM Digest*, pp.223-230, Oct. 2006.
- [30] P. Chevalier et al, "Advanced SiGe BiCMOS and CMOS platforms for Optical and Millimeter-Wave Integrated Circuits," *IEEE CSICS* 2006, pp. 12-15, Nov. 2006.
- [31] P. Chevalier et al, "230-GHz Self-Aligned SiGe:C HBT for Optical and Millimeter-Wave Applications," *IEEE JSSC*, vol. 40, no.10, pp. 2025-2034, Oct. 2005.
- [32] S.P. Voinigescu, et al., "RF and Millimeter-Wave IC Design in the Nano-(Bi)CMOS Era," in *Si-Based Semiconductor Components for Radio-Frequency Integrated Circuits (RFIC)*, 2006, W. Z. Cai, Ed. New Delhi, India: Transworld Research Network, ISBN 81-7895-196-7, 2006.
- [33] S.P. Voinigescu et al., "A Comparison of Silicon and III-V Technology Performance and Building Block Implementations for 10 and 40 Gb/s Optical Networking ICs," *IJHSES*, Vol.13, No.1, and book chapter in *Compound Semiconductor Integrated Circuits*, pp.27-58, 2003.
- [34] P. Chevalier et al, "High-Speed SiGe BiCMOS Technologies: 120-nm Status and End-of-Roadmap Challenges," *IEEE SiRF Digest*, pp.18-23, Jan. 2007.
- [35] T.O. Dickson et al., "The Invariance of Characteristic Current Densities in Nanoscale MOSFETs and its Impact on Algorithmic Design Methodologies and Design Porting of Si(Ge) (Bi)CMOS High-Speed Building Blocks," *IEEE JSSC*, pp.1830-1845. Aug. 2006.
- [36] T.H. Lee, "The Design of CMOS Radio-Frequency Integrated Circuits" Cambridge, 2nd. Edition, 2004.
- [37] B. Floyd et al., "Silicon Millimeter-Wave Radio Circuits at 60-100 GHz," *IEEE SiRF Digest*, pp.213-218, Jan. 2007.
- [38] A. Mangan et al., "De-Embedding Transmission Line Measurements for accurate Modelling of IC Designs," *IEEE Trans. Electron. Dev.*, Vol. ED-53, pp.235-241, No.2, 2006.
- [39] T.O. Dickson, et al., "30-100 GHz Inductors and Transformers for Millimeter-wave (Bi)CMOS Integrated Circuits", *IEEE Trans. MTT*, Vol.53, No.1, pp.123-133, 2005.

A Comparison of CMOS and BiCMOS mm-Wave Receiver Circuits for Applications at 60GHz and Beyond

Sharon Malevsky and John R. Long
Electronic Research Laboratory / DIMEs
Delft University of Technology, the Netherlands
s.malevsky@tudelft.nl

Abstract

Implicit in the design of a 60GHz receiver front-end is the requirement for broadband amplification and gain flatness in order to support advanced modulation schemes. As we have seen from lower frequency wireless systems, the choice of technology often dictates the system architecture. While either silicon CMOS or silicon Bipolar technology could be used for implementation, cost, availability, performance and time to market constraints typically dictate the technology choice. The trade-offs between CMOS and Bipolar/BiCMOS technologies for mm-wave receiver circuits are outlined in this paper. Realizing low-noise performance and gain flatness over a wide bandwidth is used as a case study from both the circuit and system points of view.

1. Introduction

Continued scaling of CMOS and BiCMOS technologies will eventually allow RFIC designers to exploit the unlicensed 7GHz of free spectrum around 60GHz. The proposed 802.15.3c standard for short-range, gigabit/s communication in the unlicensed bands at 60GHz may be used an adapter for any gigabit/s data source or sink requiring a short-range wireless link. The spectral allocations vary with region as shown in the following table.

The small physical size of passive circuit elements at mm-wave frequencies enables more cost-effective integration of distributed elements such as antennas or baluns. A second benefit of mm-wave operation is immunity to interferers due to attenuation (see Fig. 1) of the transmitted signal

Table 1: Unlicensed Bands

Region	Unlicensed Band
USA	57 – 64 GHz
Europe and Japan	59 – 66 GHz

at 60GHz caused by absorption of the RF energy by oxygen (O_2). This natural immunity to interferers can relax the demand on the front-end circuit linearity, thereby conserving current and power consumption. Early

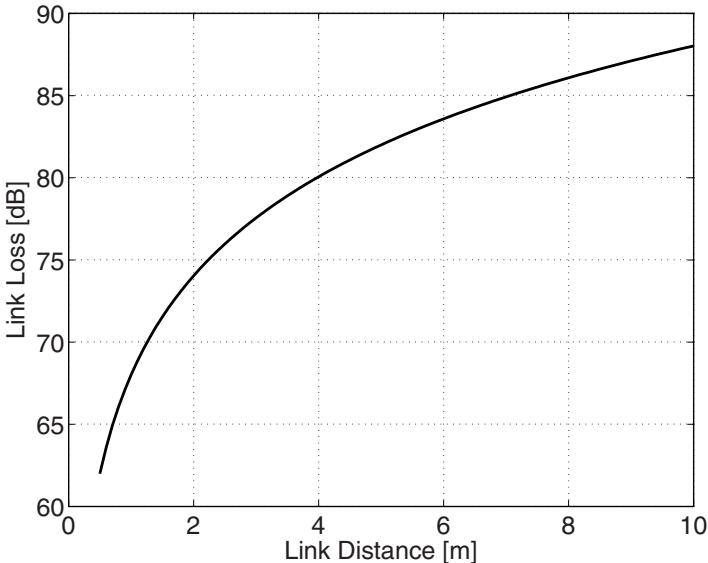


Figure 1: 60GHz Link loss according to Friis' prediction for an isotropic radiator and receiver [1]

demonstrations of 60GHz communication transceivers used silicon bipolar devices in an advanced BiCMOS technology [2, 3]. However, CMOS examples of 60GHz receiver front-ends in both $0.13\mu m$ [4–6] and 90nm [7] technologies have also been reported. The potential for lower cost and higher integration (e.g., higher packing density for digital circuits at the latest CMOS technology node) for CMOS relative to BiCMOS make it the preferred technology for a consumer application. With deep sub-micron CMOS devices now offering transit frequencies well above 100GHz, there is a strong motivation to focus research in this area on CMOS implementations. However, the choice of technology also influences the eventual transceiver architecture, because many of the common receiver architec-

tures cannot be implemented effectively in all technologies. This aspect is reviewed in Section 5. of this paper.

The 60GHz band is ideally suited to an indoor application, such as a wireless feedlink to a high definition TV, or other application that requires a relatively large channel bandwidth (i.e., 7GHz is available). Multipath interference resulting from non-line-of-sight propagation indoors of a wideband signal [8], may require adoption of an advanced modulation scheme, such as OFDM. Systems employing OFDM are sensitive to the gain flatness over bandwidth and linearity, especially for the transmitter power amplifier.

In this paper, a brief comparison of RF characteristics of CMOS and Bipolar devices is presented in the following section. The effect that the gain flatness constraint has on the receiver front-end circuit is then examined within the context of a broadband LNA design (see Section 3.). Finally, Section 4 the demands of a potential 60GHz design on the down-converting mixer and the LO source are considered.

2. BiCMOS/CMOS Technology and Passive Devices at 60GHz

Two factors that play a role in circuit design choices are the system specification and the capabilities of a given technology. The following section will review current deep sub-micron silicon VLSI technologies and their applicability to circuit design for the 60GHz band.

2.1 Passive Devices

Passive components that are currently used in commercial RFICs, such as MIM capacitors, inductors and transformers, are still useful for 60GHz designs. The relatively high operating frequency ($\frac{\lambda}{4} = 600\mu m$ on-chip) permits the use of transmission lines, which are easier to model accurately and design than spiral inductors or transformers. One consideration for the designer when dealing with such distributed elements, is to know the return path (e.g., the electrical ground). The use of a metal shield layers in order to avoid substrate losses is common. Realizing a well-defined ground for the shield constrains the design and its ultimate performance. The example shown in Fig. 2b illustrates the behavior of electric field lines close to a neighboring ground that may be more conductive and at lower potential. The resulting crosstalk changes the properties of the transmission line and affects the circuit behavior. On the other hand, symmetry defines the 0V

reference in a differential design (see Fig. 2c) and smart shield schemes that reduce the eddy currents effects [9] with lower parasitic capacitive and losses maybe implemented.

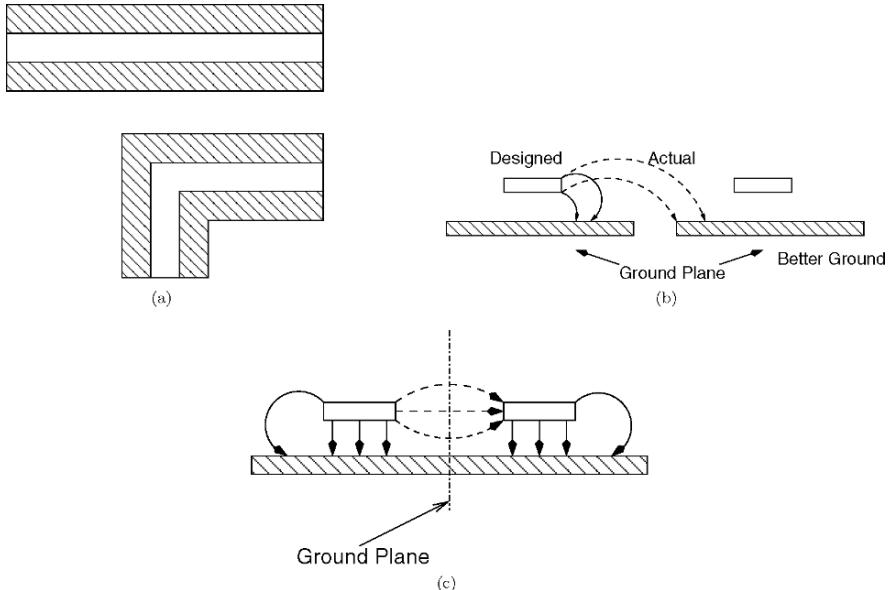


Figure 2: In the case of two close transmission lines (a) the electric field lines may be closed by the neighboring ground and thus influence the properties of the line. Differential design (c) has an implicit ground reference.

Furthermore, many foundries do not supply adequate support for passive components in their design kits. Thus, the design and modeling of passive components becomes the designer's task, and if the passives are not modeled accurately, circuit behavior becomes unpredictable. Unfortunately, quasi-static approximations, which are typically robust and relatively fast, are inaccurate when the passive device dimensions are greater than approximately $\lambda/10$. Full-wave EM simulators, such as EM-Sonnet [10] and ADS-Momentum (in microwave mode [11]) can provide more accurate results. In the past, one drawback of EM simulators was long simulation time, however, personal computers are now fast enough to permit simulation of multi-turn spirals or other complex component designs relatively quickly. In most cases, passives that consist of transmission lines are easier to simulate and model, and often yield a more robust design.

2.2 Active Devices

For the following comparison, state-of-the-art 90nm bulk CMOS 90nm and $0.13\mu\text{m}$ SiGe-BiCMOS technologies are compared.

One advantage of BJT (Fig. 3(b)), over an NMOS device (Fig. 3(a)) is simplified matching. This can be easily illustrated by examining the “passive gain”, or the improvement in the S_{21} gain (i.e., voltage gain in a 50Ω system) and the maximum available gain MAG) of the transistor (note that for the NMOS, MAG and maximum stable gain (MSG) are identical over the plotted frequency range). This advantage diminishes above 20GHz, as the gain of both devices rolls-off quickly at -20dB/decade.

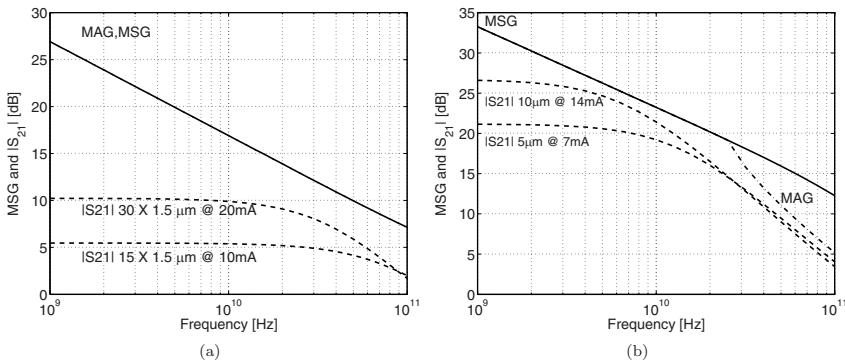


Figure 3: Maximum stable gain (MSG) and $|S_{21}|$ for (a) 45 μm wide and 22.5 μm wide 90nm NMOS transistors (1.5 μm wide fingers), and (b) 10 μm and 5 μm emitter length (130nm emitter width) BJTs. Both devices are biased at peak f_T .

However, there are others advantages of the BJT to consider. For example, the higher transconductance efficiency (i.e., bias current/transconductance ratio, I/gm) leads to reduced current consumption in a low-noise amplifier (LNA) design. Higher absolute gm is advantageous, but since voltage gain can be realized at baseband (i.e., after mixing), a higher mixer load impedance (rather than a larger gm) also increases gain, but without additional current consumption. However, there is a trade-off between bandwidth and load impedance at the mixer output to consider. For an active mixer design, the BJT has a clear advantage because of its lower flicker ($1/f$) noise corner frequency compared to a MOS transistor. The key parameter differences between the technologies are summarized in the following Table.

Table 2: Parameters Comparison Between a 90nm NMOS and 0.13m BJT

Parameter	MOS	BJT
I/gm	V_{eff}	V_T
C_{in}/C_μ	2.3	>10
1/f Corner	1GHz	1kHz
ro @ 5mA bias	220Ω	$6.5k\Omega$
Breakdown Voltage	1.1V	2-5V

Furthermore, when working with gigabit/s signals at baseband, a considerable portion of the power consumption comes from the digital circuitry (e.g., the modem, PLL, and I/Os). In order to minimize total power consumption and the chip area consumed by the digital circuits, a CMOS technology offering the shortest gate length possible is necessary.

3. Low Noise amplifiers

As mentioned in Sec. 1., gain flatness is very important when employing advanced modulation schemes such as OFDM. Hence, a key specification for the LNA is bandwidth, which, for an amplifier covering the entire 60GHz band implies a circuit that is quasi-broadband ($\frac{f_{center}}{Bandwidth} = \frac{61GHz}{7GHz} < 10$). One possible approach to designing a broadband LNA is to employ various methods to increase the bandwidth of a conventional narrowband low noise amplifier (LNA) topology. A second approach is to take a broadband circuit and narrow its bandwidth in an effort to optimize performance.

3.1 Bandwidth Enhancement

The cascode amplifier (Fig. 4a) is often used for narrowband LNA designs. The main advantage of this topology at high frequencies are:

- Increase in bandwidth by decreasing Miller capacitance
- Improved isolation
- Unconditional stability

Which is the best approach to take at 60GHz? Observing Fig. 4b, the Miller effect depends upon the ratio between g_m to c_{gd} , which at microwave frequencies is very small. However, at 60GHz $c_{gd} \approx 10fF$ and $g_m \approx 10m\text{U}$, which makes suppression of Miller effect relatively poor. Furthermore, the transimpedance stage adds capacitance at the interstage node, thereby degrading the gain-bandwidth and noise performance of the circuit. The

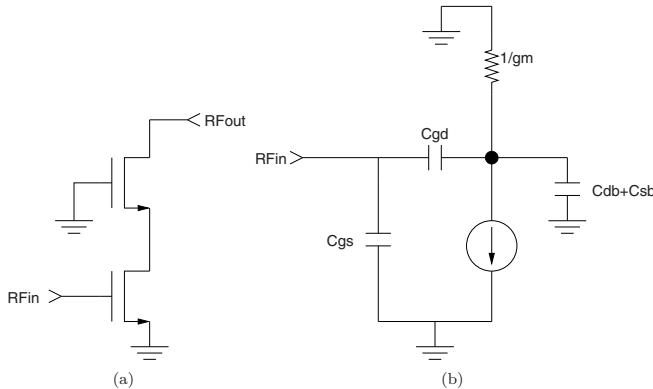


Figure 4: A CMOS Cascode stage (a) and the small signal equivalent circuit (b) of the transconductance stage

interstage pole may be moved to a higher frequency by a inductive peaking (e.g., using a transmission line) [12] as shown in Fig. 5. The inductor negates some of the capacitive parasitic effect, but a low-Q inductor is needed to realize a wide bandwidth, which degrades the noise performance. One common rationale for high frequency circuit choices is that “simple

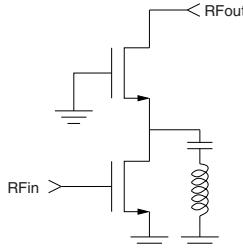


Figure 5: Cascode stage with a shunt peaking inductor

circuits are the best solution”. However, are the “**simple circuits**”, simple solutions? Consider the stability of a cascode amplifier. It depends upon realization of a very low (i.e., ideally zero) impedance at the gate of the transimpedance (common gate or common base) stage. At 60GHz this requirement becomes difficult to fulfill, because a $30\mu m$ interconnection at the common node input of the cascode may be enough in some cases to render the entire amplifier unstable. An alternative is applying negative feedback to broaden the bandwidth of an amplifier stage. However, with a maximum stable gain less than 10dB for a $90nm$ common-source stage at 60GHz, there is virtually no gain that may be traded for increased bandwidth. A third alternative is the parallel path or distributed amplifier.

3.2 Distributed Amplifier

The motivation to use a distributed amplifier (DA) as an LNA for a 60GHz application comes from its inherent broad bandwidth. This class of amplifiers can operate from DC to the mm-wave frequency band with a relatively flat gain. Designers often avoid distributed topologies due to their many disadvantages, such as, relatively high current consumption, poor noise performance, design complexity and low overall efficiency. However, the 60GHz band specifications require operation over 7GHz of bandwidth and not the entire spectrum, and therefore a trade-off can be made between operating bandwidth and other aspects of the amplifier's performance.

A common source (CS) topology for a DA is shown in Fig. 4. The amplifier consists of three CS stages which are distributed along a transmission line (TRL) at the input and at the output. The parasitic components of each stage are absorbed as part of the TRL, thereby extending the bandwidth of the amplifier. One of the drawbacks of a distributed topology is the large chip area required by the amplifier due to the physical size of the passive elements (inductors and/or transmission lines). This difficulty may be resolved if the bandwidth is restricted to the 57-64 GHz band, where the inter-stage transmission lines (typically a fraction of a wavelength long) are relatively small as the wavelength on-chip (λ) is less than 2.5mm. The LNA can drive a second-stage mixer with a relatively high (i.e., non- 50Ω) input, which permits narrow-width output transmission lines (with higher attenuation) that have lower parasitic capacitance.

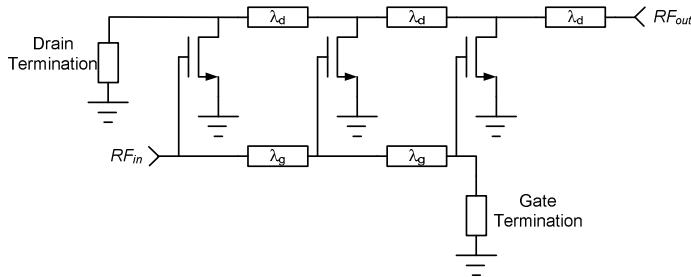


Figure 6: Distributed Amplifier based on common-source stages

Another drawback of the DA is low efficiency (AC power output/DC power consumed), which is often degraded by the use of a resistor as the drain termination in order to realize a broad bandwidth. An approximate gain calculation using the following expression from [13] gives,

$$G = \frac{g_m^2 Z_d Z_g}{4} \left[e^{\gamma_g l_g - N \gamma_d l_d} \sum_{n=1}^N e^{-n(\gamma_g l_g - \gamma_d l_d)} \right]^2 \quad (1)$$

where G is the amplifier gain, Zd and Zg are the drain and gate transmission line characteristic impedances, respectively, d and g are the respective drain and gate line propagation coefficients, ld and lg are the drain and gate line lengths, gm is the transistor transconductance and N is the number of gain stages (i.e., 3 in Fig. 4). An example of this topology is presented in [14], where the current consumption is 132mA for a gain of 8dB. Examining equation 1, one can observe that the gain is reduced by a factor of 4 due to drain current which flows away from the output (i.e., towards the termination). Replacing the output line termination with a TRL (e.g., close to $\lambda/4$), can produce a higher impedance at the termination, and more current will flow toward the load. This solution is band-limited. However, DC-to mm-wave bandwidth, which requires a broadband output line termination, is not required as previously discussed. If a transmission line is used to replace the resistive termination, the impedance at each drain node can be written as

$$Z_{T_n} = j Z_d \tan [2\text{Im}(\gamma_d)(l_T + (n-1)l_d)], \quad (2)$$

where lt is the terminating transmission line length. Substituting equation (2) into equation (1) results in the modified gain expression:

$$G = g_m^2 Z_d Z_g \left[e^{\gamma_g l_g - N \gamma_d l_d} \sum_{n=1}^N \frac{Z_{T_n}}{Z_{T_n} + Z_d} e^{-n(\gamma_g l_g - \gamma_d l_d)} \right]^2 \quad (3)$$

As seen from equation (3), the current flowing toward the load is now double the previous result, giving a 6dB increase in the in-band gain. In Fig. 7, the gain of the modified DA is plotted versus an identical design using a resistor termination. The gain in the vicinity of 60GHz is 6dB higher, however, the amplifier bandwidth is now narrower by an order of magnitude. The resistive gate termination also contributes to the relatively poor noise performance (e.g., noise figure) of a conventional DA. Replacing it by a transmission line termination may result with a narrowband matching at the driving stage, or even radiate energy (as an antenna) when used in a 60GHz LNA. In [15], an active termination is suggested. The termination

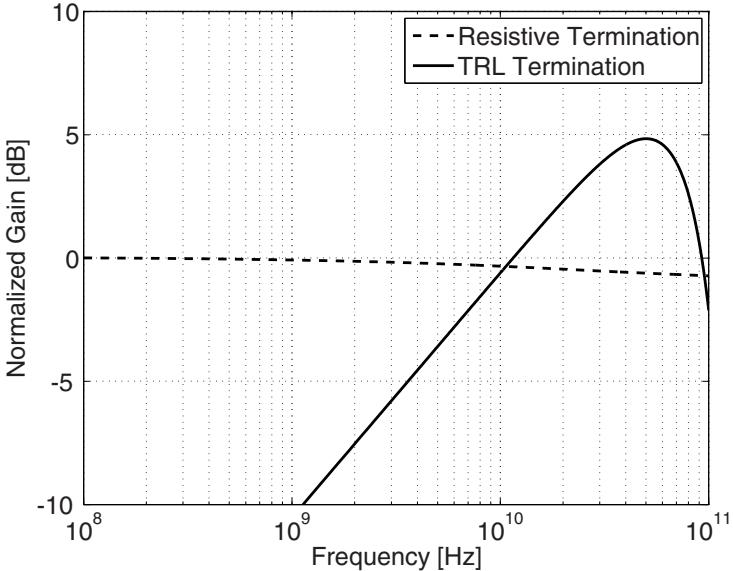


Figure 7: The Normalized gain of a CS-DA compared with a modified DA with a TRL termination

resistor is replaced by an active amplifier (e.g., a common-source stage with feedback, or a common-gate stage), where the amplifier input impedance provides a broadband termination for the input (i.e., gate) transmission line. This approach is also band-limited, which as mentioned before is not a problem for a 60GHz LNA.

4. Mixers

Implementation of circuits at mm-wave frequencies requires short-channel (CMOS) or narrow-base (BJT) devices. These devices pose a limitation on the supply voltage (e.g., 0.9-1.2V for 65-90nm CMOS). Another constraint imposed by a low-voltage supply is circuit linearity, which affects the ability of the receiver to operate properly when interfering signals are present. A logical choice for the mixer topology is the simple switching quad, as shown in Fig. 8. The mixer can be operate as a passive (i.e., resistive) mixer if no drain-source bias voltage is applied (i.e., $V_{DS} = 0V$), or it may be biased if more conversion gain is required. This circuit is commonly used as an RF mixer in both CMOS and BiCMOS technologies. However, for deep submicron technologies operating at high frequency, there is a large difference in mixer noise performance between the 2 technologies. The $1/f$

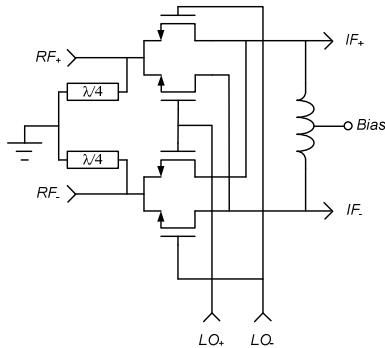


Figure 8: A Switching Quad mixer

noise produced by small-geometry CMOS devices is typically much higher than for a bipolar device, where the 1/f noise corner frequency may be as high as 1GHz for a gate width of a few microns at gate lengths of $0.13\mu m$ and below.

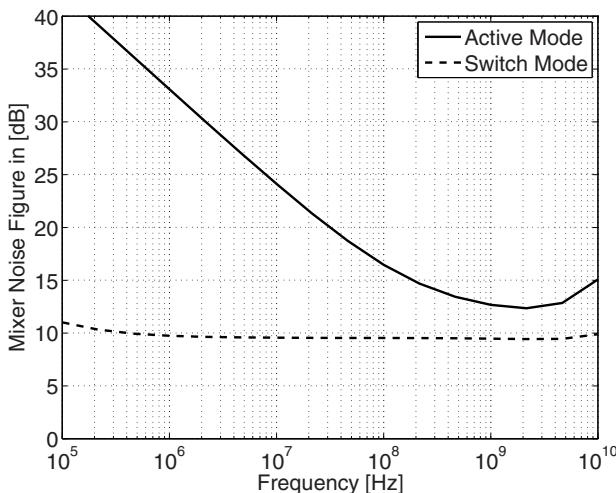


Figure 9: Simulated mixer noise figure for active and passive mode CMOS mixers. Since CMOS technology is often preferable (cost and density-wise), the mixer trade-off between mixer noise performance and conversion gain at baseband must be addressed at the system design level, which will be discussed in the following section

Considering a single conversion receiver, the poor noise performance of the mixer can affect the receiver performance. In Fig. 9, the simulated noise figure (NF) of a switching mixer implemented in 90nm technology, with $7.5\mu m/90\mu m$ devices, is plotted. It can be seen from the figure that

the noise figure when the transistors are biased in the active mode is much larger than for the passive mixer up to about 1GHz, which precludes its use in a low IF receiver.

5. System

The 60GHz unlicensed band allows the RFIC designer a larger degree freedom in choosing the architecture of the receiver. The small size of the passive components allows greater on-chip integration, possibly including the antenna. The main constraint comes from the physical behavior of the active devices, which may restrict the designer to a certain topology. Assuming BiCMOS technology is selected, the high RF performance of the devices (e.g., f/T and gm) allows latitude in the system design. For example, a BJT local (fundamental) oscillator implemented at 60GHz combined with an active mixer (made feasible by low 1/f noise corner of the BJT), enables a homodyne receiver implementation. The designer may then choose to relax the demands on the system blocks and adopt a heterodyne receiver architecture.

The last statement may sound surprising, as it was innovations in single conversion receivers that enabled integration of all the transceiver components onto a single die (excluding the PA). However, an IF filter can now be implemented at 10GHz or above, making it smaller in size and easier to implement on-chip in a heterodyne receiver.

Selecting CMOS technology leads to a different scenario. As shown in Fig. 9, the 1/f noise contribution of a short-channel MOS active mixer is relatively high and adversely affects the front-end noise performance of a homodyne receiver. For gigabit communication, a 500MHz IF (center of the IF band) is a logical assumption. The averaged noise figure of the active MOS mixer up to this frequency is approximately 23dB. If a LNA with a noise figure of 4dB is used, the gain required for a total front-end NF of 5dB is $GLNA=25$ dB. The demand for gain makes attaining broad bandwidth and sufficient linearity difficult. However, if a passive mixer is chosen, the gain of the front-end LNA may be insufficient unless a large number of stages in cascade (e.g., 3 or more) are used, compromising linearity. Consequently, a heterodyne receiver is preferable for a CMOS front-end implementation. Choosing a heterodyne architecture also eases the demands on the LO circuit, which now operates in a lower frequency.

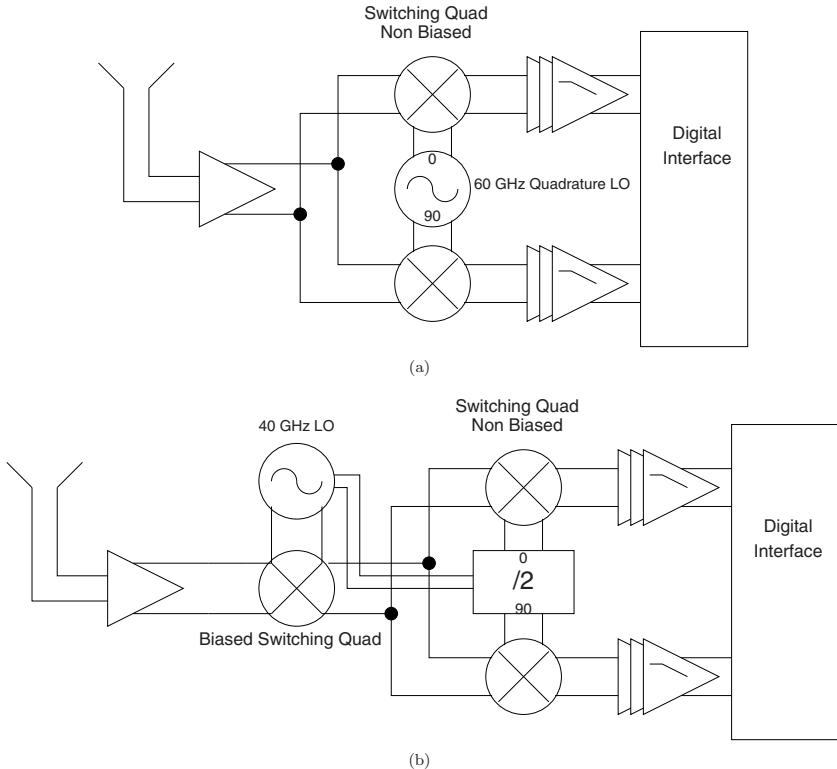


Figure 10: (a) Homodyne Receiver (b) Heterodyne Receiver

Also, quadrature phase local oscillator signal generation is simplified, as seen from Fig. 10 and demonstrated in [7].

Selection of the VCO topology, LO buffering and distribution, and the trade-off between fundamental and harmonic multiplication (or division) in the design of the local oscillator chain, requires careful planning at the system level, and efficient implementations at the circuit level.

6. Conclusions

Deep sub-micron CMOS technology is likely sufficient for the implementation of 60GHz transceivers, as performance limitations may be overcome by utilizing existing circuit topologies and careful system design. BiCMOS technology offers a lower-risk path, especially for the transmitter, where sufficient power to drive an antenna is required. Since the digital circuitry

is usual the prime drive, very deep submicron, (i.e., $< 45nm$) will be necessary.

Standard narrow-band amplifier topologies should be replaced by a broadband topology in order to make maximum use of the spectrum available at 60GHz and to support advanced modulation schemes such as OFDM. Acknowledging that passive components are not modeled well in VLSI silicon technologies at these (and other) frequencies (e.g., due to substrate losses), a fully-differential topology reduces the risk of implementation at the cost of increased power consumption. Also, simple passive structures, such as transmission lines, are physically small on-chip at mm-wave and show reduced parameter variation when driven differentially.

References

- [1] C. A. Balanis, Antenna Theory Analysis and Design, 2nd ed. John Wiley and Sons. Inc., 1997.
- [2] B. A. Floyd, "V-band and w-band sige bipolar low-noise amplifiers and voltage controlled oscillators," in Radio Frequency Integrated Circuits (RFIC) Symposium, 2004. Digest of Papers. 2004 IEEE, June 2004, pp. 295–298.
- [3] B. A. Floyd et al., "Sige bipolar tansceiver circuits operating at 60ghz," IEEE J. Solid-State Circuits, vol. 40, pp. 156–167, 2005.
- [4] B. Razavi, "A 60-ghz cmos receiver front-end," IEEE J. Solid-State Circuits, vol. 41, pp. 17–22, 2006.
- [5] H. D. et al., "A 60ghz cmos differential receiver front-end using on-chip transformer for 1.2 volt operation with enhanced gain and linearity," in Symposium on VLSI Circuits, 2004. Digest of Technical Papers. 2006 IEEE, June 2006, pp. 144–145.
- [6] S. Emami et al., "A 60ghz cmos front-end receiver," in International Symposium on Solid State Circuits, 2007. Digest of Technical Papers. 2007 IEEE, Feb 2007, p. TBD.
- [7] B. Razavi, "A mm-wave cmos heterodyne receiver with on-chip lo and divider," in International Symposium on Solid State Circuits, 2007. Digest of Technical Papers. 2007 IEEE, Feb 2007, p. TBD.

- [8] G. Allen and A. Hammoudeh, “60 ghz propagation measurements within a building,” in European Microwave Conference. Digest of Technical Papers. 1990 IEEE, Oct 1990, pp. 1431–1436.
- [9] T. Cheung and J. Long, “Shielded passive devices for silicon-based monolithic microwave and millimeter-wave integrated circuits,” IEEE J. Solid-State Circuits, vol. 41, pp. 1183–1200, May 2006.
- [10] C. Wang and W. Ruey-Beei, “Modeling and design for electrical performance of wideband flip-chip transition,” IEEE Trans. Adv. Packag., vol. 26, pp. 385–391, Nov 2003.
- [11] C. K. H. Zirath et al., “Flip chip assembly of a 40-60 ghz gaas microstrip amplifier,” in Microwave Conference, 2004. 34th European, vol. 1, Oct 2004, pp. 89–92.
- [12] M. Sanduleanu, “31-34 ghz low noise amplifier with on-chip microstrip lines and interstage matching in 90-nm baseline cmos,” in Radio Frequency Integrated Circuits (RFIC) Symposium, San Fransisco, CA 2006. Digest of Technical Papers. 2006 IEEE, June 2006.
- [13] D. Pozar, *Microwave Engineering*, 3rd ed. John Wiley and Sons. Inc., 2005.
- [14] F. Ellinger, “60-ghz soi cmos traveling-wave amplifier with nf below 3.8 db from 0.1 to 40 ghz,” IEEE J. Solid-State Circuits, vol. 40, pp. 553–558, 2005.
- [15] P. Ikalainen, “Low-noise distributed amplifier with active load,” IEEE Microwave Guided Wave Lett., vol. 6, no. 1, pp. 7–9, 1996.

INTEGRATED FRONTENDS FOR MILLIMETERWAVE APPLICATIONS USING III-V TECHNOLOGIES

Herbert Zirath*#, Sten E. Gunnarsson*, Camilla Kärnfelt*, Toru Masuda⁺,
Mattias Ferndahl*, Rumen Kozhuharov*, Arne Alping#

*Chalmers University of Technology, Department of Microtechnology and Nanoscience
Microwave Electronics Laboratory, Göteborg, Sweden.

Ericsson AB, Microwave and High Speed Electronics Research Centre, Mölndal, Sweden.
+Hitachi, Central Research Laboratory, Tokyo, Japan

Abstract

In order to reduce the manufacturing cost for future 60 GHz products, a high integration level is necessary. Recent results on mHEMT and pHEMT multifunctional receiver/transmitters utilizing are reported. The building blocks for highly integrated millimeterwave front-end circuits based on III-V-technologies such as mixers, amplifiers, frequency multipliers, and VCOs are presented in this work. Balanced and single ended 7–28 GHz MMIC frequency multipliers are described and compared. Multifunctional MMICs utilizing single ended, subharmonically pumped, balanced and single sideband mixers are reported. Examples of multi-stage mHEMT and pHEMT wideband amplifier for example covering 43–64 GHz with a gain of 24 dB, a minimum noise figure of 2.5 dB and ripple of 2 dB are shown.

1. Introduction

A key word for future wireless communication systems is “multimedia communications”, which can provide a variety of services from voice to high definition videos by establishing high data rate channels. High data-rate requires broad frequency bands, and sufficient broadband frequency can be obtained in higher frequency bands such as mm-wave bands. The mm-wave band has several advantages: large spectral capacity, compact and light equipment and for the 60 GHz band (where the oxygen absorption has its maximum, 10–15 dB/km) also the benefits of reduced co-channel interference providing dense, short reach (< 1 km) wireless communication due to shorter cell re-use distance, as well as

access to worldwide allocated non-regulatory frequency bands. The ultimate application area would certainly be for 60 GHz WLAN & WPAN, which would require mass production of small, low-cost and highly integrated transceiver products. However, to provide interoperability with the legacy WLAN at 2.5 GHz or 5 GHz it is necessary to develop a hybrid dual-band system. This would extend existing broadband WLAN systems providing high-speed hot spot access points (APs), as well as a fall back option for the 60 GHz WLAN during temporary worsen channel conditions due to wall attenuation or shadowing. The European IST project Broadway [1] is addressing these issues for scenarios including hot spots in vendor areas and cyber-cafés, high-density residential dwellings and flats, and corporate environments. This report presents the latest development of MMICs, which have been concentrated on VCOs, LNAs, and frequency multipliers. Since we use a GaAs PHEMT technology, our primary goal is to design all MMICs using the same process. It is then possible to integrate the front-end in a few chips. The GaAs PHEMT technology is well suited for 60 GHz but the phase noise of GaAs PHEMT based oscillators is generally quite high due to the high 1/f noise in HEMTs. There is a general belief that HEMT based VCOs cannot compete with HBT-based VCOs. In this work, PHEMT-based VCOs with phase noise comparable with many HBT-based VCOs are presented. This result was obtained by using two tightly coupled grounded gate Colpitt VCOs and maximizing the Q-value of the tank. This topology was also used in a similar HBT-based VCO, with state-of-the-art results in phase noise.

Balanced doublers have been reported [2-7] in the literature, in this work we report a balanced frequency quadrupler. The balanced configuration was considered because it provides an efficient rejection of the fundamental and odd-harmonic frequency, and it is appropriate for implementation of a balanced VCO at the input. The choice of VCO-frequency was determined by the availability of suitable frequency dividers and PLL-circuits on the commercial market.

2. The HEMT technology

The used process is a commercial foundry process from OMMIC [8], D01PH, based on a 0.14 μm gate length, double delta doped technology for high drain current density with high breakdown voltage. The maximum current density and maximum transconductance is 700 mA/mm and 700 mS/mm respectively. The f_t and f_{\max} is 100 and 180 GHz respectively. The process contains two metal layers, via-hole, GaAs and NiCr thin film, and two different MIM-capacitors. Some of the reported designs in this paper are based on a newly developed metamorphic process D01MH, also from OMMIC. For the multifunctional receiver/transmitter developed, we used a PHEMT and mHEMT processes from WIN Semiconductor.

3. Circuit designs and measurements

The first generation MMICs for the transceiver topology according to Fig. 1 have been designed and characterized [9-13]. Lately, significant improvement in the VCO, mixer and amplifier design was achieved, and different multifunctional MMICs are now being developed based on these designs.

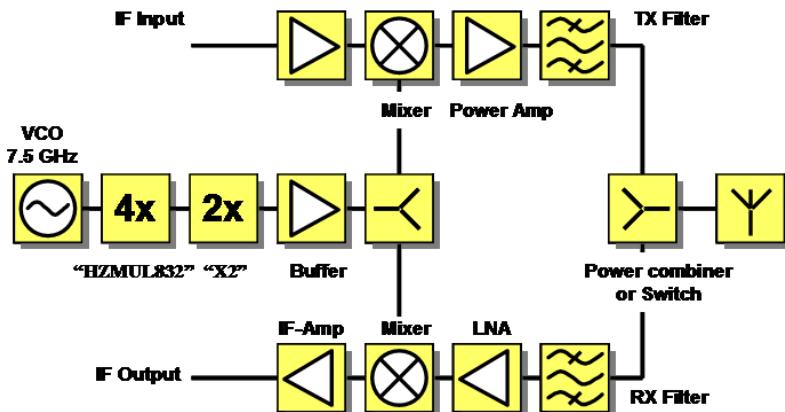


Fig. 1 Transceiver front end design

3.1 VCO

Our previous VCO designs were based on a reflection type oscillator topology utilizing one transistor [13]. These oscillators were designed for various frequencies from 7 GHz up to 56 GHz. An output power of the order 10dBm is obtained but the phase noise is quite high. For more advanced digital modulation schemes, a phase noise of -110dBc@100kHz frequency offset might be necessary. Such low phase noise can be realized by using a dielectric stabilized oscillator but is hard to realize with MMIC-based VCOs with on chip varactor. The HBT technology is regarded to be the best technology choice due to a low 1/f noise. The status of MMIC-based VCOs obtained from the literature is presented in Fig. 2 and Table 1.

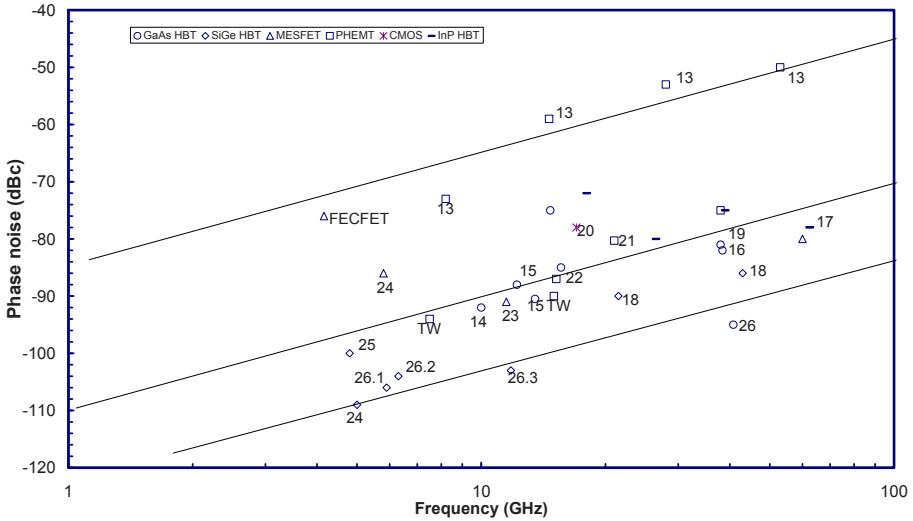


Fig. 2 Phase noise of MMIC-based reported VCOs

The obtained phase noise as a function of frequency at an offset frequency of 100 kHz is plotted along with 20dB/decade slope-lines. Line 1 represents our previously obtained results using a reflection topology, line 2 represents the best III-V HBT VCOs [14-19] and a few MESFET and HEMT based VCOs [20-23]. SiGe HBT VCOs have shown impressive performance, see ref [25]. Ref. [24] is a SiGe-HBT MMIC oscillator based on a balanced Colpitt design serving as a pre-study to this work showing very low phase noise. Although this design has no varactor, the tank circuit is completely integrated. For even lower phase noise, combining methods can be considered. This was demonstrated by Jacobsson et al. [26] showing that phase noise can be improved 1-3 dB, and 3-6 dB for double and quadruple VCO-design respectively. An InGaP/GaAs HBT of common emitter reflection type [27] has shown similar phase noise levels but this design has very limited tuning range. Line 3 represents state-of-the-art in phase noise of MMIC oscillators with integrated tank. In our lab, various PHEMT VCO designs have been simulated, optimized and characterized such as balanced Colpitt, negative transconductance, and second harmonic balanced Colpitt. A reason for investigating the Colpitt oscillator topology is its favorable impulse sensitivity characteristics [28], which is important for achieving low phase noise.

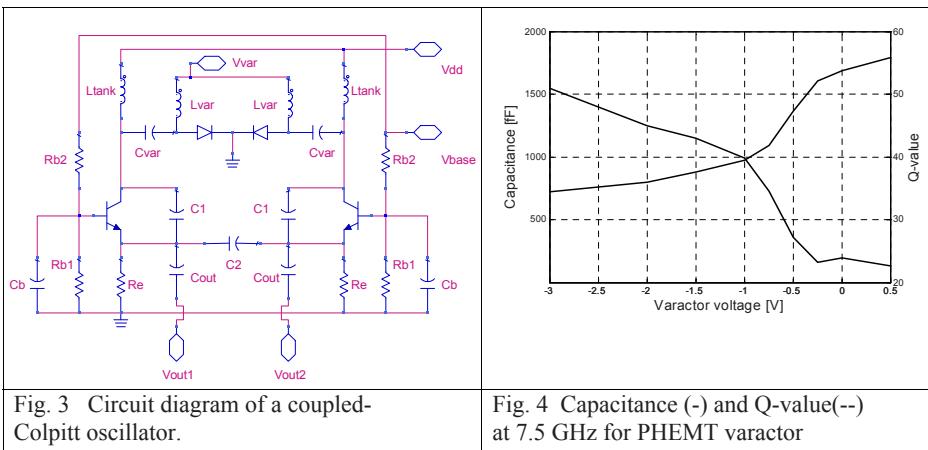
Table 1 MMIC VCO performance

Ref	type	Pout dBm	Freq. GHz	Tuning range GHz	Phase noise dBc@100kHz	Pdc mW
14	InGaP-HBT	11-13	10	1.5	-92	ND
15	InGaP-HBT	-1.5	12.2	0.6	-88	48
15	InGaP-HBT	0	13.5	0.8	-90.5	36
16	InGaP-HBT	0	40.5	0.2	-83	ND
17	InP HBT	-4 to 4	62.4	0.3	-78	ND
18	SiGe HBT	-6	21.5	1	-90	130
18	SiGe HBT	-17	43	2	-86	130
19	AlInAs/InGaAsHBT	10	38	0.85	-82	ND
20	Si CMOS	ND	17	1.4	-78	10.5
21	GaAs HEMT	13.7	21	1.6	-80.3	ND
22	GaAs HEMT	17	15.2	0.6	-87	ND
23	GaAs MESFET	11.5	11.5	0.55	-91	ND
24	SiGe HBT	-4	5	-	-109	96
25	SiGe HBT	-13	4.8	0.27	-100	46
26.1	SiGe HBT	-4	6.3	1	-104	30
26.2	SiGe HBT	-6	5.9	1	-106	53
26.3	SiGe HBT	-7	11.8	2	-103	106
27	InGaP-GaAs	5.3	40.8	0.012	-95	ND
Fig 3	GaAs PHEMT	3-6	7.5	0.4	-94 to -89	150
Fig 1	GaAs PHEMT	-3 to 2	15	0.8	-90	150

The balanced Colpitt topology was first implemented in a SiGe HBT technology with very impressive results; at 5 GHz, a phase noise of -109dBc@100kHz was achieved [24]. A schematic diagram of the balanced Colpitt VCO is shown in Fig. 3. This VCO consists of two separate grounded gate Colpitt VCOs which are tightly coupled. The arrangement with two coupled oscillators gives a 3 dB improvement in phase noise compared to a single oscillator [29]. Instead of using two separate feedback grounding capacitors to the source (one for each oscillator), one single capacitor is ‘cross connected’ between the sources. Since the oscillators are oscillating out of phase, a virtual ground is created inside the capacitor. This topology has the advantage saving one capacitor, no RF-ground currents have to circulate due to this capacitor, and the size of the cross coupled feedback capacitance is reduced by a factor of two. The phase noise of an oscillator can be described by Leeson’s formula

$$L(\Delta\omega) = 10 \cdot \log \left\{ \frac{2 \cdot F \cdot k \cdot T}{P_{sig}} \left[1 + \left(\frac{\omega_0}{2 \cdot Q \cdot \Delta\omega} \right)^2 \right] \cdot \left(1 + \frac{\Delta\omega / f^3}{|\Delta\omega|} \right) \right\}$$

F is the noise figure, T is the temperature, k is Boltzmann's constant, P_{sig} is the signal power, Q is the loaded Q-value of the tank, ω_0 is the oscillation frequency, $\Delta\omega$ is the offset from the oscillation frequency, $\Delta\omega_{1/f}^3$ is the corner frequency for 1/f noise. The equation is phenomenological and useful in order to get an understanding how phase noise can be minimized. In general, the Q-value of the tank should be maximized, the oscillation amplitude in the tank should be maximized, and the noise generated by the transistor should be as small as possible. The Q-value of the tank is determined by the tank inductor, the L_{tank} , the tank capacitance C_{tank} , and the loading by the transistor's source through the Colpitt feedback capacitors C_1 , C_2 , and the output load. The tank capacitance in our designs is the gate-Schottky diode of a HEMT, acting as a voltage controlled capacitor (varactor), in series with a MIM-capacitor. This varactor is not optimum in terms of Q-value but is the only choice when using a PHEMT-process; if not an advanced varactor epilayer is available in the process. Such processes have been reported and VCOs utilizing 'epi-varactors' have been demonstrated [21], but this inevitable increases the complexity of the process and the cost. In our design, the series resistance of the varactor was minimized by using a multi finger layout ($N=20$) and using the same gate length as the HEMTs, 0.14 μ m. Separate S-parameter measurement on this varactor yielded a Q-value of 40 and 1 pF capacitance at 7.5 GHz, see Fig. 4.



The oscillator design includes an overall optimization using harmonic balance simulations where the feedback network, the size of the transistors, and the dc-current are parameters. In the optimization, the voltage swing of the tank is maximized, while the drain current is designed to have a narrow pulse shape i.e. the transistor is forced to operate in Class B or C with a conduction angle of less than 180 degree. Fig. 5 shows the simulated drain current and the drain voltage in the time domain.

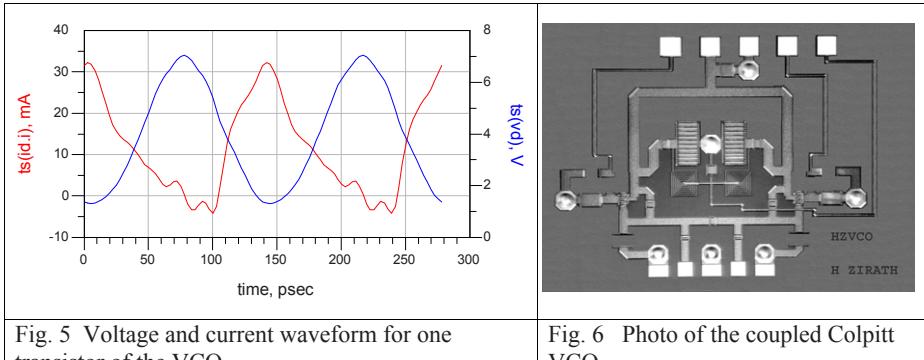


Fig. 5 Voltage and current waveform for one transistor of the VCO

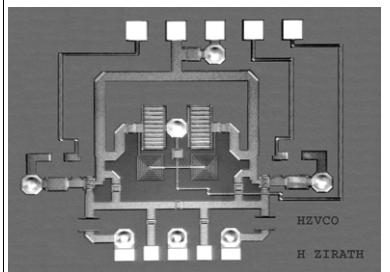


Fig. 6 Photo of the coupled Colpitt VCO

The layout of the VCO is shown in Fig. 6. The chip size is 2x1.5 mm. The outputs are balanced with CPW-pads located at the bottom of the chip. The supply voltage, varactor control voltage, and gate bias are connected at the top of the chip. The gates are connected to 0V through resistors but it is possible to adjust the gate voltage for an adjustment of output power etc. The oscillation frequency versus varactor voltage and output power versus varactor voltage is plotted in Fig. 7 and Fig. 8.

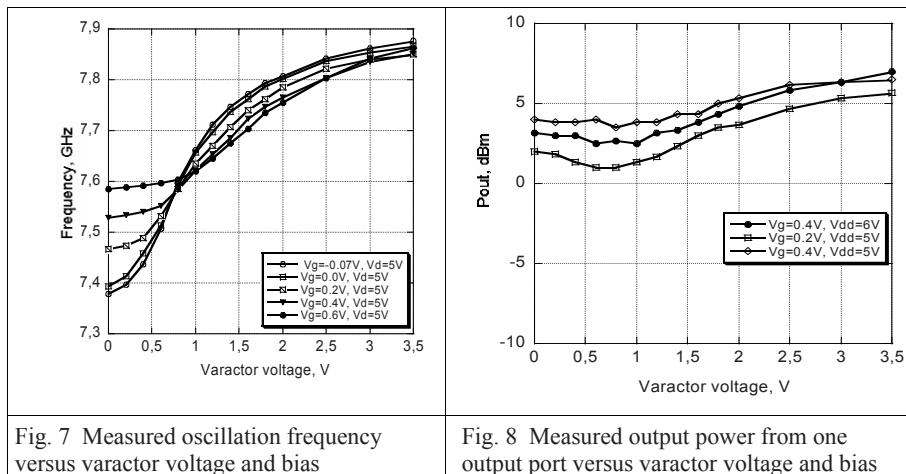


Fig. 7 Measured oscillation frequency versus varactor voltage and bias

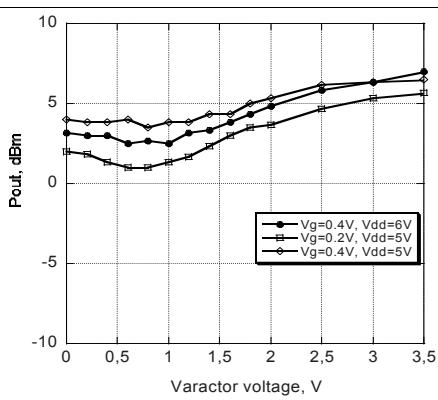


Fig. 8 Measured output power from one output port versus varactor voltage and bias

The phase noise was measured with a spectrum analyzer and with an Agilent E 5500 phase noise system ‘on wafer’ in a shielded probe station. A minimum phase noise of -94 dBc is obtained at an offset frequency of 100kHz from the carrier at 7.52 GHz. Across a 7.4-7.8 GHz tuning range, the phase noise is below -89 dBc, see Fig. 9. To our best knowledge, this result represents state-of-the-art for a PHEMT VCOs with on-chip tuning varactor and is comparable with

many VCOs based on HBTs. The oscillator shows the expected 30 dB/decade slope between LF and 10kHz, above 10kHz the slope is 25dB/decade. The output power is 5 dBm when combined in a balun.

A second harmonic oscillator based on basically the same topology was also investigated. The second harmonic output is taken from the virtual ground node of the feedback network, see Fig. 10.

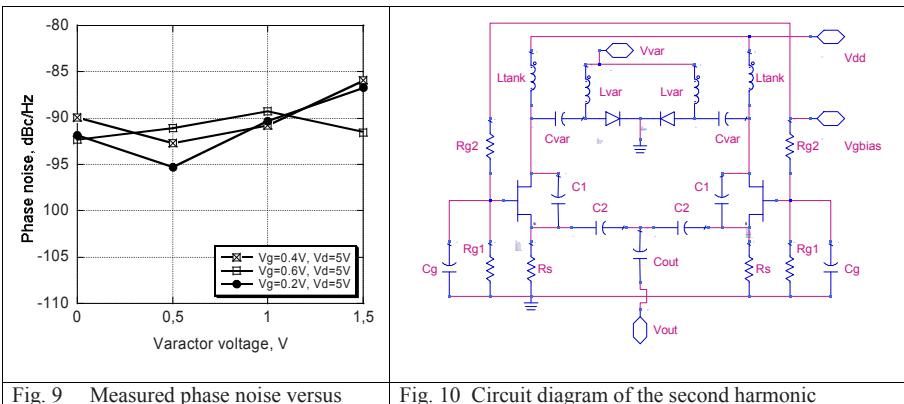


Fig. 9 Measured phase noise versus varactor voltage and bias

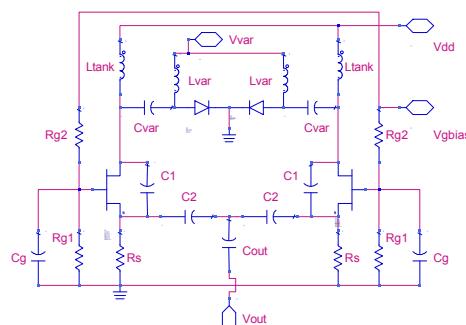


Fig. 10 Circuit diagram of the second harmonic coupled-Colpitt oscillator.

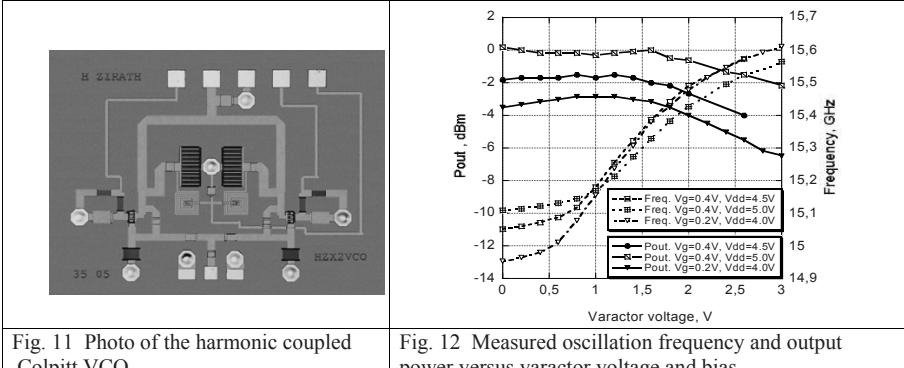
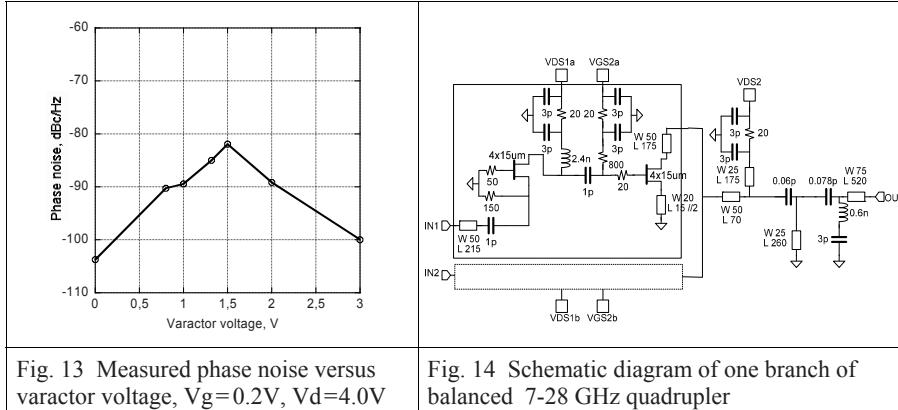


Fig. 11 Photo of the harmonic coupled Colpitt VCO

Fig. 12 Measured oscillation frequency and output power versus varactor voltage and bias.

By this arrangement, the output load has negligible effect on the loaded Q-value of the tank and the phase noise can be expected to be improved compared to the previous design. The drawback is a decreased output power. A photo of this VCO is shown in Fig. 11. The measured oscillation frequency and output power versus varactor voltage is shown in Fig. 12. The frequency range of the VCO is 14.9-15.8 GHz. The output power is between 0 and -2 dBm. The phase noise was measured as a function of varactor voltage which is plotted in Fig. 13.



The phase noise is improved compared to the fundamental VCO as expected, $-90 \text{ dBc}@100\text{kHz}$ is measured. This result represent the lowest phase noise reported for a PHEMT based VCO with an on-chip varactor. Recently, we have reported coupled Colpitt VCOs based on an InGaP-GaAs HBT process [32]. This VCO was optimized for low phase noise considering the tank design, choice of transistor technology, transistor size, impulse sensitivity, and oscillator amplitude (and the ability to control it). The measured phase noise is less than $-112 \text{ dBc}@100\text{kHz}$ at an oscillator frequency of 6.4 GHz, with an output power of 6dBm for a fundamental VCO with the two outputs power combined. For the second harmonic VCO a minimum phase noise of -120 dBc is obtained at 12.9 GHz. The output power from the second harmonic VCO is approximately 6dB lower, compared to the fundamental VCO. To our best knowledge, the reported phase-noise represents state-of-the-art for a VCO with a fully integrated tank described in the research literature. The frequency tunability is however limited to a few percent and we now intend to increase the tunability. Although it was shown that the phase noise is much lower for the InGaP-GaAs HBT technology, the PHEMT based VCO is useful in many cases where the higher phase noise can be accepted.

3.2 Frequency multipliers

The balanced outputs from the previously described VCO can directly be used for balanced frequency multipliers thus saving the input balun. In this work, we have investigated novel single ended and balanced quadruplers. The schematic diagram of one branch of the balanced quadrupler 7-28 GHz is shown in Fig. 14. Both the single ended and balanced quadruplers are based on a two-stage configuration. The first stage is a grounded gate active input impedance matching circuit. By choosing the appropriate source and gate resistances to

ground, an optimum matching condition over a large frequency range can be obtained. In order to reduce the dc power consumption and generate the fourth harmonic signal, the second stage transistor is biased near the pinch off region, where the transfer nonlinearity is used for frequency multiplication. At the output, a 2-pole high pass filter is used for suppression of the lower harmonics, and matching at the output frequency. The transistors in this design have a width of $4 \times 15 \mu\text{m}$. A photo of the single ended quadrupler is shown in Fig. 15. The active input matching was evaluated by measuring the small signal S-parameters, S_{11} is found to be below -10 dB from 6 to 45 GHz. The large signal measurement is carried out with a HP 8565E spectrum analyzer. The measured and simulated characteristics of the output power of the fundamental and all harmonics up to 4-th versus input frequency are shown in Fig. 16. The input power is 0 dBm. The effective rejection of unwanted harmonics is larger than 15 dB for the third and more than 25 dB for the second and the fundamental. A maximum output power of -7.7 dBm in a bandwidth of 1.8 GHz was obtained. The measured and simulated power dependence of the output power at 7.5 GHz input frequency is shown in Fig. 17. A photograph of the balanced quadrupler is shown in Fig. 18. Symmetrical bias circuits are designed to keep the balance in the circuit. The size of the chip is the same as of the single ended multiplier. Compared to a single ended frequency quadrupler, a balanced quadrupler shows an excellent rejection of the first and third harmonic, see Fig. 19. The optimized results are obtained by decreasing the drain voltage of the second stage ($V_{d2}=1\text{V}$). In Fig. 20, the spectrum of the single and the balanced quadrupler are compared. The balanced quadrupler shows an excellent suppression of first and third harmonics (50 dB compared to 30 dB, and 30dB compared to 15 dB respectively). The balanced quadrupler exhibits also higher output power -2.5 dBm, compared to -7.7 dBm with similar power dissipation less than 50 mW. The measured frequency bandwidths are equal - approximately 22%.

3.3 Broadband millimeter wave amplifiers

A design methodology was developed for applications where a higher bandwidth is needed. The basic idea is to use interstage equalizers that compensate for the intrinsic gain roll-off of the transistor; the basic concept is described in detail in [28]. In this work, it is accomplished by loading the drain of each transistor by a shorted stub. In addition, a series element, which consists of a parallel RC-network, is used in the interstage networks for improving the stability at lower frequencies by decreasing the gain outside the pass band as described by Ono et al. [29]. A photo of a three stage V-band design is shown in Fig. 21. The metamorphic 0.1 μm gate length D01MH-process from OMMIC was used in this investigation.

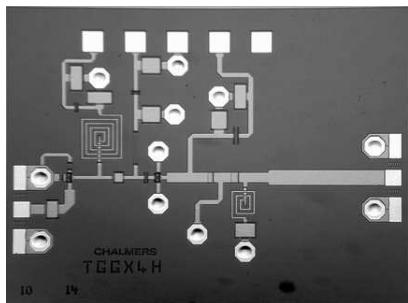


Fig. 15 Photograph of the single ended frequency quadrupler. The effective chip size is 1.5×1.5 mm

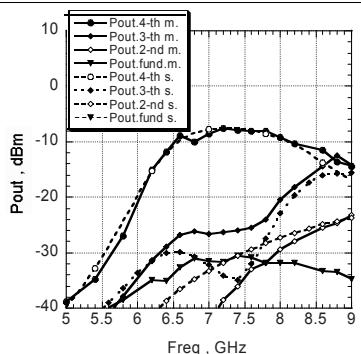


Fig. 16 Measured (line) and simulated (dot) output power vs. input frequency of single ended frequency quadrupler

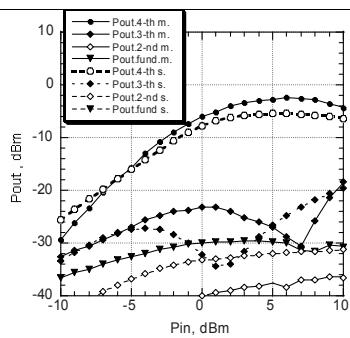


Fig. 17 Measured (line) and simulated (dots) output power dependence on input power at 7.5 GHz input frequency

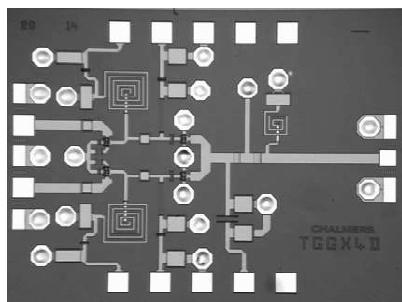


Fig. 18 Photograph of balanced quadrupler. Chip size is 2×1.5 mm

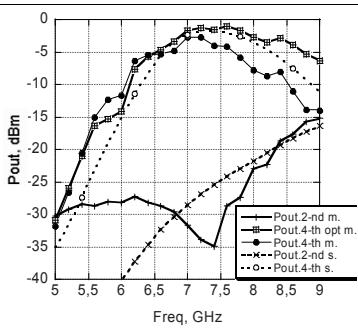


Fig. 19 Measured (m), optimized (line) and simulated (ds) output power vs. input freq. of quadrupler

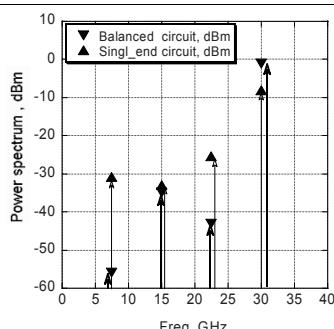
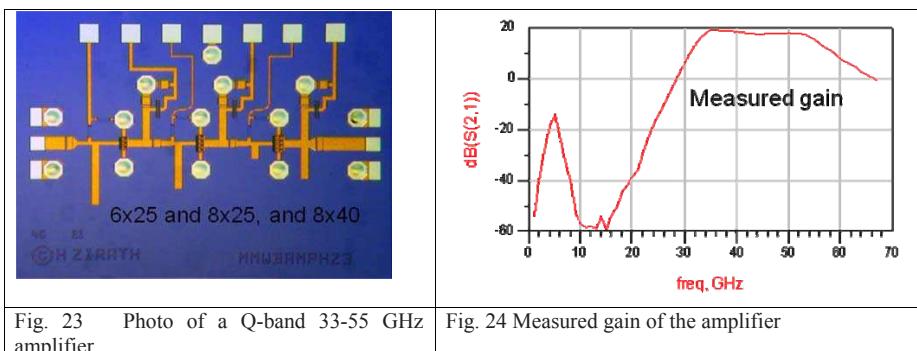
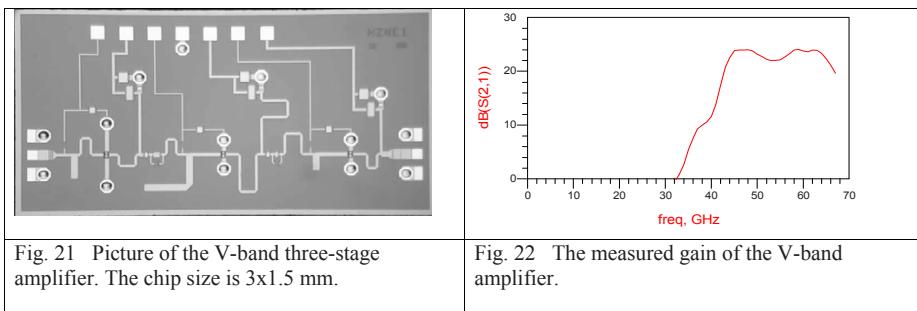


Fig. 20 Harmonic spectrum comparison of single ended (\blacktriangle) and balanced (∇) quadrupler

The chip was characterized by using an Agilent PNA network analyzer from 0.1 to 67 GHz. The measured S_{21} is shown in Fig. 22. The amplifier covers 43-64 GHz with a gain of 24 dB, and the ripple within the band is 2 dB. The noise figure was also measured and found to be approximately 2.5 dB at 50 GHz. Another example of a broadband Q-band (33-55GHz) 3-stage amplifier is shown in Fig. 23 (photo), designed for an output power of approximately 100mW. The D01PH process from OMMIC was used for this design. The gate widths of the transistors are selected to prevent saturation in the stage 1 and 2. Gate widths of 150, 200 and 320 μm were chosen. Stabilizing resistors are used in the drain bias circuits to prevent oscillation. The measured gain as a function of frequency is shown in Fig. 24.



4. Multifunctional 60 GHz MMICs

60 GHz transmit and receive chipset

Based on the previously described work, a receiver and a transmitter chip was recently developed [33] based on PP15-20, a 0.15 μm PHEMT process from WIN

Semiconductors. The receiver (RX) chip is designed in a similar way as the transmitter chip and consists of an X8 LO-chain, image reject mixer and three stage amplifier.

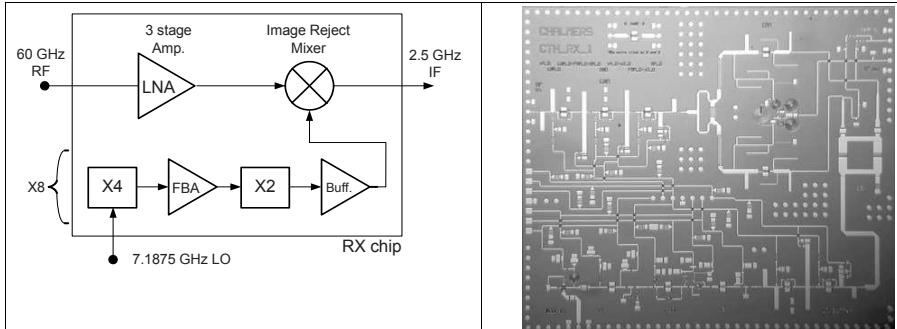


Fig. 25 Block diagram of the receiver MMIC

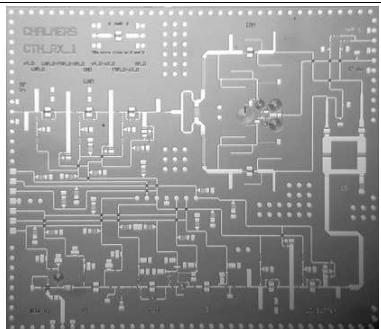


Fig. 26 Photo of the receiver chip.
The chip measures 5.7×5.0 mm

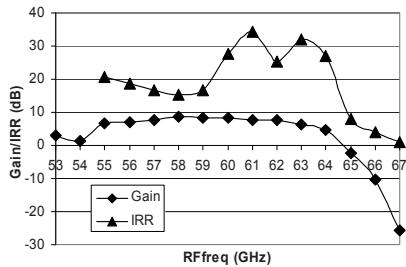


Fig. 27 Conversion Loss and Image Rejection for the RX chip versus RF frequency, $f_{IF}=2.5$ GHz

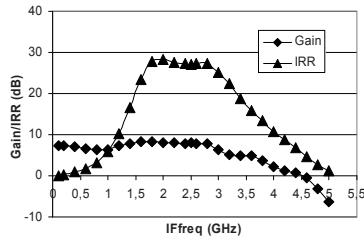


Fig. 28 Conversion gain and IRR versus IF-frequency for the RX chip, $f_{LO}=57.5$ GHz

The block diagram is shown in Fig. 25 and a photo of the circuit is presented Fig. 26. The chip measures 5.7×5 mm. The measured GC and IRR for the RX chip versus RF frequency is plotted in Fig. 27. The RX chip possesses a 3 dB RF bandwidth of 8 GHz between 55 and 63 GHz with an optimal G_c of 8.6 dB at 58 GHz. The IRR is larger than 20 dB between 59.5 and 64.5 GHz. In Fig. 28 the measured G_c and IRR are plotted versus IF-frequency. The 3 dB IF bandwidth ranges from 0 to 3.2 GHz and the IRR is larger than 20 dB between 1.5 and 3.3 GHz. The block diagram of the transmitter is shown in Fig. 29, a photo of the chip is depicted in Fig. 30. The output power of the chip was measured and is plotted Fig. 31 versus RF-frequency and Fig. 32 versus IF-frequency.

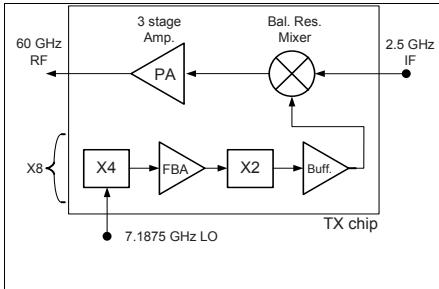


Fig. 29 Block diagram of the transmitter chip

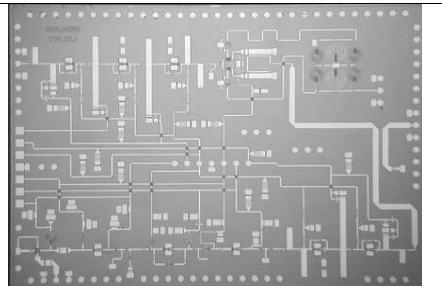


Fig. 30 The transmitter chip. The chip measures 5.0 x 3.5 mm

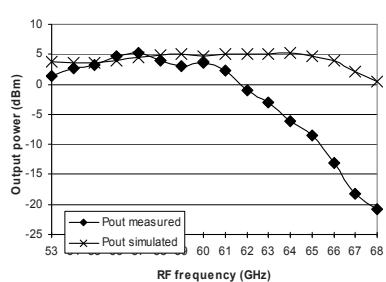


Fig. 31 Output power versus RF-frequency

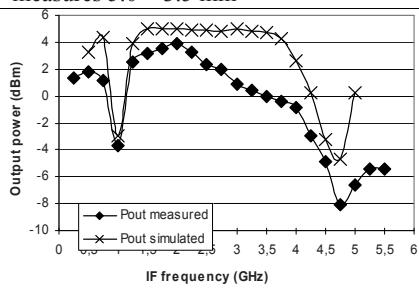


Fig. 32 Output power versus IF-frequency

Due to the general architecture of the chipset any modulation format can be used. In Fig. 33, a test bench for system tests is depicted, general modulation signals can be calculated and loaded to the ESG. Due to the limitation in the measurement setup, we used ASK for higher bitrates than 200Mbit/s, the test bench for this setup is shown in Fig. 34 together with measured eye-diagrams and measurements of bit-error rates. A summary of the results from the measured transmitter and receiver MMICs are shown in Table 2 and 3.

Table 2
Summary of measured results for the TX chip.

Conversion Gain, dB	Output power (dBm)	Max output power (dBm)	3 dB RF Bandwidth, GHz
-1.4	3.3	5.2 (@ 57 GHz)	7 (54 - 61)
3 dB IF Bandwidth, GHz	1dB Compression Point, dBm	LO Power, dBm	Power consumption (mW)
1.5 (1.25 - 2.75)	0	-3	820

Table 3
Summary of measured results for the RX chip.

Conversion Gain, dB	3 dB RF Bandwidth, GHz	3 dB IF Bandwidth, GHz	Image Rejection Ratio (dB)
8.5	8 (55 - 63)	3.2 (0 - 3.2)	> 20 (59.5 - 64.5 GHz)
IIP3, dBm	1dB Compression Point, dBm	LO Power, dBm	Power consumption (mW)
-11	-19	-2	990

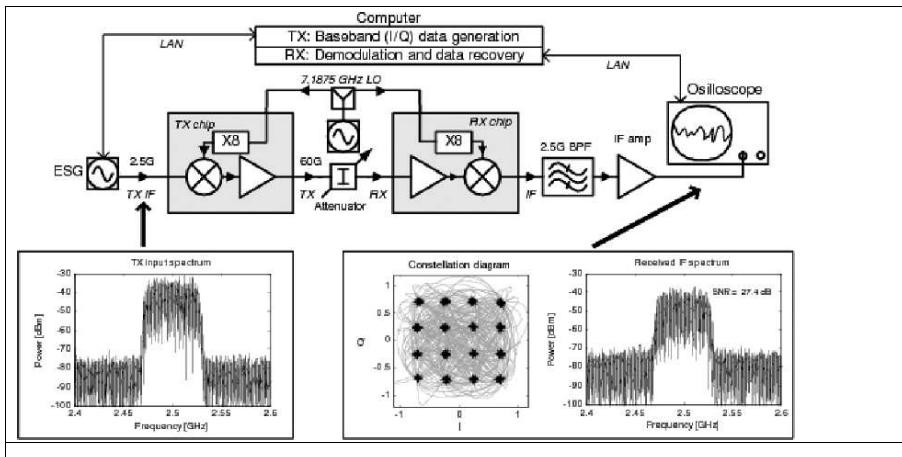


Fig. 33 Testbench for general system test. 16 QAM 200Mbit/s was investigated in this setup

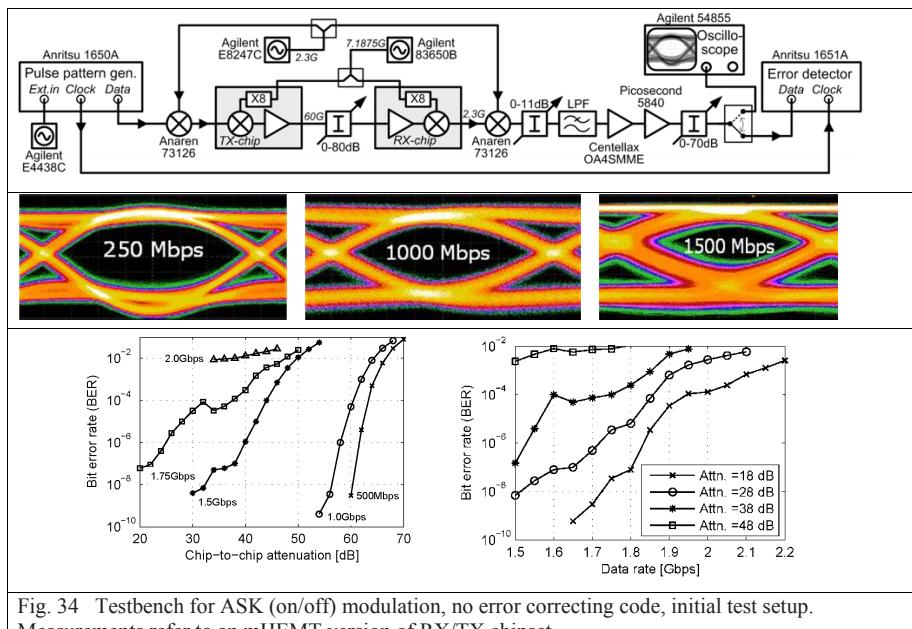
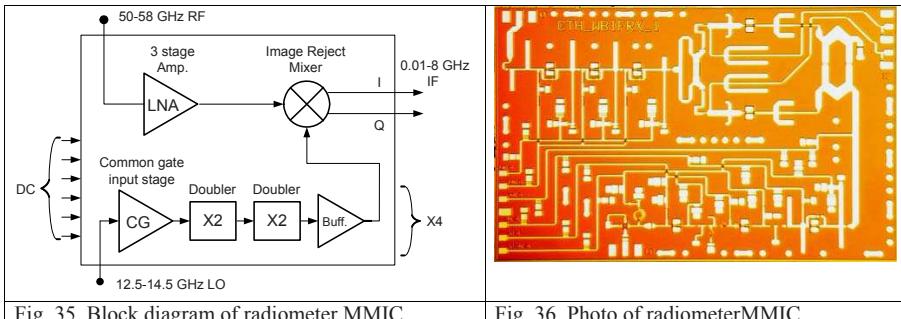


Fig. 34 Testbench for ASK (on/off) modulation, no error correcting code, initial test setup. Measurements refer to an mHEMT version of RX/TX chipset.

An integrated radiometer with a wideband IF for 53 GHz

A modified version of the 60 GHz receiver was designed in WINs mHEMT process for a 53 GHz radiometer application. The mixer has I-Q wideband intermediate frequency output which is sampled by a high speed spectrometer processor. The block diagram of the radiometer is shown in Fig. 35 and a photo in Fig. 36. The measured 3-dB bandwidth is 7 GHz.



A subharmonically pumped receiver for 60GHz

A 58-66 GHz subharmonically pumped receiver is shown in Fig. 37 (schematic) and Fig. 38 (photo). The receiver contains a two-stage amplifier and a subharmonically pumped mixer intended to be used for an ASK/FM multifunction TX/RX chip with integrated VCO. An active balun is used for the generation of the 0/180° LO-signals from a single LO-input. In Fig. 39, the gain and isolation characteristics are plotted, and in Fig. 40, the conversion gain versus gate bias for the mixer transistors versus LO-power.

A 24 GHz FMCW radar chip

Frequency modulated continuous wave (FMCW) radar sensors are widely used for both range and velocity determination. This integrated FMCW-sensor contains a local oscillator (LO) which provides a transmit signal to the antenna, LO signal to a receive mixer and a signal to an off-chip PLL-circuit for phase locking of the LO. In addition, a buffer amplifier, a Wilkinson power splitter, and a receiver mixer are integrated. A push-push VCO topology is used for low phase noise. The chip is intended to be used in radar transceiver (FMCW) for distance sensor application. The block diagram is shown in Fig. 41 and a photograph of the chip in Fig. 42. The dc current requirement of the VCO and buffer amplifier is 100mA and 20mA respectively with a bias voltage 2.2V. The complete Pdc is 260 mW, the mixer is a resistive mixer. In Fig. 43, the bandwidth and output power measurements are plotted. Varying the varactor voltage between 0V and 2.8V, the output frequency changes from 23GHz to 24.3 GHz respectively. In this frequency range the oscillator's output power is

10 ± 1.2 dBm and the measured phase noise level is less than -92 dBc/Hz at 1 MHz offset frequency as depicted in Fig. 44.

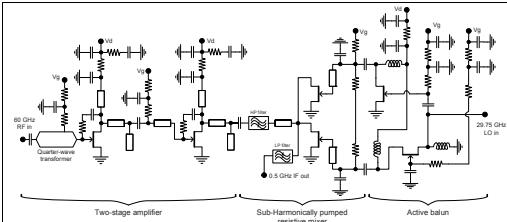


Fig. 37 Schematic of the subharmonically pumped downconverter

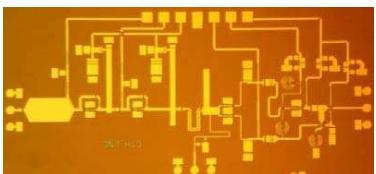


Fig. 38 Photo of the downconverter

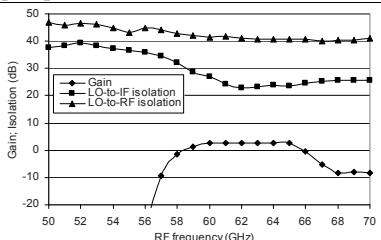


Fig. 39 Gain and isolation characteristics versus frequency

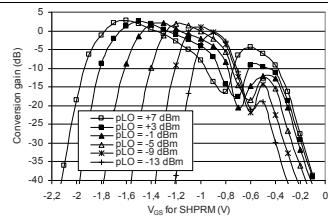


Fig. 40 Conversion gain versus gatebias

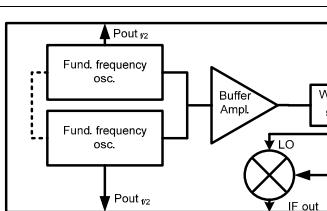


Fig. 41 Block diagram of the 24 GHz FMCW chip

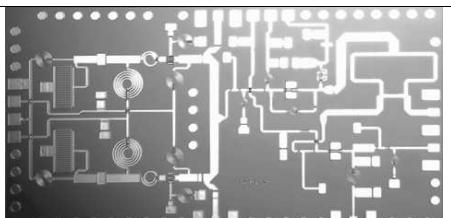


Fig. 42 Photograph of the 24 GHz mHEMT MMIC. The chip size is 3.8 x 2mm

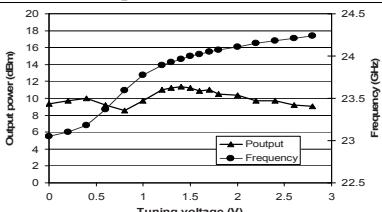


Fig. 43 Output power and oscillation Frequency versus tuning voltage

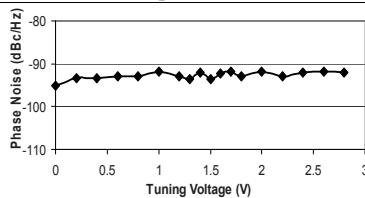


Fig. 44 Phase noise at 1-MHz offset versus tuning voltage

5. Conclusions

Millimeterwave transceiver front end MMICs have been realized in GaAs PHEMT and mHEMT technologies and multifunctional MMICs like a 60 GHz transceiver chip set, 53 GHz radiometer, a subharmonically pumped 60 GHz frontend, and a 24 GHz FMCW-radar have been developed. PHEMT-based VCOs have shown a phase noise comparable with many HBT-based VCOs described in the literature. In applications where phase-noise is a very critical parameter, the InGaP-GaAs HBT VCO is a better choice. A PHEMT or MHEMT technology is suitable for ‘one-chip’ solutions for 60 GHz front-ends for less critical modulation formats like ASK since they offer overall high performance like low noise figure for RF-amplifiers, good linearity for mixers, and high output power for power amplifiers. MMIC processes which combine HBT and FET/HEMT are being developed at some foundries at the moment thus enabling the possibility to combine HBT-based low phase noise VCOs with low noise HEMT amplifiers and switches for next generation of high performance multifunctional MMICs.

Acknowledgement

Dr Thomas Lewin from Ericsson AB and Dr Jan Grahn from Chalmers University of Tech. are acknowledged for their support in this work. Dr Christian Fager is acknowledged for help with the system measurement setup. MMIC foundries OMMIC and WIN are acknowledged for processing of the circuits.

References

- [1] BROADWAY homepage <http://www.ist-broadway.org>
- [2] B. Piernas et al., ‘A broadband and Miniaturized V-Band PHEMT Frequency Doubler’, IEEE Microwave and guided wave letters, v.10, N.7 July 2000.p.276
- [3] T. Hiraoka et al., ‘A Miniaturized Broad-Band MMIC Frequency Doubler’, IEEE Microwave theory and techniques, Vol.38, No.12, p.1939, Dec.1990.
- [4] P. Kangaslahti et al., ‘Miniaturized artificial-transmission-line monolithic millimeter wave frequency doubler’, IEEE Microwave theory and techniques, Vol.48, No.4, Part 1, p.510, April 2000.
- [5] P. Kangaslahti et al., ‘Low phase noise signal generation circuits for 60 GHz wireless broadband system’, IEEE MTT-S International Microwave Symposium Digest, vol.1, p.43, 2000.
- [6] K. Deng et al., ‘A miniature broad-band pHEMT MMIC balanced distributed doubler’, IEEE Microwave theory and techniques, Vol.51, No.4, p.1257, April 2003.
- [7] D. Klymyshyn et al., ‘Active Frequency-Multiplier Design Using CAD’, IEEE Microwave theory and techniques, Vol.51, No.4, p.1377, April 2003.

- [8] OMMIC homepage <http://www.ommic.com>
- [9] H. Zirath, C. Fager, M. Garcia, P. Sakalas, L. Landen, A. Alping, ‘Analog MMICs for Millimeter-Wave Applications based on a Commercial 0.14 μm PHEMT Technology’, IEEE Transactions on Microwave Theory and Techniques, pp. 2086-2092, Vol. 49, No. 11, November 2001.
- [10] H. Zirath, K. Yhland, D. Marcouly, ‘An ultra wideband balanced resistive mixer, based on a deltadoped 140 nm AlGaAs-InGaAs-GaAs HEMT MMIC process’, Asia Pacific Microwave Conference, Kyoto, Nov 2002.
- [11] T. Masuda, L. Landen, H. Zirath, ‘Low power single ended active frequency doubler for a 60 GHz band application,’ Proceeding of GaAs 2002. EUMW 2002
- [12] R. Kozuharov, T. Masuda, H. Zirath, V. Lövenmark, ‘A 55GHz HEMT Monolithic Voltage Controlled Source’, GaASIC 2003.
- [13] H. Zirath, M. Hasselblad, R. Kozuharov, L. Landén, T. Masuda, K. Yhland, ‘MMIC-based nonlinear circuits for millimeter wave applications with low power dissipation’, Asia Pacific Microwave Conference, Taipei, Taiwan Dec 2001.
- [14] Z. Ouarch, F. Arlot, M. Borgarino, M. Prigent, L. Bary, M. Camiade, ‘Low phase noise integrated monolithic VCO in X-band based on HBT technology’, 2001 IEEE MTT-S Digest, pp. 1415-1418, 2001.
- [15] D-H. Baek, J-G. Kim, S. Hong, ‘A Ku band InGaP-/GaAs HBT MMIC VCO with a Balanced and a Differential Topologies’, 2002 IEEE MTT-S Digest, pp. 847-850, 2002.
- [16] H. Do-Ky, M. Stubbs, T. Laneve, C. Glaser, D. Drolet, ‘Ka-band MMIC Voltage Controlled Oscillators’, 1997 Asia Pacific Microwave Conference Digest, APMC 1997, pp. 545-548, 1997.
- [17] H Wang, et al., ‘A 62-GHz Monolithic InP-based HBT VCO’, IEEE Microwave and Guided Wave Letters, Vol. 5, No. 11, pp. 388-390, Nov. 1995.
- [18] M. Bao, Y. Li, and H. Jacobsson, “a 21.5/43 GHz dual-frequency balanced Colpitts VCO in SiGe technology”, *IEEE J. solid-state circuits*, vol. 39, pp. 1352-1355, Aug. 2004.
- [19] A. Kurdoghlian, M. Sokolich, M. Case, M. Micovic, S. Thomas III, C.H. Fields, ‘30 GHz Low Phase Noise CPW Monolithic VCOs implemented I Manufacturable AlInAs/InGaAs HBT IC Technology’, 2000 IEEE GaAs Digest, pp. 99-102, 1999.
- [20] C. R. C. De Ranter, M. S. J. Steyaert, ‘A 0.25 mm CMOS 17 GHz VCO’, 2001 International Solid-State Circuits Conference Digest, pp. 370-372, 2001.
- [21] O. Sevimli, J.W. Archer, G.J. Griffiths, ‘GaAs HEMT Monolithic Voltage-Controlled Oscillators at 20 and 30 GHz Incorporating Schottky-Varactor Frequency Tuning’, IEEE Trans Microwave Theory and Techniques, Vol. 46, No.10, pp. 1572-1576, Oct. 1998.
- [22] J. Portilla, M.L. de la Fuente, J.P. Pascual, E. Artal, ‘Low-Noise Monolithic Ku-Band VCO using Pseudomorphic HEMT Technology’, IEEE Microwave and Guided Wave Letters, Vol. 7, No. 11, pp. 388-390, Nov. 1997.

- [23] C.H. Lee, S. Han, B. Matinpour, J. Laskar, 'GaAs MESFET-based MMIC VCO with low phase noise performance', 2000 GaAs Digest, pp. 95-98, 2000.
- [24] H. Zirath, unpublished results
- [25] H. Jacobsson, S. Gevorgian, M. Mokhtari, C. Hedenäs, B. Hansson, T. Lewin, H. Berg, W. Rabe, A. Schüppen, 'Low-Phase-Noise Low-Power IC VCOs for 5-8-GHz Wireless Applications', IEEE Trans. Microwave Theory and Techniques, Vol.48., No. 12, pp.2533-2539, Dec. 2000.
- [26] H. Jacobsson, B. Hansson, H. Berg, S. Gevorgian, 'Very Low Phase-Noise Fully-Integrated Coupled VCOs', 2002 Radio Frequency Integrated Circuits Symposium, paper IF-TU-52, pp. 467-470, 2002.
- [27] M.S. Heins, D.W. Barlage, M.T. Fresina, D.A. Ahmari, Q.J. Hartmann, G.E. Stillman, M. Feng, 'Low Phase Noise Ka-band VCOs using InGaP/GaAs HBTs and Coplanar Waveguide', 1997 IEEE MTT-Digest, pp.255-258, 1997.
- [28] A. Hajimiri, T Lee, 'Low Noise Oscillators', ISBN 0-7923-8455-5, Kluwer Academic Publishers, 1999.
- [29] B. Razavi, 'RF Microelectronics', ISBN 0-13-887571-5, Prentice Hill PTR, p. 224, 1998.
- [30] P. H. Ladbroke, 'MMIC Design:GaAs FETs and HEMTs', ISBN 0-89006-314-1, Artech House, 1989.
- [31] N Ono, K Onodera, K Arai, K Yamaguchi, Y Iseki, 'K-band Monolithic GaAs HEMT Driver Amplifiers', Proceedings of 2002 Asia Pacific Microwave Conference, pp. 1390-1392, 2002.
- [32] H. Zirath, R Kozuharov, M. Ferndahl, 'Balanced Colpitt Oscillator MMICs designed for ultra-low phase-noise', IEEE Journal of Solid State Circuits, Vol.40, no. 10, pp. 2077-2086, October 2005
- [33] S. E. Gunnarsson, C. Kärnfelt, H. Zirath, R Kozuharov, D. Kyulenstierna, A. Alping, C. Fager, 'Highly Integrated 60 GHz Transmitter and Receiver MMICs in a GaAs pHEMT Technology', IEEE Journal of Solid State Circuits, Vol.40, no. 11, pp. 2174-2186, October 2005.